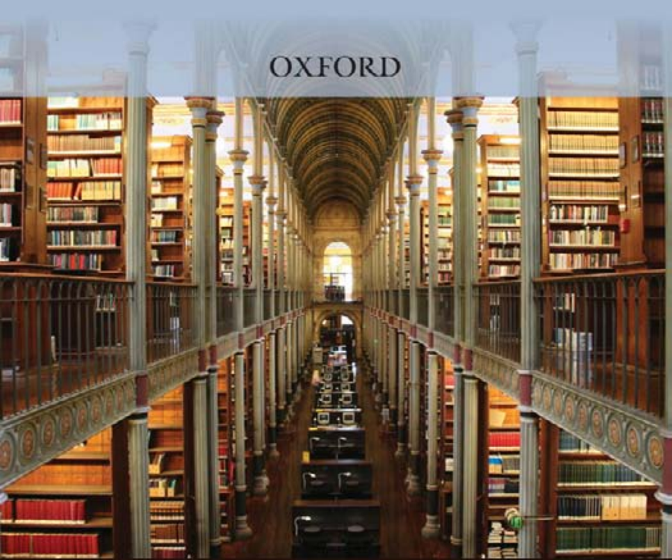
OXFORD



Leila Haaparanta

The Development of Modern Logic

The Development of Modern Logic

This page intentionally left blank

The Development of Modern Logic

Edited by

Leila Haaparanta

OXFORD
UNIVERSITY PRESS

2009

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016

www.oup.com

Library of Congress Cataloging-in-Publication Data
The development of modern logic / edited by Leila Haaparanta.
p. cm.
Includes bibliographical references.
ISBN 978-0-19-513731-6
1. Logic—History. I. Haaparanta, Leila, 1954–
BC15.D48 2008
160.9—dc22 2008016767

9 8 7 6 5 4 3 2 1
Printed in the United States of America
on acid-free paper

Preface

This volume is the result of a long project. My work started sometime in the 1990s, when Professor Simo Knuuttila urged me to edit, together with a few colleagues, a volume on the history of logic from ancient times to the end of the twentieth century. Even if the project was not realized in that form, I continued with the plan and started to gather together scholars for a book project titled *The Development of Modern Logic*, thus making a reference to the famous book by William and Martha Kneale. Unlike that work, the new volume was meant to be written by a number of scholars *almost as if* it had been written by one scholar only. I decided to start with thirteenth-century logic and come up with quite recent themes up to 2000, hence, to continue the history written in *The Development of Logic*. My intention was to find a balance between the chronological exposition and thematic considerations. The philosophy of modern logic was also planned to be included; indeed, at the beginning the book had the subtitle "A Philosophical Perspective," which was deleted at the end, as the volume reached far beyond that perspective. The collection of articles is directed to philosophers, even if some chapters include a number of technical details. Therefore, when it is used as a textbook in advanced courses, for which it is also planned, those details are recommended reading to students who wish to develop their skills in mathematical logic.

In 1998, we had a workshop of the project with most of the contributors present. It was a fine beginning, organized by the Department of Philosophy at the University of Helsinki and by the Philosophical Society of Finland. We got financial support from the Academy of Finland and from the Finnish Cultural Foundation, which I wish to acknowledge. I moved to the University of Tampere in the fall of 1998. Unlike logic perhaps, life sometimes turns out to be chaotic. As we were a large group, it was no surprise that various personal and professional matters influenced the process of writing and editing. Still, we

happily completed the volume, which became even larger than was originally intended.

I wish to thank the contributors, from whom I have learned a great deal during the editorial process. It has been a pleasure to cooperate with them. Renne Pesonen and Risto Vilkko kindly assisted me with the editorial work. I am very grateful to my colleagues for useful pieces of advice. There are so many who have been helpful that it is impossible to name them all. My special thanks are due to Auli Kaipainen and Jarmo Niemelä, who prepared the camera-ready text for publication. Jarmo Niemelä also assisted me with compiling the index. I wish to thank Peter Ohlin, editor at Oxford University Press, who has been extremely helpful during the process. I have benefited considerably from the help of my editors, Stephanie Attia and Molly Wagener, of Oxford University Press. The financial support given by the Academy of Finland is gratefully acknowledged. I have done the editorial work at the University of Tampere, first at the Department of Mathematics, Statistics and Philosophy and then at the Department of History and Philosophy. Finally, I wish to express my deep gratitude to my mother and to my husband, whose support and encouragement have been invaluable.

L. H.

Contents

Contributors

ANDREW ABERDEIN, Humanities and Communication, Florida Institute of Technology, aberdein@fit.edu.

TUOMO AHO, Department of Philosophy, University of Helsinki, tuomo.aho@helsinki.fi.

CALIXTO BADESA, Department of Logic, History and Philosophy of Science, University of Barcelona, cbadesa@ub.edu.

SIBAJIBAN BHATTACHARYYA, Department of Philosophy, University of Calcutta (died in 2007).

MIRELLA CAPOZZI, Department of Philosophical and Epistemological Studies, University of Rome "La Sapienza," capozzi@uniroma1.it.

AMITA CHATTERJEE, Department of Philosophy and Center for Cognitive Science, Jadavpur University.

LEILA HAAPARANTA, Department of History and Philosophy, University of Tampere, leila.haaparanta@uta.fi.

RISTO HILPINEN, Department of Philosophy, University of Miami, hilpinen@miami.edu.

WILFRID HODGES, Department of Mathematics, Queen Mary, University of London, w.hodges@qmul.ac.uk.

TAPIO KORTE, Department of Philosophy, University of Turku, tapio.korte@utu.fi.

ALESSANDRO LENCI, Department of Computational Linguistics, University of Pisa, alessandro.lenci@ilc.cnr.it.

PAOLO MANCOSU, Department of Philosophy, University of California, Berkeley, mancosu@socrates.berkeley.edu.

ARI MAUNU, Department of Philosophy, University of Turku, ari.maunu@utu.fi.

J. N. MOHANTY, Department of Philosophy, Temple University, Philadelphia.

VOLKER PECKHAUS, Department of Humanities, University of Paderborn, peckhaus@hrz.upb.de.

STEPHEN READ, Department of Philosophy, University of St. Andrews, slr@st-andrews.ac.uk.

GINO RONCAGLIA, Department of Humanities, University of Tuscia, Viterbo, mc3430@mclink.it.

S. R. SAHA, Department of Philosophy, Jadavpur University.

GABRIEL SANDU, Department of Philosophy, University of Helsinki; Department of Philosophy, Sorbonne, gabriel.sandu@helsinki.fi.

TUSHAR KANTI SARKAR, Department of Philosophy, Jadavpur University.

GÖRAN SUNDHOLM, Department of Philosophy, University of Leiden, B.G.Sundholm@let.leidenuniv.nl.

CHRISTIAN THIEL, Department of Philosophy, University of Erlangen, Thiel-Erlangen@t-online.de.

RICHMOND H. THOMASON, Department of Philosophy, University of Michigan, rich@thomason.org.

RISTO VILKKO, Department of Philosophy, University of Helsinki, risto.vilkko@helsinki.fi.

JAN VON PLATO, Department of Philosophy, University of Helsinki, jan.vonplato@helsinki.fi.

MIKKO YRJÖNSUURI, Department of Social Sciences and Philosophy, University of Jyväskylä, mikyrjo@jyu.fi.

SANDY ZABELL, Department of Mathematics, Northwestern University, zabell@math.nwu.edu.

RICHARD ZACH, Department of Philosophy, University of Calgary, rzach@ucalgary.ca.

The Development of Modern Logic

This page intentionally left blank

1

Introduction

Leila Haaparanta

1. On the Concept of Logic

When we state in everyday language that a person's logic fails, we normally mean that the rules of valid reasoning, which ought to guide our thinking, are not in action for some reason. The word "logic" of our everyday language can usually be analyzed as "the collection of rules that guide correct thinking or reasoning." That collection is assumed to be known naturally; a rational human being follows those rules in normal circumstances, even if he or she could not formulate them, that is, express them in language. When the word "logic" (in Greek *logos* "word," "reason") refers to one subfield of philosophy or of mathematics, it usually means the discipline concerning valid reasoning or the science that studies that kind of reasoning.

In his logical studies, Aristotle (384–322 b.c.) considered inferences, which are called syllogisms. They consisted of two premises and a conclusion, and the validity of the argument of a syllogistic form was determined by the structure of the argument. If the premises of a syllogism were true, the conclusion was also true. According to Aristotle, the basic form of a judgment is "A is B," where "A" is a subject and "B" is a predicate. Forms of judgments include "Every A is B," "No A is B," "Some A is B," and "Some A is not B." Unlike Aristotelian logic, modern formal logic is called symbolic or mathematical, as it studies valid reasoning in artificial languages.

Until the nineteenth century logic was mainly Aristotelian. Following Aristotle, the main focus was on judgments that consisted of a subject and a predicate and that included such words as "every," "some," and "is" in addition to letters corresponding to the subject and the predicate. The Stoics, for their part, were interested in what is nowadays called propositional logic, in which the focus is on such words as "not," "and," "or," and "if–then."

It was not until the nineteenth century that symbolic logic, which had its model in mathematics, became a serious rival of Aristotelian logic. The grammatical analysis of judgments was challenged in the late nineteenth century by logicians who took the model of analysis from mathematics. The words "function" and "argument" became part of the vocabulary of logic, and predicates that expressed relations as well as quantifiers were included in that vocabulary. In the new logic, which was mostly developed by Gottlob Frege (1848–1925) and Charles Peirce (1939–1914) and which was codified in *Principia Mathematica* (1910, 1912, 1913), written by A. N. Whitehead (1861–1947) and Bertrand Russell (1872–1970), the rules of logical inference received a new treatment, as the pioneers of modern logic tried to give an exact formulation of those rules in an artificial language.

Except for the collection of the rules of valid reasoning and the discipline or the science that focuses on those rules, the word "logic" means a specific language that fulfills certain requirements of preciseness. It also means a field of research that focuses on such a language or such languages. Since the seventeenth century, it has been typical of the field called logic to construct and study a formal language or formal languages called logic or logics.

The old Aristotelian logic heavily relied on natural language. Aristotle and his followers thought that natural language reflects the forms of logical inference and other logical relations, even the form of reality. The pioneers of modern logic sought to construct an artificial language that would be more precise than natural languages. In the twentieth century those languages called logics have been used as models of natural languages; hence, modern logic that rejected the grammatical analysis of judgments has, among other things, served as a tool in linguistic research. It is important to note that the pioneers of modern logic, such as G. W. Leibniz (1646–1716) and Frege, did not intend to present any tools of studying natural languages; they wished to construct a symbolic language that would overcome natural language as a medium of thought in being more precise and lacking ambiguities that are typical of natural language.

As the views of the tasks and the aims of logic have varied in history, we may wonder whether Aristotle and the representatives of modern logic, for example, Frege, were at all interested in the same object of research and whether it is possible to talk about the same field of research. In spite of differences, we may name a few common interests whose existence justifies the talk about research called logic and the history of that field. In each period in the history of logic, researchers called logicians have been interested in concepts or terms that are not empirical, that is, whose meanings are not, or at least not incontestably, based on sensuous experience, and that can be called logical concepts or terms. What concepts or terms have been regarded as logical has varied in the history, but interest in them unites Aristotle, William of Ockham, Immanuel Kant, and Frege as well as logicians in the twentieth and twenty-first centuries. Other points of interest have been the so-called laws of thought, for example, the law of noncontradiction and the law of excluded middle. A third theme that unites logicians of different times is the question of the validity of reasoning.

In several chapters of the present volume, the question concerning the nature and the scope of logic is discussed in view of the period and the logicians that are introduced to the reader.

2. What Is Modern Logic?

The starting point of modern logic is presented in textbooks in various ways depending on what features are regarded as the characteristics of modernity. Some say modern logic started together with modern philosophy in the late Middle Ages, while others think that it started in the seventeenth century with Leibniz's logic. Still others argue that the beginning of modern logic was 1879, when Frege's *Begriffsschrift* appeared.

If the beginning of modern logic is dated to the seventeenth century, its pioneers include Leibniz, Bernard Bolzano (1781–1848), Augustus De Morgan (1806–1871), George Boole (1815–1864), John Venn (1834–1923), William Stanley Jevons (1835–1882), Frege, Peirce, Ernst Schröder (1841–1902), Giuseppe Peano (1858–1932), and Whitehead and Russell. Unlike many contemporary logicians, modern logicians believed that there is one and only one true logic. Leibniz was the most important of those thinkers who argued that the terms of our natural language do not correspond to the objects of the world in a proper way and that therefore we have to construct a new language, which mirrors the world correctly. Following Leibniz, modern logicians sought to construct an artificial language that would be better than natural languages. If we think that this kind of effort is an important feature of modern logic, then we may say that modern logic started with Leibniz. The idea of calculus has also been an important feature of modern logic. Logic has been considered a system which consists of logical and nonlogical vocabulary, formation rules, and transformation rules; the formation rules tell us what kind of sequences of symbols are well formed, and the transformation rules are the basis on which logical reasoning is performed like calculating.

Many early pioneers of modern logic relied on the grammatical subject-predicate analysis in analyzing sentences that was also part of traditional logic, as mentioned above. It was not until Frege's logic that this division was rejected. The division between arguments and functions thus became central in logic. Frege also stressed that it was the distinction between individuals and concepts that he wants to respect. If we stress that feature, we may say that the philosophical ideas of modern logic can be found in medieval nominalists, but that they did not become codified in formal languages until the latter half of the nineteenth century in Frege's and Peirce's discoveries. Those two logicians also made quantifiers into the basic elements of logic. As modern thinkers, many late medieval philosophers were interested in individuals, but the distinction between an individual and a concept was not taken into account in logic until Frege's and Peirce's discoveries. Frege regarded his logic as an axiomatic theory. That feature can also be considered a typical feature of modern logic.

As was said before, it is often thought that Frege's *Begriffsschrift* gave birth to modern logic. In that book there were many logical discoveries, such as the theory of quantification and the argument-function analysis. Frege's book was both philosophical and mathematical. Later, in the first volume of his *Grundgesetze der Arithmetik* (1893), Frege states that he is likely to have few readers; all those mathematicians stop reading who, when seeing the words "concept," "relation," and "judgment" think: "It is metaphysics, we do not read it," and those philosophers stop reading, who, when seeing a formula, shout: "It is mathematics, we do not read it" (p. xii).

Charles Peirce discovered the logic of relatives in the 1870s. That logic was inspired by Boole's algebra of logic and De Morgan's theory of relations. Peirce's articles "The Logic of Relatives" (1883) and "On the Algebra of Logic: A Contribution to the Philosophy of Notation" (1885) contain the first formulation of his theory of quantification that he calls his general algebra of logic. Peirce's algebra differed from that of Boole's especially in that Peirce introduced signs that refer to individuals in addition to signs that signify relations. Second, he introduced the quantifiers "all" and "some." Frege only used the sign for generality and defined existence by means of generality and negation. Both the logicians rejected Boole's idea that judgments are formed by combining subjects and predicates. Frege and Peirce, who made their important discoveries independently of each other, Peirce maybe with his group of students and Frege alone, had common features. They were both philosophers and mathematicians and could combine philosophical ideas with technical novelties in their logical thought.

Frege and Peirce both invented a notation for quantifiers and quantification theory almost simultaneously, independently of each other. Therefore they can be regarded as the principal founders of modern logic. However, as many scholars have emphasized, most notably Jean van Heijenoort in his paper "Logic as Calculus and Logic as Language" (1967), Jaakko Hintikka in his papers "Frege's Hidden Semantics" (1979) and "Semantics: A Revolt Against Frege" (1981), and Warren Goldfarb in his article "Logic in the Twenties: The Nature of the Quantifier" (1979), the two logicians seem to be far apart philosophically. The division between the two traditions to which the logicians belong has also been emphasized by a number of authors of the present volume. The distinction between the two conceptions of logic, namely, seeing logic as language versus seeing it as calculus, has been suggested from the perspective of twentieth-century developments, but the origin of the division has been located in nineteenth-century logic. Different interpretations of the history of logic follow depending on how the distinction is understood. According to van Heijenoort, Hintikka, and Goldfarb, those who stressed the idea of logic as language thought that logic speaks about one single world. It is certain that Frege held that position. He thought that there is one single domain of discourse for all quantifiers, as he assumed that any object can be the value of an individual variable and any function must be defined for all objects. On the other hand, those who supported the view that logic is a calculus gave various interpretations or models for their formal systems. That was Boole's and his followers' standpoint. Several other features of the two traditions are mentioned in the chapters of the present volume.

The volume titled *Studies in the Logic of Charles Sanders Peirce* (1997) introduces another pair of traditions, which are mathematical logic and algebraic logic and which are also touched upon in the present collection of articles. Ivor Grattan-Guinness states in his contribution to the volume on Peirce that the phrase "mathematical logic" was introduced by De Morgan in 1858 but that it served to distinguish logic using mathematics from "philosophical logic," which was also a term used by De Morgan. However, in Grattan-Guinness's terminology, De Morgan's logic was part of the algebraic tradition; using algebraic methods in logic would be typical of what he calls algebraic logic. The most common phrase used in the nineteenth century was "the algebra of logic" or sometimes "logical algebra."

In the figure which Grattan-Guinness presents to us, Boole, De Morgan, Peirce, and Schröder belong to the tradition of algebraic logic, while Peano and Russell belong to the tradition of mathematical logic. It seems that many of those who belong to the tradition of logic as calculus belong to the tradition of algebraic logic in Grattan-Guinness's division, and that many of those who think that logic is a language belong to what Grattan-Guinness calls the tradition of mathematical logic. Grattan-Guinness gives us a few typical features of the two traditions that he discusses. In algebraic logic, laws were stressed, while in mathematical logic axioms were emphasized. Moreover, he states that in mathematical logic, especially in the logicist version represented by Russell, logic was held to contain all mathematics, while in algebraic logic it was maintained that logic had some relationship with mathematics. In Grattan-Guinness's view, algebraic logic used part-whole theory and relied on a basically extensionalist conception of a collection, while in mathematical logic the theory of collections was based on Cantor's *Mengenlehre*. In addition, there was, in his view, an important difference between the traditions concerning quantification; the interpretation of the universal and existential cases as infinite conjunctions and disjunctions with the algebraic analogies of infinite products and sums was typical of the algebraic tradition. Grattan-Guinness also notes that the questions addressed in mathematical logic were more specific than those addressed in algebraic logic.

Frege's and Peirce's logical views are discussed in several chapters of the present volume. Many contributors also touch on the more general question concerning the borderline between traditional and modern logic, the divisions between the traditions of modern logic, and the shift from the modern logic of the late nineteenth century and the early twentieth century till twentieth-century logic. The periods of Western logic that are studied in the present collection of articles extend from the thirteenth century to the end of the twentieth century. Unlike the rest of the contributions, the chapter on Indian logic covers several schools whose history reaches far back in the history but which are also living traditions in contemporary Indian logic.

3. Logic and the Philosophy of Logic

Besides the term "logic," the terms "philosophical logic" and "philosophy of logic" have various uses. Philosophy of logic can be understood as a subfield of philosophy that studies the philosophical problems raised by logic, including the problem concerning the nature and the scope of logic. Those problems also include metaphysical, or ontological, and epistemological questions of logic, problems related to the specific features of logical formal systems (e.g., related to the basic vocabulary of logic) and logical validity, questions concerning the nature of propositions, judgments, and sentences, as well as theories of truth and truth-functions, and the questions concerning modal concepts and the alternatives of classical logic, which some call by the name "deviant logics." The term "philosophical logic" is often used as a synonym of "philosophy of logic"; occasionally it means the same as "intensional logic," or it is used as an opposite to "mathematical logic." By metalogic, one normally means the study of the formal properties of logical systems, such as consistency and completeness, and thus distinguishes it from the philosophy of logic, which studies their philosophical aspects.

The present volume deals with the history of modern logic and pays attention both to the core area of logic and to the philosophy of logic. Such terms as "classical logic," "modal logic," "alternative logics," and "inductive logic" are also used and explained in the chapters of the volume. The variety of logics raises the problem of demarcation that is essential to the philosophy of logic: which formal systems belong to the objects of logical research, and which ought one to exclude from the field of logic? For example, the program of logicism, which was supported by Frege, among others, was a position taken in the discussion concerning the demarcation of logic.

Logic and philosophy have complicated relations. Nowadays logical tools are often used as the methods of philosophy. Logical discoveries have also been motivated by philosophical views, and philosophers have changed their opinions because of logical discoveries. Logic can be said to have a philosophical basis, and likewise there are philosophical doctrines that rely on developments of logic. The present collection of articles studies some of those relations. To some extent, it also pays attention to the relations between logic and mathematics and logic and linguistics.

Logic and rationality are often tied together, but the concept of rationality has many uses in everyday language and in philosophical discussion. We talk about logical or argumentative rationality and refer to one's ability to reason or to give arguments, and we also think that one who is rational is able to evaluate various views critically and independently of authorities; in this latter meaning, logic is considered to play a significant role. Moreover, rationality is both theoretical and practical, the latter form of rationality being related to a person's actions, and philosophers also tend to regard one's ability to control one's volitional and emotional impulses as a sign of rationality. There is no one concept or "essence" of reason that can be detected in philosophical or in everyday discussion. However, what we can find in most uses of the concept is the general idea of control (control of thought, actions, passions, etc.), which is also central in logical rationality.

Even if rationality as control or as rule-following seems to be crucially important, rationality as a faculty of judgment is also in everyday use in the practice of logicians as in all science. In the tradition of logic, it has been important both to be able to follow rules or repeat patterns and to be able to evaluate the commands and prohibitions. It is important both to be able to think inside a given system and to be able to evaluate the very system from the outside. The history of modern logic is a history of these two huge projects.

Philosophers and logicians have used the volume titled *The Development of Logic* by William and Martha Kneale (1962) for decades. The ambitious idea behind the present work was to write a book on the development of modern logic that would bring the history of modern logic till the end of the twentieth century and would also pay attention to the philosophy of logic and philosophical logic in modern times. The idea was not to bring about a handbook but a volume that would be as close as possible to a one-author volume, that is, a balanced whole without serious gaps or overlaps. It was taken for granted in the very beginning that that goal cannot be reached in all respects. Each author has chosen his or her style, some wish to give detailed references, others are happier with drawing the main lines of development with fewer details; some express their ideas in many words, while others prefer a concise manner of writing. However, what has been reached is a story that covers a number of themes in the development of modern logic. The history begins with late medieval logic and continues with logic and philosophy of logic from humanism to Kant, that is, with two chapters whose scope is chronologically determined. Chapters 4–7 cover the nineteenth century and early twentieth century in certain respects, namely, they focus on the emergence of symbolic logic in two ways, first, by paying attention to the relations between logic and mathematics, second, by emphasizing the connections between logic and philosophy. That discussion is completed by a chapter that focuses on the themes of judgment and inference from 1837 to 1936.

The volume contains an extensive chapter of the development of mathematical logic 1900–1935, which is continued by a discussion on main trends in mathematical logic after the 1930s. The subfields of logic that are called modal logic and philosophical logic are discussed in two separate chapters, one dealing with the history of modal logic from Kant until the late twentieth century and the other discussing logic and semantics in the twentieth century. Separate chapters are reserved for the philosophy of alternative logics, for the philosophical aspects of inductive logic, for the relations between logic and linguistics in the twentieth century, and for the relations between logic and artificial intelligence. Eastern logic is not covered, but the main schools of Indian logic are presented in the last chapter of the volume. While the former part of the volume is chronologically divided, the chapters of the latter part follow a thematic division.

Note

I have used extracts from my article "Peirce and the Logic of Logical Discovery," originally published in Edward C. Moore (ed.), *Charles Peirce and the Philosophy of Science* (University of Alabama Press, Tuscaloosa, 1993), 105–118, with the kind permission of University of Alabama Press. The chapter also contains passages from my review article "Perspectives on Peirce's Logic," published in *Semiotica* 133 (2001), 157–167, which appear here with the kind permission of Mouton de Gruyter.

References

Frege, Gottlob. [1879] 1964. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. In Frege (1964), 1–88.

Frege, Gottlob. 1964. *Begriffsschrift und andere Aufsätze*, ed. Ignacio Angelelli. Hildesheim: Georg Olms.

Frege, Gottlob. 1893. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet, I. Band*. Jena: Verlag von H. Pohle.

Goldfarb, Warren D. 1979. Logic in the Twenties: the Nature of the Quantifier. *Journal of Symbolic Logic* 44: 351–368.

Grattan-Guinness, Ivor. 1997. Peirce between Logic and Mathematics. In *Studies in the Logic of Charles Sanders Peirce*, ed. Nathan Houser, Don D. Roberts, and James Van Evra, 23–42. Bloomington: Indiana University Press.

Hintikka, Jaakko. 1979. Frege's Hidden Semantics. *Revue Internationale de Philosophie* 33: 716–722.

Hintikka, Jaakko. 1981. Semantics: A Revolt against Frege. In *Contemporary Philosophy*, vol. 1, ed. Guttorm Fløistad, 57–82. The Hague: Martinus Nijhoff.

Kneale, William and Martha Kneale. [1962] 1984. *The Development of Logic*. Oxford: Clarendon Press.

Peirce, Charles Sanders. 1931–1935. *Collected Papers of Charles Sanders Peirce*, vols. 1–6, ed. Charles Hartshorne and Paul Weiss. Cambridge, Mass.: Harvard University Press.

van Heijenoort, Jean. 1967. Logic as Calculus and Logic as Language. *Synthese* 17: 324–330.

Whitehead, Alfred North, and Bertrand Russell. [1910, 1912, 1913] 1925–1927. *Principia Mathematica I–III*. London: Cambridge University Press.

2

Late Medieval Logic

Tuomo Aho and Mikko Yrjönsuuri

1. The Intellectual Role and Context of Logic

Our aim is to deal with medieval logic from the time when it first had full resources for systematic creative contributions onward. Even before that stage there had been logical research and important logicians. The most original of them, Abelard, achieved highly significant results despite having only a very fragmentary knowledge of ancient logic. However, we shall concentrate on the era when the ancient heritage was available and medieval logic was able to add something substantial to it, even to surpass it in some respects. A characterization such as this cannot be adequately expressed with years or by conventional period denominations; we hope though that the grounds for drawing boundaries will become clearer during the course of our story.

1.1. Studies

It was characteristic of later medieval logic that it was pursued as an academic discipline, as a major component in an organized whole of studies. Indeed, after the Middle Ages, logic has never been allotted so large a share in the activities of the universities. Moreover, logic was connected to certain classical texts that were seen as natural foundations of this science. Thus, it is reasonable first to say something about the system of studies in general and about the nature of these works in particular.

Ever since Rome, school teaching had always centered on the *trivium* of grammar, rhetoric, and dialectic. When schools developed and the most prominent clusters of schools began to turn into universities, these disciplines found their place in the faculty of arts (*artes*). Dialectic, the art of arguing and reasoning, was largely concerned with logical issues, and was often taken to be the most important art of the *trivium*. Thus, the outcome was that every student had to take extensive courses in logic. Perhaps the dialectical background can throw some light upon the linguistic and semantic tone of medieval logical thought.

The faculty of arts was always much bigger than the higher faculties (theology, law, medicine). If there was a theological faculty in a university, it was associated with advanced studies and required a preliminary education in arts. But philosophical and logical research was pursued by theologians even after proceeding to the higher faculty; in fact, the most competent scholars often preferred the privileged higher faculty. Thus the history of logic must take into account the production of both faculties. Many commentaries on Peter Lombard's theological *Sentences* contain important passages on logic, and topics related to logic are often dealt with in the so-called *quodlibetal disputations*, to mention just two examples.

We cannot pay much attention to the history of universities, though we can say that the process of university education started in Italy in the twelfth century, Bologna being the oldest university. Paris, however, was undoubtedly the most important university for philosophy, and it received its official statutes in 1215. Paris was a permanent international center for current philosophical and theological discussion. Another place where logical research was often especially popular was Oxford. These were the two capitals of medieval logic, although the center of gravity shifted to Italy in the less innovative period toward the end of the fourteenth century. During the fourteenth century, universities spread to the east and to the north. There were 15 universities in 1300, 30 in 1400, and about 60 in 1500, naturally of very different size and quality, though one component of studies was standard everywhere, and that was logic.

1.2. The Growth of Logic

Medieval philosophers normally made use of an array of authoritative classical texts, which were taken to be trustworthy, though not infallible. The curriculum was organized around these texts, and very often the problems discussed were put forward as questions of interpretation and explication of the texts. Hence the general breakthrough of Aristotelianism in the thirteenth century represented a great change, establishing Aristotle as the main source of academic studies. But in *logic* Aristotle had even before that been regarded as the greatest of authors, and anti-Aristotelian reactions did not seriously extend to logic. Rather than being rejected in the Middle Ages, Aristotle's own work in logic was built upon and developed ever further toward the end of the period.

The famous standard translation for most of Aristotle's texts was that by William Moerbeke. With the logical works the case was different: Though

Moerbeke translated some of them in the 1260s, the authority of the old Roman translation by Boethius (c. 480–524) remained unquestioned. The *Organon* that late medieval logicians used was the Latin text of Boethius. (For *Posterior Analytics*, no translation by Boethius is known; its standard rendering was made by James of Venice before 1150.) These translations are actually quite accurate, although written in a very formal and literal idiom.

Three concise basic works belonged to the kernel of logic throughout the Middle Ages. These were Aristotle's own short *Categories* and *De interpretatione*, and Porphyry's introduction, *Isagoge*. In addition to these, the so-called old logic (*logica vetus*) used Boethius's logical works and a few minor ancient texts (by, e.g., Apuleius and Augustine). The shape of logic changed considerably when Aristotle's complete works of logic became known in the middle of the twelfth century. That opened the way for the new logic, *logica nova*, and in a relatively short time the corpus of *logica vetus* was practically replaced by new works. Even Boethius's treatises on syllogisms fell into disuse. Except for Aristotle and Porphyry, the only work that retained its place was *Liber sex principiorum*, a treatise explaining the categories that Aristotle himself does not dwell upon.

The period of *logica nova* used as its authoritative corpus all the six works in Aristotle's *Organon: Categories*, *De interpretatione*, *Prior Analytics*, *Posterior Analytics*, *Topics*, and *Sophistici elenchi*. At first, dialecticians were especially fascinated by fallacies and sophisms (*Soph. el.*), but gradually the investigation turned more toward the formal theory of syllogism (*Pr. Anal.*). During the thirteenth century, they encountered problems that could not be answered by straightforward Aristotelian principles, and were thus drawn to new fields of logic. After the introduction of such new subjects, logic came to be called *logica moderna*, in contrast to *logica nova*, now called *logica antiqua*. This way of speaking, however, did not imply any break with the earlier Aristotelian tradition, only an expansion of investigation.

The first complete handbooks of *logica moderna* date from the second quarter of the thirteenth century. The earliest known overview is *Introductiones in logicam* by William of Sherwood from the 1230s, but the greatest success of all was the *Tractatus*, also called *Summulae logicales*, by Peter of Spain (probably from the 1240s). This comprehensive work maintained its status as a famous standard textbook throughout the later Middle Ages and the Renaissance, even in the time of printed books. It also served as the source for numerous shorter courses. Similar ambitious textbooks were written by Roger Bacon (*Summulae dialectices*, 1250s) and Lambert of Auxerre (*Logica*, 1250s). In a way, these works can be seen as a synthesis of the founding period of *logica moderna*: On the one hand, they were the first systematic presentations of whole logic, on the other hand, they completed the new so-called terminist logic. Simultaneously a more profound philosophical discussion was started by the influential Robert Kilwardby, who wrote one of the first commentaries on *Prior Analytics* (1240s).

As to the logical content in the overall presentations of logic, significant advance comes only much later, in the generation of Walter Burley (c. 1275–1344) and William Ockham (c. 1285–1347). Ockham's *Summa logicae* is from the early 1320s, Burley's *De puritate artis logicae* from the late 1320s. These works manifest a turn in logical literature toward new problems and to a more theoretical way of thinking. The greatest representative of the next period is John Buridan (c. 1300–1361?); a comprehensive picture of his teaching in Paris is given in his *Summulae de Dialectica*. In the latter half of the fourteenth century, logic was already highly technical. In particular, a series of Englishmen distinguished themselves, among them William Heytesbury (d. 1372?), Ralph Strode (d. 1387?), and Richard Lavenham (d. 1399). A kind of summary of this stage is the enormous *Logica magna* (c. 1400) by the Italian Paul of Venice (c. 1369–1429).

1.3. Non-Latin Traditions

Our account will be only about the Latin West. The significance of Arabic philosophy must be emphasized, and yet we shall not discuss the Arabic logic per se, as it had its creative phase long before the time of Western late medieval philosophy. Aristotle's *Organon* was translated into Arabic in the ninth century in Baghdad, and a commentary tradition started soon after that. Logic was honored as a kind of grammar of reasoning, and for example, al-Farabi (c. 870–950) underscored its importance as the "forecourt of all philosophy." Avicenna (Ibn Sīnā, 980–1037), on the other hand, was already a brilliant, independent exception: During his time, logical research was already declining, and commentaries were replaced by handbooks. His work had a profound influence on Western theories of meaning. In the twelfth century, the Spanish Arabic school revived commentaries, and the last commentator, Averroes (Ibn Rushd, 1126–1198), was also the greatest. The works of Averroes, "the Commentator," were soon translated and became highly appreciated in Europe. In logic he was not as dominant as in metaphysics or in natural philosophy, but undoubtedly his works belong to the background that was always present. Averroes's thought survived mainly in the West. In the Islamic world, logic was integrated into studies of theology and law, and even handbooks were gradually replaced by more or less elementary textbooks. During the period we describe, from the thirteenth century on, Arabic logic no longer produced anything but new versions and editions of established textbooks.

On the other hand, a rich tradition of Jewish philosophy was alive in Europe through the late Middle Ages. Logic was not its favorite field, but some Jewish authors paid considerable attention to logical questions. However, these studies had little interaction with Latin logic, and thus had to rely solely on Aristotle as commented by Averroes and al-Farabi. Still, there were some innovations, the most interesting figure being Gersonides (1288–1344). Writing in a rigorous manner, he made a number of criticisms of traditional doctrines; among other things, he rejected the old Averroistic construction of modal syllogistics in *Prior Analytics* (thus paralleling contemporaneous Latin developments).

Even today, very little is known about Byzantine logic. Apparently an uninterrupted interest in logic, “the instrument of philosophy,” existed among Byzantine scholars. It produced mainly Aristotelian commentaries, often in the neo-Platonist spirit. Its independent progress was severely hindered by a conservative, philological approach to Greek sources, and occasionally also by religious scruples against the pagan heritage.

1.4. Texts

Aristotle was the essential basis of later medieval logicians, but other classical ideas also played their part. First, Greek Aristotelian commentators had discussed various problems in Aristotle’s logic and its correct systematization, and their work became partly known (either directly or through Arabic sources). Second, the Stoics had argued that Aristotelian predicate logic was insufficient and required some background from the propositional logic that they studied as the real logic. No complete Stoic works were preserved, but these Stoic themes were transmitted, for example, by Augustine’s *Dialectica* and by Boethius. We shall meet similar problems in the medieval theory of demonstration in topics and consequences. Because of the Stoic influence, medieval logicians were always in a different position from the ancient Peripateticians in that they were aware of the necessity of essentially nonsyllogistic inference. Furthermore, logic was obviously influenced by classical grammar, which provided it with categories like nouns, verbs, and other parts of speech, as well as central syntactical notions, the main authority being Priscian’s grammar of Latin. Finally, some logical material had found its way into the work of famous ancient authors, among them Cicero, and the Christian fathers.

From the middle of the thirteenth century, there was a rapid increase both in logical studies and in Aristotelian studies in general. Soon the obligatory logical curriculum included the whole of the *Organon*. Aristotle’s text is so concise and difficult that it was always accompanied by commentaries and explanatory texts. It was required that students mastered this material thoroughly, and practical logical exercises became very popular as a supplement to lectures. A major and growing part of studies was dedicated to logic. If we understand logic in the widest sense, it appears that more than half of the program of an arts faculty could be about logic. At least we may note that logic had an undisputed place in medieval learning, and that it was not a specialist subject since almost all leading philosophers wrote about logic.

Most logical works were closely connected to university teaching. The usual teaching method in medieval universities was that a text was lectured on and explained in detail. The intention was to build a consistent interpretation of the text, to eliminate ambiguities and to resolve the problems and conflicts the text gave rise to, and a typical medieval method of study included disputations where some theses were argued for and against. The character of university teaching goes some way to explaining the literary types that became widespread. In addition to simple lecture drafts, there were all kinds of commentaries, ranging from elementary glosses to large systematic books. There were treatises (*tractatus*), that is, manuals or more advanced surveys of some field, which gradually became more independent of the underlying texts. The liking for argumentation and disputation produced *quaestiones*, analytic works where some specific question is resolved or a thesis defended. (Later, systematic studies were organized in the form of a series of questions even though they were often referred to as commentaries.) We must also remember that logical subjects are often encountered as digressions in other works, for example, in the extensive sentence commentaries of theologians.

There has been quite a decisive improvement in the accessibility of medieval logic over the last three decades, when numerous texts have been published. However, a large amount of material still remains unpublished and even completely unstudied. In fact, it is quite possible that our whole view of the outlines of medieval logic will undergo a change; indeed, such changes have occurred before, and systematic historical research of this logic is still a very young enterprise.

1.5. Interdisciplinary Relations

Obviously, logic had a well-established place in the system of disciplines in the Middle Ages. But what kind of interaction did logic have with the other sciences? Unfortunately, it is not easy to say anything definite about this. First of all, formal philosophy of science was studied by logicians in connection with the *Posterior Analytics*, which discusses the correct form and nature of deductive theories. In this way, the methodology and philosophy of science were a part of medieval logic. Also, the occasional attempts to create calculative scientific speculations used heavy logic, but in general there was little concrete connection of logic to particular natural sciences, which took care of their own subjects. On the other hand, metaphysics—universally considered a real science—was always relevant for logic. Thus, semantic theory, so prominent in medieval logic, is immediately bound to metaphysical questions. Just as early supposition theory employs a metaphysical basis, so in the late-medieval nominalist trend it is impossible to separate logical from ontological thought.

The role of theological matters is less transparent. Obviously theology needed systematic thought and conceptual analysis, and was hence favorable to logic. The conceptual examples and difficulties that logicians examined were very often drawn from theology. Generally, the significance of theology for logic must have been positive. Their union was made problematic, however, when many philosophers began to think that some mysteries of faith, such as the Trinity, were not only inscrutable but literally beyond logic—even if their exact formulation could be a task for logic. Thus Ockham and Buridan thought that certain theological notions had to be explicitly declared unsuitable for use as substitution instances in ordinary logical principles, and a few authors were even more radical on this point. On the other hand, there was—of course—some religious hostility which regarded logical reasoning as an unhealthy method in theological matters.

2. Language as the Subject Matter of Logic

Thinkers in the Middle Ages were anxious to discuss the correct system and classification of sciences. Since their philosophy of science was realist, they believed that the classification should be based on the order of nature. Logic, however, clearly has special features that make its place in this scheme problematic. Is it a science that has as its subject matter some part or aspect of reality? Or is it merely the art of using linguistic idioms? Or is its function something else altogether?

2.1. A Science "of Words" or "of Reason"

The first known medieval textbook of logic, Garlandus Compotista's *Dialectica* from the late eleventh century, already sets the discussion of this topic on a track that was to have crucial influence on the kinds of innovations that were to be achieved in medieval logic. Throughout the Middle Ages, logical theories had a very intimate relation to actual language use. According to Garlandus, logic is concerned with actual utterances (*voces*). After Garlandus, Abelard, for example, continues on the same track, but refines the position: As he sees it, statements are not built from mere spoken sounds but from words that have a signification (*sermones*). Thus, they also constitute the subject matter of logic. Logic is "a science of words" (*scientia sermonicalis*).

It seems that well into the thirteenth century the idea that logic studies actual language use remained basically unchallenged. Teachers and students of logic considered that their studies helped in the acquisition of argumentative skills for actual scientific disputations. Given the status of Latin as the language of all medieval learning, it was natural to make the appropriate logical distinctions from the viewpoint of spoken Latin. This gave an important status to essentially linguistic structures even in the later developments of medieval logic.

In approaching many particular features of medieval logic, it is crucial to remember this pragmatic way of looking at the subject matter of logic. In the Middle Ages, the art of logic was not taken to be concerned with abstract structures in the way modern logic and modern mathematics are, but with actual linguistic practices of reasoning. It was generally accepted that logic is, at least in some sense, a practical science giving advice on how to understand and make assertive statements and how to argue and reason in an inferential manner—though opinions varied whether this practical characterization of logic was accurate in any deeper sense.

In the Arab world logic was thought of in a different manner, and thus toward the thirteenth century under Arabic influences the Latin world became aware of a different way of looking at the character of logic as a field of research. According to al-Farabi, the *logos* (in Arabic, *al-nutq*) discussed in logic occurs on two levels, one inscribed in the mind, and the other existing externally in spoken sounds. Thus, we may even separate different senses of the Greek word *logos* in accordance with the level of discourse at issue. Avicenna was also influenced by al-Farabi's discussion, and gave even further impetus to the idea that logic is concerned with intellectual structures rather than with what we do in spoken discourse. Thus, logic should be called "a science of reason" (*scientia rationis*), as the Latin world translated the idea.

In the thirteenth-century Latin tradition, both the idea of logic as "a science of words" and as "a science of reason" had a foothold. In his major classification of all the university disciplines, *De ortu scientiarum*, Robert Kilwardby (c. 1215–1279) gave a definition of the nature of logic that combined the two views. It is worth taking a closer look at his definition, because it also clarifies the medieval way of locating branches of logic in terms of Aristotle's logical works in the *Organon*.

According to Kilwardby, logic is "a science of words" (*scientia sermonicalis*) in the sense that "it includes grammar, rhetoric and logic properly so-called." But as Kilwardly immediately points out, "in the other sense, it is a science of reason," and in this sense it is "distinguished from grammar and rhetoric." It may seem that here Kilwardby would be demarcating two different disciplines both ambiguously called "logic." But this is not really his intention, as he hastens to explain: Logic properly so-called must in his opinion be listed as one of the "sciences of words"; it is the science of words that attends to their rational content. As he sees it, logic does not study arguments as mere words nor as mere rational structures, but as rational structures presented in linguistic discourse. The grammatical and rhetorical features of these arguments, for example, do not pertain to the art of logic. Logic studies the rational structures expressed and understood in linguistic discourse—neither rational structures as such, nor linguistic structures as such.

The core of logic can in Kilwardby's view be found in Aristotle's *Prior Analytics*. This is because at its core, logic is concerned with reasoning, and this is the main topic of *Prior Analytics* and its system of syllogistic reasoning. It is of some interest to note that Kilwardby is very Aristotelian in claiming that all forms of valid reasoning can be reduced to the categorical syllogism discussed in *Prior Analytics*. This was not the received view at the time, and Kilwardby's position did not win unconditional approval. Abelard had discussed the theory of conditional inference and clearly would not have accepted such a principle. Indeed, conditional inferences were throughout the Middle Ages a standard part of logical curriculum. Soon after Kilwardby, toward the end of the thirteenth century, the theory of consequences (*consequentiae*) grew into a self-conscious general theory of inference that had no specific reference to the syllogistic system; syllogism was increasingly presented as a special case of inference.

Kilwardby pushes aside one aspect of Aristotle's discussion in the *Prior Analytics.* According to Kilwardby, only dialectical and demonstrative syllogisms are relevant to logic, while the rhetorical syllogisms discussed by Aristotle fall out of the scope of logic because they take "the form that is suited to the singular, sensible things considered by the orators." Logic as a science is concerned with universal rational structures as captured in discursive reasoning.

In Kilwardby's presentation of the structure of logic, the system developed in the *Prior Analytics* is put to further use in the *Posterior Analytics* and the *Topics.* As Kilwardby sees it, the division into different works is based on the matter to which the syllogistic structures are applied. The *Posterior Analytics* discusses the way in which the syllogistic form is applied to "specific matter" and yields scientific demonstrations. For its part, *Topics* is concerned with "common matter" and shows how we can construct good inferences relying on generic considerations. Aristotle's *Sophistici Elenchi*, for its part, plays in Kilwardby's view the role of considering what can go wrong in constructing an inference.

As Kilwardby shows, the role of *De interpretatione* and *Categoriae* can also be considered in terms of the syllogism. *De interpretatione* considers the propositional structures that are essential for constructing syllogisms. A syllogism must be construed so that it has a middle term, and for this purpose it is necessary to see how assertive statements usable as premises can be built to consist of two terms conjoined affirmatively or disjoined negatively. *Categoriae* goes into the structure even more deeply, considering the terms and their signification in reality.

From the mid-thirteenth century onward, Avicenna's conception of logic as a science of reason gained increasing currency in philosophical discussions on the subject matter of logic. To some extent this happened at the expense of the earlier view of logic as a science of words. As we have already seen, Kilwardby restricts the meaning in which logic is a science of words so that it no longer carries much weight. Albert the Great's general position concerning the nature of logic is similar, but in the beginning of his commentary on Aristotle's *Categories* he takes the explicit position that logic is strictly speaking not a "science of words" at all. Rather, logic is concerned with argumentation, and argumentation should be referred to reason rather than to words.

Albert's student Thomas Aquinas (1224?–1274) followed him in this matter. In his more elaborate system, the subject matter of logic consists of three conceptual operations of the mind, namely, formation of concepts, of judgments, and of inferences. This systematization can be traced back to Plotinus and the neo-Platonic commentators of Aristotle's logic in a more explicit way than Kilwardby's system. The first two operations are discussed, respectively, by Aristotle's *Categories* and *De interpretatione*, and the third by the other four works included in the *Organon.* As Aquinas saw it, making a judgment—and, in fact, anything that logic is concerned with—requires an intellectual act of understanding. Thus, making a judgment is not primarily to be understood as a speech act but as a mental act. According to Aquinas, externally expressed linguistic structures should be seen as results and representations of intellectual acts, and only in this intermediate way does logic come to be concerned with linguistic structures. In the first place, logic is concerned with the intellectual operations by which the universal features of material reality are understood.

The detailed structure of Aquinas's presentation of the subject matter of logic has the crucial feature that it relies heavily on the Aristotelian idea that all inferences can be presented as syllogisms. As Aquinas saw it, all of logic can be understood in terms of syllogistic structures. Since he thought that logic deals with the three basic operations of the intellect, any inference would have to be based on them. However, there are understandably quite stringent limitations on the extent to which logic can be derived from these basic operations. For example, with the claim that all assertions are made by the composition of a predicate with a subject, Aquinas was almost forced to reject conditionals as assertions. However, hypothetical propositions had a long tradition deriving from Stoic logic and had been dealt with already by Boethius, and thus Aquinas was compelled to comment on them. As he put it in the first section of his commentary on *De interpretatione*, hypothetical propositions "do not contain absolute truth, the understanding of which is needed in demonstration ... but they signify something to be true on condition." According to Aquinas's logic, conditionals could not be used as premises in scientific demonstrations.

Neither Albert nor Aquinas worked much with the actual details of logical systems, and their discussion has more of the character of the philosophy of logic. However, the distinctive flavor of medieval logic showed itself in its close connections to actual language use, and it incorporated analysis of a much wider variety of linguistic structures than the simple predications included in the syllogistic presented in the *Prior Analytics*. Moreover, Abelard's work had already made medieval logicians acutely aware of a concept of inferential validity that was essentially unconnected to the syllogistic structure. While Kilwardby, Albert the Great, and Aquinas defended the strong Aristotelian program of reducing all inferential validity and thereby all logic to an analysis of the syllogistic system, actual work in logic was taking another course.

In the subsequent development, Aquinas's three operations of the mind were often referred to, but usually understood in a loose and suggestive manner. It became standard to treat logic with the organizational principle that *Categories* studies concepts and *De interpretatione* propositions, while *Prior Analytics* and the three subsequent Aristotelian works concentrate on inferences. It was, furthermore, commonly accepted that there are many traditional logical genres inherited from the twelfth century that do not fit into this basic scheme. For example, there was an abundance of literature on the so-called syncategorematic terms, analyzing the logical properties of words such as "except" (*praeter*), "begins" (*incipit*), "whole" (*totum*) and "twice" (*bis*). Such problems had little connection to the development of syllogistic systems. Furthermore, it remained a problem to explain how such logical genres could fit into the description of logic as a science of reason, because many of them were quite clearly motivated by the analysis of linguistic structures.

2.2. Mental Language

An interesting alternative way of characterizing logic as concerned with mental concepts rather than Latin words was being developed at the time Aquinas was working, and it gained momentum among logicians in the latter half of the thirteenth century. It was based on quite a different understanding of the workings of the human mind from that of Aquinas's Aristotelian outlook. Roger Bacon (c. 1214–c. 1293) rejected the idea that the human understanding works only with real universals existing intentionally in the mind. Rather, the mind should be understood in terms of a discourse consisting of singular acts of intellection whereby different singular things are understood in different ways. According to Bacon, logic is not concerned with an external discourse but with the internal discourse of the mind, with "mental expressions and terms" (*dicciones et termini mentales*). In other words, Bacon posits a mental language to serve the role of the subject matter of logic. As we shall see, this approach was to play a major role in later developments.

First, however, we must take a closer look at the content of Bacon's suggestion. One of the central classical texts that Bacon refers to was Boethius's distinction between three levels of discourse (*oratio*): intellectual, spoken, and written. In making the distinction, Boethius was commenting on Aristotle's *De interpretatione* 16a10, and Boethius's way of reading the passage was well known in the late Middle Ages, but it remained a debated issue how one should understand the intellectual level of discourse and how one should relate logic as a discipline to these levels.

It seems clear, though, that Bacon understood the intellectual discourse in a way that can with good reason be called linguistic. He even takes pains to show how word order functions in this discourse. Without going into details, it is sufficient here to point out that he looked at the structure of mental sentences in terms of Aristotelian predication: The subject comes first, then the predicate, both with their "essential determinations." They are then followed with the various "accidental parts" of the composition. Especially his way of dealing with these "accidental parts" shows how looking at thought as a linguistic phenomenon gives Bacon a clear advantage in comparison to Aquinas from the logician's point of view. Through his theory of mental language, Bacon is able to attribute considerably more logically relevant linguistic structure to the intellectual level. One of the aims of this enterprise was—as is evident to any logician—to show how to solve ambiguities of scope arising in Latin through the relatively loose rules concerning word order. In this way, Bacon worked toward a theory of an ideal language to serve logical functions as early as the 1240s, if the current scholarly opinion of the date of his *Summa de sophismatibus et distinctionibus* is correct.

The mental discourse that Bacon was after is abstracted from spoken languages like Latin and overcomes their arbitrariness. However, there is also the other side of the coin: He specifically wanted to find the basis for certain logically relevant Latin structures from mental discourse. Although he thought that grammatical gender has no correlate in mental discourse, the subject-predicate structure and many syncategorematic expressions have. Indeed, Bacon seems to find from the mental discourse even more than a logician would need. In many issues, it becomes apparent that he was working more as a linguist than as a logician. His aim was a universal grammar rather than a universal language suitable for logic.

Commentators have, accordingly, connected Bacon to the movement of speculative grammar emerging in the latter half of the thirteenth century. The approach to linguistic analysis employed by this school is often called "modist." The label reflects the specific use of a threefold series of concepts: "Modes of being" (*modi essendi*) in reality were paralleled in language by "modes of signification" (*modi significandi*) and in the mind by "modes of understanding" (*modi intelligendi*). The movement was more closely connected to language theory than logical theory, and accordingly we will only discuss it briefly here.

The main idea of modist theory was to approach Latin expressions as generated from a universal grammatical structure accurately reflecting the structure of reality. That is, they thought that grammar is (in the words of Bacon) "substantially one and the same in all languages, although varied in its accidents." Other central figures of this movement include Boethius of Dacia, Martin of Dacia, and Radulphus Brito. At the beginning of the fourteenth century, the program lost ground, although much of the terminological innovations, including the term "mode of signifying," survived until the Renaissance in the standard vocabulary of logicians.

According to the modists, all words have two levels of meaning. Words have in addition to their own specific meanings certain more general meanings, or so-called modes of signifying. To be more exact, a phonological construction gains a special meaning when it is connected to a referent that it "is imposed" (*imponitur*) to mean (in the so-called first imposition). Furthermore, the word is also "imposed" (in the second imposition) to mean its referent in a certain grammatical category with certain modes of signifying. For example, pain can be referred to by a variety of Latin words in different grammatical categories: *dolor* refers to it as a noun, *doleo* as a verb, *dolens* as a participle, *dolenter* as an adverb, and *heu* as an interjection. In all these words the special signification is the same, but the modes of signifying are different.

The modists found no theoretical use for the most central logical term of the terminists, "supposition" (*suppositio*; it will be described with more detail in the next section). In their view, the varieties of ways in which words are used in sentential contexts are based on modes of signifying contained in the words, and thus they were not willing to admit that the sentential context as such would have an effect on how the term functions—which is one of the leading principles of the supposition theorists. Rather, their approach was generative in the sense that the sentences were to be generated from words that have their signification independently. This approach made it unnatural to distinguish the sentential function of a term from its signification. It may, however, be noted that the term "consignification," meaning the function of syncategorematic terms in the terminist approach, was used by modists to express the way in which phonological elements of actually used words mean modes of signification: For example, the Latin ending *-us* "consignifies" nominative case, singular number, and masculine gender. The thirteenth-century grammarians recognized the congeniality of syncategorematic terms and modes of signifying: Both are understood as the elements of discourse that show how the things talked about are talked about and what in fact is said about them.

From the viewpoint of the history of logic, it is important to recognize that from the twentieth-century viewpoint, the modist conception of grammar can be characterized as making the subject a "formal science." The criteria of congruence were taken to depend solely on the grammatical structure, or the consignifications of the elements of the sentence, regardless of the special significations of the terms used in the sentence. Modists thought of the generation of language as putting semantically significant elements into grammatical structures. It seems that at least the Parisian master Boethius of Dacia wanted to develop also logic into this direction and wanted to make a distinction somewhat like the twentieth-century distinction between logical form and semantic content. Nevertheless, it was only some decades later at the time of John Buridan that the substance of logic was thoroughly reconsidered from this viewpoint.

2.3. The Universality of Logic

From the viewpoint of practicing logicians, the debate concerning the subject matter of logic at the end of the thirteenth century probably seemed like a search for a credible account of the universal basis of the invariable features of argumentation found in the logical analysis of actual use of language. That is, what is the universal basis on which the validity of an inference formulated in a particular language is grounded?

It was accepted as relatively clear that logic is about actually or potentially formulated tokens of terms, propositions, and arguments that are linguistic in some sense of the word. It was clear that such discursive arguments existed in such external media as spoken or written Latin expressions. However, logic aimed at, and appeared to have found, some kind of universality, and such universality apparently could not be achieved if logic was tied to a particular spoken language. Instead, thirteenth-century discussions converged in finding the universality of logic in intellectual operations. But what are these intellectual operations? Can we speak of a mental language serving as the domain of logic? In particular, is a mental proposition linguistic in any relevant sense? And because it was assumed that an affirmative predication is based on or performs a composition, one had to ask what exactly does this composition put together.

At the turn of the fourteenth century, we find different logicians giving different answers to these specific questions. At that time, the most common answer was the one inspired by Bacon. It was based on looking at the mental discourse from the viewpoint of "imagined spoken words," and accepting it as the privileged medium of logical arguments. This kind of explanation is straightforward and relatively acceptable from the metaphysical viewpoint, but is, of course, less satisfactory in explaining the kind of universality achieved in logic. If mental language is nothing but imagined Latin words, there seems to be little reason for assuming it to have any more universal status than Latin has. Yet that appears to be what Bacon wished to propose.

The realist Walter Burley seems to have approached the problem from the viewpoint of the universality of logic. Given his realist metaphysics, it is understandable that he contributed the concept of "real proposition" (*propositio in re*). He aimed at explaining mental propositions as consisting of real external things, which are conceived and propositionally combined in the mind. This model of the metaphysical basis of mental language of course works only if conceptual essences are understood in a realist way without separating them from the things themselves. Also, such "real propositions" are not very language-like.

The nominalist William Ockham formulated the most innovative and by far the most influential theory of mental language. He ridiculed the position of Burley by asking how it could be that the subject of a proposition formulated in Oxford is in Paris while the predicate is in Rome. A suitable example of such a proposition would be "Paris is not Rome." Ockham seems to have gone back to Bacon's theory, but with the awareness of some of its shortcomings.

With his nominalist metaphysical outlook, he strongly held the view that all the metaphysically real things involved in mental propositions are particular mental acts or states. But the substantial logical strength of his theory of mental language was really that it was formulated in a way that was sufficiently neutral from the metaphysical point of view. Indeed, Ockham himself started with the idea that mental language consists of *ficta* (that is, of intellectually imagined objects of thought that do not have any kind of existence outside the mind but are simply "made up" by the mind) but ended with the view that mental language is better understood as consisting of intellectual acts intentionally directed at real or possible things. At one stage of his career he was working on the theory of mental language without being able to make up his mind which of these two rather different views would provide the appropriate metaphysical foundations.

In the first chapters of *Summa Logicae*, William Ockham addresses the Boethian idea of three levels of language. In opposition to Aquinas's treatment of the same topic, Ockham claims that written language is subordinated to spoken language rather than signifies it. Similarly, spoken language is subordinated to mental language rather than signifies it. That is, according to Ockham, all three languages similarly signify things in the external world. They are, furthermore, all equally languages. Written language is inscribed on external material things, and spoken language exists as a continuum of sounds. Similarly, mental language consists of real qualities of the thinking mind. Furthermore, in Aquinas's picture intellectual acts were the significations of linguistic expressions and by their nature could not serve as a medium of communication. For Ockham, mental language could by its nature serve equally as a medium of communication if only there were beings who could perceive its expressions apart from the "speaker" him- or herself. In fact, Ockham thought that we have every reason to suppose that the angels described in the Christian doctrine communicate in the same language in which we think.

The main difference between mental language and the two other kinds of language is the naturalness of mental language. Unlike spoken ordinary languages, which we nowadays call natural, Ockham's mental language is natural in the sense of not being conventional. The expressions of mental language have their significations naturally, without explicit or implicit consent or any other kind of conventionality involved. A mental word is capable of signifying only the things it really signifies, and it signifies exactly those things to all competent users of the word. (It may be noted that Ockham admits that in angelic communication some mental expression may be unfamiliar to the perceiver and thus unintelligible to him.) In principle, there are no ambiguous terms in the mental language. This is one of the central features that make Ockham's mental language an ideal language, which is then suitable for the purposes of a discipline like logic.

There are also two other senses in which Ockham aims at description of an ideal universal language. On the one hand, he tries to describe in general terms what must be required of any language that is used for thinking, and assumes that mental language has only such necessary features without any accidental "ornaments of speech." Since these features are necessary requirements of thought, all thought must comply with them. Thus, Ockham constructs a theory of a language that is universal in the sense of being used by all intellects that think discursively.

On the other hand, according to Ockham, mental language is directly related to the constitution of the world. It reflects accurately mind-independent similarities between real things. Thus, a fully developed mental language would be universal in its expressive power: There cannot be any feature of the world that could be conceived by an intellectual being but not expressed in mental language. Everything that can be thought can also be cast in terms of mental language. From this principle it also follows that all linguistic differences between expressions of spoken languages that result in different truth values (which are not "ornaments of speech") have their correspondents on the level of mental language.

From the logical point of view, perhaps the most interesting ideal feature of Ockham's mental language is its compositionality, which makes it a recursive system. Complex expressions get their meaning from their constituent parts in a systematic way. In this respect, mental language shows similarities to twentieth-century formal calculi, although it is much more complex.

According to Ockham, the expressions of mental language consist of categorematic and syncategorematic parts with specific linguistic roles (we will return to this distinction more fully in the next section). A categorematic term (e.g., "animal") signifies real individuals and refers to them as the other elements of the propositional context determine. A syncategorematic term (e.g., "every") does not signify any external things but rather, as Ockham puts it, "performs a function with regard to the relevant categorematic term." Typically, syncategorematic terms affect the way in which the significations of the categorematic terms result in reference (or *suppositio*) in the sentential context. We may say that the categorematic terms of a sentence determine which things are talked about, whereas the syncategorematic terms determine how they are talked about and what is actually said about them. The number of basic categorematic terms of the ideal mental language accords to the variety of things that could exist in the world; they express the natural kinds of possible things. Ockham's view of the number and selection of syncategorematic terms is more difficult to determine. On the one hand, it is clear that he is thinking of a much wider variety of such logical constants than twentieth-century logic used. On the other hand, it is equally clear that most of his logical rules concern the effects of syncategorematic terms on logical relations between sentences.

Because the compositional characteristics of mental language depend on the distinction between categorematic and syncategorematic terms, Ockham's mental language seems to conform to the twentieth-century ideal principle of logical formalism, namely, the idea that all sentences directly reveal their logical form. This seems to be one of the features of the mental language that Ockham is most interested in, and much of his logic is devoted to systems elaborating on the functions of syncategorematic expressions. However, Ockham's theory has interesting details that reflect a conscious decision not to accept logical form (as we nowadays understand it) as the guiding universal principle in determining the logical validity of an inference.

The theory of mental language was also discussed and developed after Ockham, but without major revisions. The most important innovator was John Buridan, who altered much of the terminology used in defining language and gave a rather different account of how the simple terms of language are learned, but these revisions resulted in few changes that would be relevant to our purposes here. After Buridan, some minor topics like the role of proper names and individual terms, and the nature of word order as explanatory of issues of scope were discussed. These can hardly be called revolutionary with regard to the nature and purposes of logic.

At the peak of its success, medieval logic had thus found a definition of its subject matter that provided a relatively reasonable explanation both of its universality and of its dependency on discursive linguistic structures. For the logicians of the second quarter of the fourteenth century, logic was the art of constructing and using mental propositions. It studied the basic syntactic features of mental language, the ways and forms of assertions that can be produced in it, and the ways these assertions can be organized in inferential relations. Because mental language was understood as capable of expressing all possible universal structures of discursive thought, logic studied the universal art of reasoning.

3. Terms

3.1. The Notion of a Term

Textbooks of "traditional logic" used to divide their material into three sections: the doctrines of terms, propositions, and inferences. This practice is based on ancient grounds, of course, but Aristotle nowhere says that all logic should be so divided, and medieval logic did not at first do so. In thirteenth-century logic books, often the chapters are still relatively independent, or at least not organized according to such a general plan. But then, at the turn of the century, this idea soon became dominant. We find it, for example, in both Burley and Ockham, in spite of their sharp disagreement. We are going to follow this familiar order, starting with *terms.*

Everybody agreed that terms were the ultimate units of discourse. In a way this is obvious, but the emphasis on this fact in logical contexts also has a nontrivial sense which shows the Aristotelian character of medieval logic. For the logic that Aristotle had developed was term logic, unlike that of the Stoics. But Aristotle gave two different explanations of terms. In *Categories* he speaks about noncomposite expressions ("such as 'man', 'ox', 'runs', or 'wins' "). In *Prior Analytics* he says (24b16–18): "I call that a term into which a proposition is resolved, i.e., the predicate or that of which it is predicated, when it is asserted or denied that something is or is not the case." These explanations lead to very different uses of the word "term."

In the first sense, a term is simply any word. Many medieval logicians mentioned even meaningless words, like "ba," "bu," but only to concentrate on ordinary words. In this sense, which is that of grammarians, it is only required that a term is a noncomposite significant element of the language. Or it can be a composite expression signifying one thing.

In the second sense, which is more exciting for the logicians, a term is something that can stand as a subject or a predicate of a proposition. This excludes wide classes of words from the status of terms. According to the strictest definition, a term is only that type of nominal expression that can figure as S or P in a categorical proposition "S is P."

This leads to a question concerning the structure of terms because S and P can be complex expressions. Buridan, for instance, took a strictly propositional view and argued that a simple proposition has exactly two terms. In this usage a term is identified with an extreme (*extremum*) of a proposition. But since the various words occurring within such terms can be extremes in other propositions, authors kept on saying that propositions can have complex terms that are composed of simple terms.

The theory of terms is obviously connected to grammar, and Priscian's classifications had a strong influence on earlier writers. But logically it was important to eliminate Latin contingencies and consider as general cases as possible. However, that is a problematic requirement regarding terms: What could be those language-independent terms? Different ways to tackle this question systematically were offered first by so-called speculative grammar, and then by the mentalistic interpretation of language, which was finally victorious, but both approaches emphasized the universality of language. For late medieval logicians, the terms were in the first place mental terms that occurred in mental propositions.

3.2. Categorematicity

A distinction that is especially important for logic was made between *categorematic* and *syncategorematic* terms. This distinction was well known to all logicians, and they usually introduced it immediately after the definitions of terms. The source of these notions was in grammar, but logicians gave them a new function, following a hint from Boethius. Priscian had written about "syncategorematic, i.e., consignificant, parts of speech": Most words are grammatically categorematic since they can occur as subjects or predicates, but for instance, conjunctions, prepositions, adverbs, and auxiliary verbs cannot. They are syncategorematic and signify only together with other words. Logicians proceeded from this picture to distinguish two ways of meaning and to describe the logical behavior of philosophically interesting syncategorematic words.

Syncategorematic words were first studied in special treatises. This genre of *Syncategoremata* was popular from the last quarter of the twelfth century to near the end of the thirteenth century. Well-known treatises of this kind were written by Peter of Spain, William of Sherwood, Nicholas of Paris, and even the famous metaphysician Henry of Ghent. Later, the subject was incorporated into general textbooks of logic. The distinction itself had its systematic place at the outset of the exposition of the theory of terms, since it was utilized in many questions; particular syncategoremata were then discussed in their due places.

Even in the fourteenth century most authors apparently based their definitions of syncategoremata on different ways of signifying. According to Ockham, "categorematic terms have a definite and determinate signification. . . . Examples of syncategorematic terms are 'every', 'no', 'some', 'all', 'except', 'so much', and 'insofar as'. None of these expressions has a definite and determinate signification." Buridan states: "Syncategorematic terms are not significative *per se*, as it were, but only significative with another." Paul of Venice still defended this view against "a common definition" that a syncategorematic term cannot be the subject or the predicate or a part of either. Such a purely syntactical criterion had been supported by Albert of Saxony (1316–1390).

Syncategorematic expressions were usually counted as terms. A theoretical reason for this was a slogan that was in use at least from Peter of Ailly onward: A term is a sign that in a proposition represents something or *somehow*. Syncategoremata, indeed, signify "somehow" (*aliqualiter*), for thirteenth-century treatises had already pointed out that syncategoremata serve to show how the categorematic terms ought to be understood. It is thus essential for their signifying that they are joined with other terms to elaborate their meanings.

Present-day readers will easily associate syncategorematic terms with logical constants. This is partly correct but must not be taken too literally. For one thing, the class of syncategoremata of language is much wider than the small sets of logical constants nowadays. However, the medievals ignored most syncategoremata and studied only those which seemed to be philosophically interesting. These were just words with special logical peculiarities, and hence, for these terms, the comparison with logical constants may be justified. The lists of different logicians varied greatly, but several dozens of words were thus discussed. Among them belonged sentential connectives; words like "only" and "except"; quantifiers; modal operators; words like "whole" and "infinite"; some verbs like *incipit* and *desinit* ("begins" and "ends"); and the copula *est*, that is, the copulative use of the verb "to be," *esse*. General textbooks listed them but did not usually go into details of particular syncategoremata. In the fourteenth century, such closer study often took place by means of sophismata: In this literature it was typical to analyze sentences that were problematic or ambiguous because of syncategorematic words (see section 6).

Buridan expressly said that the matter of a proposition consists of purely categorematic terms while syncategoremata belong to its *form*. From this point of view, it is interesting to notice that the notion of syncategorematicity proved difficult because it did not determine a precise class. Thus Buridan had trouble with attitude operators: Verbs like "to know" and "to promise" clearly have a formal function and yet they are independently meaningful. The two criteria, the semantical and the grammatical, did not always coincide, and Peter of Ailly suggested that they should be wholly separated. A term could therefore be syncategorematic either "by signification," or "by function," or in both ways.

3.3. Predicables

In a proposition something is said of something, as Aristotle taught. It is therefore logically important to have some idea of the various types of things that can be thus predicated, the *predicables* (*praedicabilia*). Medieval logicians based their classification here on Porphyry and Boethius. Obviously, a predicable is something that can be said (predicated) of something else, but in a stricter sense, it is only a universal term that can be predicated of many things. This distinction was made already in thirteenth-century textbooks, and it is easy to see that predicables have a close connection to the most famous medieval metaphysical problem, the problem of universals.

Explaining Aristotle's *Categories*, Porphyry mentioned five types of universal terms: *species*, *genus*, *differentia*, *proprium*, and *accidens*. These were the "five universals" (*quinque voces*) that recur in medieval discourses. They reveal various relations of the predicate to the subject: What kind of information does the predicate give us about the subject? When it is said that S is P, the predicate P may express a species to which S belongs, or a genus to which every S belongs, or a characteristic essential feature of them (*differentia*), or a nonessential but necessary property of every S and only them (*proprium*), or their accidental feature (*accidens*). (The P of species is a somewhat obscure case here because it can be predicated of individuals, too, unlike the others.) Added to a genus, a specific difference (*differentia specifica*) defines a species, which in turn can be a genus for lower subspecies. In this way, the famous "Porphyrian tree" is generated, ranging from uppermost genera down to individuals.

The doctrine of predicables was a standard part in medieval logic texts, and it was a relatively unproblematic part: The difficulty, of course, is metaphysical and concerns the essential, necessary, and accidental qualities. Logicians, however, used the five universals as metalinguistic tools to classify predicates. A more ontological question is that of categories, or *praedicamenta*, as logicians preferred to call them. The first category is substance; the other categories are ways in which something can belong to a substance. Aristotle studied quality, quantity, and relation in his *Categories*, and more briefly he discussed even place, time, position, habit (having), passion, and action. With some variants, medieval *praedicamenta* treatises give the same list of 10 members.

As Buridan says, "this treatise is found in many summulae, but in many it is not." Indeed, it is not obvious why logicians need to discuss a question that seems purely metaphysical. But there was a motive for those who included this treatise in their summulae—an assumption of the parallelism between predication and being. Except substance, all categories both "are said of things" and "are in things." Thus, a classification of ways of being in a substance also produces a classification of questions and answers that can be made concerning an entity, and this is a logically relevant achievement. Later, nominalists give up the assumed parallelism and analyze categories simply metalinguistically, as classes of terms. Ockham, for instance, has a long discussion in which he wishes to show how terms of other categories are secondary to substance and quality.

We may note in passing that predicables and categories have a very different role among the speculative grammarians of the late thirteenth and early fourteenth centuries. For them, terms are intelligible because they manifest the same characters and structures as the entities of the world; the "modes of signifying" belonging to grammatical features of lexical meaning and inflection are functions that reflect categorial features of objects. Such an approach leads to a special view of metalinguistic issues. Hence it is also natural that these authors, the modists, concentrated on rather abstract lexical contents and were not very interested in the semantic properties of concrete occurrences of terms in particular sentences.

3.4. Significance

The main body of the theory of terms consisted of *proprietates terminorum*. The tireless analysis of these "properties of terms" displays the intense interest in *philosophical semantics* that was characteristic of later medieval philosophy. This is a field that seems to be a medieval invention. In Aristotle and other ancient sources, there were only scattered remarks on semantic questions, and it can hardly be said that they attempted to establish any self-conscious theory of semantics. On the other hand, after scholastic philosophy these problems were often considered futile, and explicit philosophical semantics was largely rejected. But the medieval theory has had a striking revival in the latter half of the twentieth century, when philosophical semantics has again grown into a complex discipline, often struggling with questions that bear an obvious resemblance to medieval themes.

Undoubtedly the two most important properties of terms are *signification* and *supposition*. They have often been compared to present-day "meaning" and "reference," but this comparison must not be taken literally. For one thing, the emphasis was on the words and signs: Unlike many accounts of meaning and reference, the medieval doctrine viewed signification and supposition mainly as something that the words do or as something that is done by means of words.

Let us start with *signification*. Logicians were aware of the ambiguity of this word. Usually, instead of interpreting signification as a signified entity of some sort, they started from "acts of signifying" and assumed that terms had a property of being significant. (In this respect, terms differed from other words which had no signification by themselves.) A word signifies, or has signification, because of its "institution," or according to another common account, because of its "use" in language. In short, signification is the role of the term in language. The same idea acquires a new slant with the introduction of mental terms. It then becomes standard to claim that spoken words have their significations because of linguistic conventions, whereas the mental terms are *natural signs* that have their significations necessarily, without any stipulation. Signification is generally connected to mental acts of understanding: A linguistic term signifies that of which it makes a person (a speaker or a hearer) think, a mental term is itself an act of thinking of something, a representation. (To quote John Aurifaber: "signifying is an accident of the intellect, but a word is the thing by means of which the intellect signifies.") The thing thus signified has "intentional being."

Even before the mentalistic turn, it was usual to find the essence of words in signification. Thus Thomas Aquinas said that "signification is like the form of a word"—the matter was the phonological shape, the form was its signifying capacity. Later, it was said that mental concepts have their significations "formally," and spoken and written words essentially function as instruments of this signification.

Signification is the defining property of all terms; thus it is natural that it can be defined no further. Late medieval philosophers seem to agree that signification is a basic notion that can only be explained by illustration. For them, it was obvious that a term signifies something, but there was a great, partly metaphysical controversy about what this something is.

Boethius had already said that words signify concepts, that is, corresponding mental entities. This gave an impulse for the view that words signify concepts immediately and objects indirectly. (Such a "semiotic triangle" had been discussed earlier by Greek Aristotelian commentators.) This opinion became prevalent among the Thomists. Aquinas himself had pointed out that a term signifies a general nature that is abstracted from individual entities. The later Thomists emphasized that the concepts were signs, too: Thus the words do signify objects "principally" (most important), although they do it only "indirectly" (through the concepts).

A contrary position was championed by Bacon, and it won general support at the end of the thirteenth century. It started from the obvious fact that terms are used in propositions, and the propositions are about objects and not concepts. Thus all terms must signify objects. However, nonexistent objects cause problems which compel logicians to make reservations concerning that general principle. What is signified is, for instance, the object "regardless of its being or not being" (Kilwardby), or the object "*secundum quod* the intellect perceives it by itself" (Duns Scotus). Moreover, spoken words and mental terms signify the same objects. According to Ockham, it is a basic fact that words are "subordinated" to the corresponding mental terms in such a way that they signify the same things. He apparently did not think that this use of language could be further explained. Buridan was not satisfied with this kind of answer and again interpreted the subordination as a type of signification: Words also signify concepts, in some sense. Later discussion became rather complex when different positions were combined and refined.

Admitting then that terms signify something extramental, it is still not clear what this *significatum* is for general terms. The question is inevitably connected to the theory of universals. The realist answer is that the term signifies something general; "man" signifies a universal, a species, a property, or a common nature of "man in general." The nominalist answer is that the term signifies all relevant individuals; "man" signifies each man. Both answers cause trouble, which shows the uneasy union of signification and denotation. For it was assumed, after all, that a term signifies what it is true of, and this characterization would better suit denotation.

Syncategorematic words have no signification in the strict sense. However, most logicians were not as rigorous as Ockham, who said that they do not signify at all. Even Buridan was willing to admit that they did not signify things but ways of thinking. And both realists and nominalists agreed that syncategorematic words could "consignify," that is, participate in forming significant wholes.

There is even another sense of the word "consignification." In addition to its basic signification, a word can have some consignification that further determines its content. Especially thirteenth-century authors often use this approach to explain the role of features like case endings and—the most discussed example—tenses. The idea is that the actual occurrences of words get richer contents than bare lexical words.

3.5. Supposition

Denotation was first discussed by means of *appellation*, a notion borrowed from the "appellative nouns" of grammar. Appellation is the relation between a general term and the things actually belonging under it at the moment of utterance. Often this notion was applied only to the predicates of propositions, but at least from William of Sherwood onward it had unrestricted use.

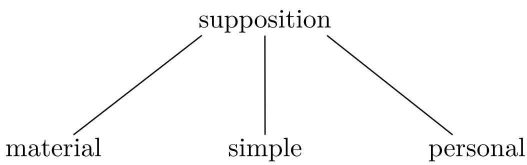
The "property of terms" that caused the most extensive study was *supposition*. The word derives from grammatical contexts. According to Priscian, a word has a supposition when it is placed as the subject of a proposition. This meaning was usual in the twelfth century. On the other hand, grammar had also formed the idea that a word supposits because it refers to an individual. Gradually this became the central aspect, and the supposition of terms was their way to denote individuals. As the supposition theory expanded, logicians had to seek for suppositions even for other terms than the subjects of propositions—for predicates and parts of complex terms. The question of supposition began to concern the denotation of terms quite generally, and at the same time appellation lost much of its importance, turning into a special case of supposition.

Supposition theory was a challenging subject especially because the supposition of terms depends on their position in a proposition. Each word that is not equivocal always has the same signification, but its supposition varies in different propositions. As Ockham said, "supposition is a property of a term, but only when it is in a proposition." This compelled the logicians to develop classifications for the several kinds of supposition. As many scholars have pointed out, precisely this propositional approach was characteristic of the theory of supposition. It must, however, be noted that Peter of Spain admitted even a "natural supposition" (*suppositio naturalis*) that belongs to a term immediately because of its own signification, and this idea was preserved by many Parisian logicians.

Thirteenth-century terminist textbooks already include a detailed and clearly developed doctrine of supposition. In Paris during the second half of the century, this tradition had to give way to the modistic influence, but it survived largely undisputed in Oxford. Subsequently, the ideology of mental terms made it again generally accepted in the beginning of the fourteenth century. After this it became part of the permanent apparatus of late medieval logic.

"To supposit" is obviously a technical term; it means something like "to stand for," and this indeed was an alternative expression. Early terminists like Sherwood thought that *supposition* belongs only to substantives that are posited as subjects (i.e., subposited under predicates), whereas the denotative function of verbs and adjectives is *copulation*. Soon, however, it became the rule to merge these cases and connect supposition to every categorematic term. The definition of this general supposition is not evident. Perhaps it is clearest simply to quote concise definitions from two authors: "When a term stands for something in a proposition in such a way that we use the term for the thing and the term (or its nominative case, if it is in an oblique case) is truly predicated of the thing (or a pronoun referring to the thing), the term supposits for that thing" (Ockham). "All and only those terms supposit which, when something is pointed out by the pronoun 'this' or several things by the pronoun 'these', can truly be affirmed of that pronoun" (Buridan).

We shall try to sketch an overview of the divisions of supposition. First of all, in some cases the supposition is "improper" because the word is used in a nonliteral or metaphoric way; let us concentrate on "proper supposition" only. The definition of its various types displays both semantic and syntactic factors. It seems that the *suppositum* of a word can be of three fundamentally different semantic kinds, and the supposition is accordingly called either *material*, *simple*, or *personal*.

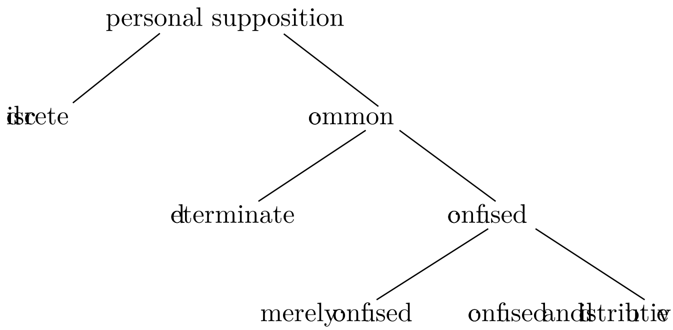


A term has *material* supposition (*suppositio materialis*) when it stands for itself. It must be kept in mind that people in the Middle Ages did not use quotation marks, and material supposition is an alternative way to cope with some problems of use and mention. Sherwood notes that material supposition can be of two types: The word supposits itself either as a sheer utterance or as something significant. His examples are "man" in "Man is monosyllabic" and "Man is a noun."

The supposition is *simple* (*simplex*) when a word stands for a concept. The classical elementary example is "Man is a species." To realists, the *suppositum* then should be equated with some extramental conceptual *significatum*. "If 'man is a species' is true, the term 'man' supposits its *significatum*.... The word 'man' does not primarily signify anything singular; thus it signifies primarily something general, and this is a species" (Walter Burley). According to nominalists, the simple *suppositum* is a mental entity, such as an intention.

In the most common case, the word supposits some things that it signifies. For historical reasons, this was called by the surprising name of *personal* supposition (*suppositio personalis*). Because both simple and personal supposition are related to the meaning, unlike material supposition, they were often together called *formal* supposition. On the other hand, nominalists liked to reduce concepts to mental words, so in a sense Buridan and Peter of Ailly are more straightforward than Ockham when they do not admit simple supposition as an independent class, counting it as material.

The main task is the classification of personal supposition, and here syntactic matters interfere. Let us start by providing the next diagram, representing the early state of the classification, and then proceed to explanations of its titles.



This scheme was in fact given by William of Sherwood, except that he makes the difference between common and discrete supposition in another context.

Discrete supposition (*suppositio discreta*) belongs to discrete terms: that is, to proper names and demonstrative expressions, like "this man" or "this." Then the *suppositum* is the unique object that is signified. All other terms have *suppositio personalis communis.*

This *common* supposition is further divided into determinate and confused kinds. The supposition is *determinate* (*determinata*) when it allows instantiation, as we might say. But medieval logicians had no such notion. Early authors thought that a determinately suppositing term stands for one determinate object. Ockham improved on this, saying instead that determinate supposition supports *descent* to singulars, that is, to sentences that are got by substituting singular terms in place of a general term. Thus, in "A man is running" the term "man" supposits determinately because we can legitimately infer that "This man is running or that man is running or. . . ," and each member of this disjunction in its turn allows *ascent* back to the original sentence.

The supposition is *merely confused* (*confusa tantum*) if the proposition does not allow instantiation but is instead implied by its particular instances. As Ockham puts it, "in the proposition 'Every man is an animal', the word 'animal' has merely confused supposition; for one cannot descend to the particulars under 'animal' by way of a disjunctive proposition. The following is not a good inference: every man is an animal, therefore, every man is this animal or every man is that animal or every man is. . ." But it is also worth noticing that "Every man is this animal or that animal or that. . ." does indeed follow.

Finally, the word has *confused and distributive* supposition (*confusa distributiva*) if it allows descent to all singulars but does not support any ascent. Here the reference concerns "distributively" each and every one of the individuals. For example, in "Every man is an animal" the term "man" has confused distributive supposition. It is correct to infer "This man is an animal and this man is an animal and…"; or as we write nowadays, "The man a is an animal and the man b is an animal and…" On the other hand, none of these singular propositions implies the original sentence.

For confused supposition—and especially for the descent to singulars—it is important to decide what the adequate class of individuals ought to be. There was some debate on this point about the correct formulation until it was agreed that the terms had to be duly *ampliated*, in other words, extended from the basic case of all individuals presently belonging under the concept, such as all actual men, to include all past and future men as well, and in later logic even all possible instances (all possible men). So terms could be examined either with their actual supposition or with an extended supposition.

It is not obvious what the motive behind supposition theory really was. Early authors possibly just wanted to capture various kinds of referring. But when Ockham and his followers started to build a more complex theory, with rules of descent and ascent, they probably did pursue something else. Thus, the supposition theory has been compared to the modern framework of quantification theory, and clearly it has something to do with the problems of multiple quantification and scope—problems that had no explicit place in Aristotle's logic. Also, it can be seen as an attempt to warrant certain inference types, like those of descent and ascent. The interpretation here is still a matter of controversy.

4. Proposition

The core of the medieval theory of judgment centers around the standard definition of proposition (*propositio*), deriving from the late ancient period through Boethius. The definition runs as follows:

> A proposition is an expression that signifies something true or false.
> *Propositio est oratio verum falsumve significans.*

This definition accords with the classical theory of definition. It consists of the generic part (expression) and the distinguishing characteristic (signifying something true or false). For our purposes, however, it seems more useful to divide it into three parts and look at the concepts of truth and falsity separately from the problem of what it is that the proposition exactly signifies.

4.1. Propositions Are Expressions

As we have already seen, medieval authors understood logic as a discipline whose subject matter is linguistic discourse. It is well in line with this general approach that they also thought of the propositions studied in logic as sentences actually uttered in some language, typically either spoken or written. As we saw in section 2, a central issue in the determination of the subject matter of logic was whether (and in what sense) we could distinguish a special class of mental propositions. Thus, thoughts can be propositional only in so far as they have linguistic structure. A proposition, as the medievals thought of it, is something that is put forward as a sentence, and thus it has actual existence in time and typically also in space. As we will soon see in more detail, it was not the case that medieval logicians would have failed to make a distinction between the actual utterance and that which it expresses. Rather, they simply thought of propositional truth as an issue that comes up in connection with claims actually put forward, not as a property of abstract entities.

From the viewpoint of twentieth-century logicians, this feature of medieval conceptual practice has some implications which are worth pointing out, although they are ultimately superficial. According to medieval parlance, a proposition has to exist (i.e., has to be actually put forward in some language) to have a truth value, and it has its truth value in respect to some specific instant and context. Thus, a proposition like "there are no negative propositions" cannot be true, since it falsifies itself, though it is clear that the case could be as it claims. Also, the same proposition can have different truth values in different situations. The truth value of "Socrates is seated" varies when Socrates either stands up or sits down. Furthermore, the truth of "this is a donkey" varies depending on what the demonstrative pronoun refers to. Indeed, all logical properties that a proposition has presuppose that it exists; thus medieval logicians often pointed out that their study applies to propositions, not eternally, but on all occasions in which they are put forward.

4.2. Propositions Carry Truth Values

Not all significative expressions are propositions. Boethius's textbook distinguishes between "perfect" and "imperfect" expressions with the idea that an imperfect expression does not make complete sense but the hearer expects something more. More important, Boethius continues by listing questions, imperatives, requests, and addresses in addition to indicative sentences that make an assertion and count as propositions. This listing of the kinds of expressions is based on grammatical categories, and similar strategies were also followed in subsequent discussions. It may be of some interest to note that Buridan, for example, takes it to be worth an argument to reject Peter of Spain's claim that sentences in the subjunctive mood (like "if you were to come to me, I would give you a horse") do not count as propositions.

It seems that medieval logicians disagreed on whether a proposition that is just mentioned without being asserted carries a truth value. The distinction between apprehensive and judicative uses of a propositional complex was rather standard. Ockham, for example, argues that it concerns propositions so that even an apprehended proposition has a truth value, although no stance is taken to it in the apprehension. Judgments, as he sees it, take stances on truth values, but propositions have them by themselves. Buridan, for his part, seems to rely on similar considerations to show that the sentential complex at issue is not a proposition. He seems to have thought that sentences are able to carry truth values through being asserted. We will return to this issue in connection with molecular propositions.

For the most part, medieval logicians accepted the laws of noncontradiction and the excluded middle. Thus, every proposition has one and only one truth value. But neither of the two principles remained unchallenged. Aristotle's famous sea battle in *De interpretatione* chapter 9 was widely discussed and within that debate it was also suggested that contingent propositions about the future perhaps do not yet have a truth value. This did not become the standard view. Similarly, in the widespread discussions concerning limit-decision problems and particularly the instant of change, it was suggested that perhaps contradictories are both true at the instant of change. Instead of accepting this, the standard line was to provide elaborate analyses of limit decision relying on mathematical considerations concerning infinitesimal magnitudes.

As is well known, in the more philosophical discussions concerning the nature of truth medieval logicians often put forward the principle of correspondence: Truth is *adaequatio rei et intellectus*. This definition was not, however, much used in the specific context of logic. There the term "truth" was mostly used with the more limited meaning of propositional truth, and it proved difficult to exemplify from the real world anything that corresponded to a propositional complex. Thus, truth could hardly be explained as a relation between a real thing and a proposition. In the Aristotelian approach, things are referred to by using simple terms, and no simple expression—a mere term—can have a truth value. Truth rather arises from "composition" or "division" of terms in a predication, and depends on how this composition or division accords with how things really are. In his *Syncategoreumata*, Peter of Spain gives an elaborate suggestion that there is some kind of real composition, typically explicable with reference to the way in which everything in the world is composed of matter and form. According to Peter's suggestion, the truth of a sentence depends on whether this "real composition" is expressed adequately.

The standard Aristotelian *dictum*, "it is because the actual thing exists or does not that the statement is called true or false" (*Cat.* 12; 14b21–22), was not always understood in this manner. A typical way of explicating the claim that a proposition is true was to say that it "signifies as it is" (*significat sicut est*) or something to the same effect. By such formulas logicians tried to avoid committing themselves to positing any real entity with which the true proposition would have direct correspondence. Instead, the expression often worked in a way analogous to what has lately been called "disquotational": allowing transformation of the claim "p is true" into the simple claim "p."

Ockham's *Summa logicae* (I, 43) contains an interesting discussion of in what sense truth is predicated of a proposition. In his opinion, it is not a real quality of the proposition. This can be proved by the fact that a proposition may change from truth to falsity by fully external change. For example, when something ceases to move, the truth value of the proposition "this thing is moving" changes without the proposition itself changing. According to Ockham's explanation, "true" is a connotative term signifying that things are as the proposition signifies. However, this remark leaves open the issue of what it is for things to be as the proposition signifies.

4.3. Are There Any Propositional Significates?

Stoic logic, and in its wake important early medieval authors like Boethius and Peter Abelard, made a distinction between a declarative sentence and its *dictum*, or that which "is said." Thus, the *dictum* expresses, or it simply *is*, the content of the proposition without being itself a proposition. For example, the proposition "Socrates is seated" (*Socrates sedet*) says or puts forward the *dictum* "that Socrates is seated," which in Latin is typically expressed as an accusative plus infinitive construction (*Socratem sedere*). Over the centuries, many logicians discussed the status of the *dictum*. Also, the related distinction between a proposition (as an expression) and its total significate (in distinction from the separate significates of its constituents) became a topic of an interesting dispute toward the second quarter of the fourteenth century.

In his early work, *Commentary on the Sentences*, Ockham puts forward a theory according to which belief always concerns a proposition formulated in mental language. That is, when a person assents to something, he has to formulate a mental proposition expressing that which he assents to. He then reflexively apprehends the proposition as a whole and assents to it. It seems that Ockham's motivation for this theory was the view that there is no way to grasp propositional content apart from formulating a proposition in mental language. Thus, if objects of beliefs are true or false, they must be formulated in mental language.

Several contemporaries of Ockham did not straightforwardly accept the idea that the object of belief must always be an actually formulated proposition. Even Ockham himself shows some hesitation toward this theory in his *Quodlibetal questions*, which he composed later. It seemed to many authors that when one believes, for example, that God exists or that a man is running, the object of belief is somehow out in the world and not merely a proposition in the mind. The idea is that people do not always believe in sentences, but at least sometimes it should rather be said that they believe things to exist in a certain way. This consideration made medieval logicians search for something like propositional content outside the mind and a number of different theories of how it could be found emerged.

In Walter Chatton's theory, the object of the assent has to be some extramental thing. If you believe that a man is running, the object of your belief is the man at issue. Thus, the significate of the proposition "a man is running" is the man. Chatton recognized that his theory has the problematic consequence that the simple term "a man" and the propositional complex "a man is running" signify the same thing. As Chatton saw it, the difference in these two expressions is not in what they signify but in how they signify it.

The terminology he used in this connection refers back to modist grammatical theories.

It seems that both Ockham and a younger contemporary, Adam Wodeham, reacted against Chatton's theory. In his *Quodlibetal disputations*, Ockham makes a further distinction concerning propositional assent, in effect allowing it to be the case that you give assent without reflexively considering a mental proposition. In such a case, you simply form the proposition and give your assent in an unreflective way as connected to rather than directed at the mental proposition. As Ockham curiously points out, this kind of assent is not at issue in scientific knowledge, only in beliefs of ordinary life. According to Ockham's obscure remarks, nothing is the object of this kind of assent.

Wodeham seems to have continued from this basis in his theorizing. He wanted, though, to allow that even the nonreflective kind of assent is about something, and the significate of the proposition appeared to be a suitable candidate for an object. However, its metaphysical status seemed quite unclear to the medieval mind. According to Wodeham, the significates of propositions need to be categorically different from the significates of the terms. As he put it, propositions do, of course, signify the things signified by their terms, but no thing or combination of things is the adequate total significate of the propositional complex. The adequate significates of propositions are such that they can only be signified by propositions; even further, they do not belong to any of the Aristotelian categories nor can they be called things.

Wodeham's theory became known as a theory endorsing "complex signifiables" (*complexa significabilia*). Such entities were rejected by most subsequent logicians, including major figures like John Buridan, but accepted by some, most famously by Gregory of Rimini—in subsequent discussion, the theory became known as his theory. In the third quarter of the fourteenth century, discussion of what propositions signify was abundant. Is it something like a mode of being? Or just a mental act of composition? Do propositions in fact signify anything more than just the things denoted by the terms, or perhaps even just the thing denoted by the subject?

The fourteenth-century discussion concerning complex signifiables seems to have made it clear to late medieval logicians that their logic was based on a metaphysical view of the world as consisting of things and not of states of affairs. The constituents of the world could be referred to by terms, but to make claims about the world, a different kind of mental act was needed. Paradigmatically, one had to construct a complex expression asserting a composition of multiple entities.

4.4. Predication

In Aristotelian logic, the ground for all judgments is laid by the predicative structure, where two terms are either joined or disjoined as the subject and the predicate. After Boethius, it remained customary in the Middle Ages to treat affirmative predication and negative predication as two different kinds of statement, and also to take negation simply as “destroying the force of the affirmation.” Thus, it is not necessary here to treat negative predication distinctly from the basic affirmative case.

The affirmative predication consists, as already Boethius recognized, not only of the two terms but also of the copula. Thus, when Aristotle remarks that a predication can be constructed either with a verb (e.g., “a man runs,” *homo currit*) or with a participle (e.g., “a man is running,” *homo currens est*), this was normally interpreted as meaning that the latter form is to be taken as primary. In the latter, the copula “is” was said to be added as a third part (*tertium adiacens*). In Latin, the copula was of course the standard verb “to be” (*esse*), which was also used in the simple existential claim “a man exists” (*homo est*). This use of *est* as *secundum adiacens* had to be explained since it appeared to lack either the copula or the predicate. As Boethius saw it, the verb serves here a double role. This solution was accepted in the Middle Ages, and thus there was no need to see it as an altogether different kind of statement. Buridan even argued against ordinary linguistic practice that logically one should prefer the formulation “a man is a being” (*homo est ens*).

Given that the copula joins the two terms into a predicative proposition and gives the sentence its assertive character, it still remains unclear exactly how it joins the terms together. It seems that this was one of the most fundamental points of disagreement among medieval logicians. For modern scholars it has proved rather difficult to find a satisfactory description of how the simple predication was understood in the Middle Ages.

One crucial nontrivial issue seemed clear, though. Throughout the Middle Ages, it was commonly assumed that in the absence of specific contrary reasons, the verb “to be” even as the copula retains its signification of being. Thus, all affirmative predications carry some kind of existential force, while negative predications do not. In an affirmation, something is affirmed to exist; a negation contains no such affirmation of existence. But beyond this simple issue, interpretations of the nature of predication seem to diverge widely.

Most of the twentieth century discussions of the exact content of the different medieval theories of predication have been based on the Fregean distinction between the different senses of “to be.” Scholars have distinguished between inherence theories and identity theories of predication, despite the evident threat of anachronism in such a strategy. For want of a better strategy, we also have to rely on that distinction here. But instead of trying to classify authors into these two classes, let us simply look at the motivations behind these two rather different ways of accounting for what happens in a predication.

The idea of the inherence theory is that the subject and the predicate have crucially different functions in the predication. While the function of the subject is to signify or pick out that which is spoken of, the function of the predicate is to express what is being claimed of that thing. The idea is, then, that the Aristotelian form signified by the predicate inheres in the thing signified by the subject. Peter of Spain seems to defend this kind of theory of predication when he tries to show that the copula signifies that relation of inherence obtaining between matter and form, or between a subject and its accident. Aquinas seems to follow this account.

Scholars have disagreed about Abelard's theory, and it indeed seems that his rich discussion of the topic provided grounds for several kinds of different subsequent theories. On the one hand, he seems to lay the basis for the inherence theory. On the other, he defends the idea that to look at the exact truth conditions of a predication like "a man is white" (*homo albus est*), it should be analyzed into a fuller form "that which is a man is that which is white" (*idem quod est homo est id quod album est*; *Logica ingredientibus* 60.13). With such a formulation he seems to have in mind the idea that for the affirmative predication to be true, the subject and the predicate must refer to the same things. This is commonly called the identity theory of predication.

Abelard's "that which is" (*quod est*) formulation remained part of the actual practice of logical writing for several centuries. It can be found commonly from logical texts throughout the Middle Ages, although it was not always offered as an explanation of the truth conditions of predication in general. The formulation has the special feature that it appears to give the subject and the predicate of a predication a similar reading. Both are to be understood as referring to some thing, and then the assertion put forward in the proposition would be the identity of these two things. This seems to amount to the identity theory of predication in Fregean terms.

In the fourteenth century, both Ockham and Buridan seem to have quite straightforwardly defended the idea that the Aristotelian syllogistic is based on identity predications. As they put it, the simple predication "A is B" is true if and only if A and B supposit for the same thing. For the most part, truth conditions of different kinds of propositions can be derived from this principle.

Somewhat interestingly, Ockham nevertheless recognizes the need of basic propositions expressing relations of inherence. For Ockham, the predicate "white" is a so-called connotative term, and therefore a somewhat special case. According to his analysis, the predication "Socrates is white" (*Sortes est albus*) should be analyzed into "Socrates exists and whiteness is in Socrates" (*Sortes est et Sorti inest albedo*). In his metaphysical picture, Ockham allows both substances and qualities to be real things, and if one is allowed to use only so-called absolute terms that supposit in a sentence only those things which they signify, the relation of inherence (*inesse*) is not expressible with an identity predication. Qualities inhere in substances, but they are not identical with substances. The whiteness at issue in the claim "Socrates is white" is not Socrates, it is a quality inhering in Socrates. Socrates is not whiteness even if he is white.

In his *Summa logicae*, Ockham has some special chapters on propositions involving terms in oblique cases (in cases other than the nominative). The just-mentioned proposition "whiteness is in Socrates" is a paradigm case of such a proposition (in Latin, the subject has to be in the dative case *Sorti*; in English, the effect of the case is represented with the preposition "in"). Furthermore, all propositions involving the terms that Ockham calls "connotative" require in their logical analysis that oblique cases are used. The main claim of the short chapters of *Summa logicae* addressing propositions containing such terms is that their truth conditions cannot be given by the simple rule of thumb that the subject and the predicate must supposit for the same thing in an affirmative sentence. Consequently, the rules for syllogisms formulated with such propositions are also abnormal. In effect, Ockham excludes propositions with oblique terms from the ordinary syllogistic system, thus leaving a surprising gap in his logical system.

In his logic, Buridan proceeds differently. For the purposes of the syllogistic system, he requires that all propositions should be analyzed into a form where truth conditions can be given through variations of the rule that in affirmative sentences the subject and the predicate supposit for the same thing. This allows him to apply the standard syllogistic system to all propositions. The solution is at the price of greater semantic complexity. Buridan has to allow so-called connotative terms (including, e.g., many quality terms like "white") as logically simple terms despite their semantic complexity.

Both Ockham and Buridan apparently thought that identity predication is the logically privileged kind of predication. Nevertheless, they also both accepted the Aristotelian substance-accident ontology to such an extent that they had to find ways of expressing the special relation of inherence. While Ockham allowed exceptions to the syllogistic through irreducible propositions expressing inherence, Buridan opted for a syllogistic system with obviously complex terms expressing inherential structures.

4.5. Negations

As the medieval logicians saw it, the simple predication "A is B" contains altogether four different places where a negation can be posited:

1. It is not the case that A is B.
2. A is not B.
3. Not-A is B.
4. A is not-B.

It is of course clear that 1 is closest to the negation used in twentieth-century logic. In it, the negation is taken to deny the whole proposition. According to Boethius's commonly accepted formulation, the force (*vis*) of the predication is in the copula, and hence denying the copula denies the whole proposition. Thus, the negation in 2 has the same effect as in 1. (As Buridan notes, for quantifiers and other modifiers, the location of the negation may still make a difference.)

2 is the standard negation of medieval logic. It is the direct contradictory of the corresponding affirmative predication. In particular, it is noteworthy that this negation does not carry any existential presuppositions. Thus, "a chimera is not an animal" is true simply because no chimeras exist.

3 and 4 are affirmative statements containing an infinite term, as terms of the type "not-A" were called. In these cases, the negation is connected directly to a term and not to a proposition. An infinite term was taken to refer to those things to which the term itself does not refer. Thus, not-man refers to anything that is not a man. Because these negations do not make the proposition negative, 3 and 4 carry existential content: Some B must exist for 3 to be true, and some A for 4 to be true.

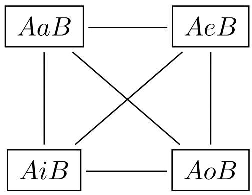
Although the syntactic idea of attaching a negation to a term was universally accepted in the Middle Ages, logicians seem to have disagreed about whether the term-negation should be taken to be essentially the same negation as the propositional one but in a different use. In his *Syncategoreumata*, Peter of Spain seems to reject this idea. He presents the two negations as genuinely different in themselves. His discussion is connected to a theory where even simple names and verbs signify in a composite sense. Thus, the idea is that "man," for example, means a composition of a substance with a quality, a substance having the quality of being human. Thus, the infinite term "not-man" signifies a substance that has not entered into a composition with the quality of being human. Ockham, for his part, preferred to reduce negating a term to ordinary propositional negation, claiming that the meaning of "not-man" can be explained as "something which is not a man." Buridan allows infinite terms a significant role in his syllogistic system, and thus seems to go back to thinking that the negation involved in them is fundamentally distinct from that which he calls "negating negation"—that is, the propositional negation that has power over the copula.

Given that negations can be put in many places even in a simple predication, medieval logicians gained skill in handling combinations of different negations. The idea that two negations cancel each other (provided that they are of the same type and scope) was also well known.

4.6. Quantifiers

Aristotelian predications typically have so-called quantity. Medieval logic commonly distinguished between universal ("every A is B"), particular ("some A is B") and indefinite propositions ("A is B"). As Boethius already pointed out, the indefinite predication that lacks any quantifier is equivalent to the particular one. Some logicians did specify certain uses that violate this rule of thumb, but such exceptions are rare. In addition to quantified and indefinite predications, singular predications were also discussed (e.g., "Socrates is running"). They had a subject term that was a proper name or some suitable demonstrative pronoun.

Basic quantified predications were given vowel symbols as mnemonic labels from the first two vowels of the Latin verbs "affirm" (*affirmo*) and "deny" (*nego*). Thus, the universal affirmative was shortened as AaB, where A is the subject and B the predicate. Similarly, the particular affirmative was AiB, the universal negative AeB and the particular negative AoB. These four predications were further organized into the so-called square of opposition to show their interrelations.



The upper two, the universal affirmative and the universal negative, were called *contraries*; they cannot be true simultaneously, but could both turn out to be false. Similarly, the particular affirmative and particular negative were *subcontraries*; they cannot be false simultaneously, but both could turn out true. The relation between the universal and the particular was called *subalternation* on both sides; the particular follows from the universal but not vice versa. The propositions in the opposite corners were called *contradictories*, since one of them had to be true and the other false.

In the Middle Ages, a substantial amount of ink was used discussing whether a universal affirmative could be true when only one thing of the relevant kind exists. The paradigm example was "every phoenix exists," and many logicians rejected it with the requirement that there must be at least three individuals to justify the use of "every." Toward the end of the thirteenth century this discussion seems to disappear, apparently in favor of the view that one referent is enough; the existential presupposition was never dropped, however.

Another issue of detail that was also widely discussed later was the case of universal predications of natural sciences, which capture some invariable that does not appear to be dependent on the actual existence of the individuals at issue. A suitable example is "every eclipse of the moon is caused by the shadow of the Earth." According to a strict interpretation of the existential presupposition, such predications prove false most of the time—which seems somewhat inconvenient. Two fundamentally different suggestions for a better reading of the predication were put forward. Ockham seems to favor the idea that what really is at issue here is the conditional proposition "if the moon is eclipsed, the eclipse is caused by the shadow of the Earth." This solution draws on the traditionally recognized idea that the conditional is implied by the universal affirmation. However, Buridan opted for another solution. As he reads the universal affirmation at issue, its verb should be read in a nontemporal sense. In such a reading, past and future eclipses also provide instances satisfying the diluted existential presupposition.

4.7. Complex Terms

Medieval logicians allowed that not all terms of a predication are simple. A predication can, of course, always be divided into the subject and the predicate together with the copula and the appropriate quantitative, qualitative, and modal modifiers. But the two terms may possibly be further analyzable (e.g., "a just man is talking," where the subject "just man" consists of two parts), and indeed this was a topic that attracted much attention during the Middle Ages. Toward the middle of the fourteenth century, discussion on this topic resulted in a detailed theory on the interaction between different kinds of combinations of categorematic and syncategorematic elements that can be found in a predication. To tackle with issues of scope an elaborate system of word order rules was introduced for the technical Latin used by logicians.

It seems that thirteenth-century logicians did not take it to be a serious problem that complex predications do not behave in ways that suit the needs of syllogistics. Following Aristotle's remark (*Analytica priora* I, 36; 48b41–49a5), syllogisms with oblique terms in the various cases were usually discussed separately, and thus it seems that the thirteenth-century logicians probably thought that more complicated predications do not necessarily fit into the ordinary syllogism. As we already noted, Ockham makes this slight inconvenience clear in his *Summa logicae*. It seems that Ockham fully understood that the traditional syllogistic logic does not always work if actually used linguistic structures are given full logical analyses. Also, he explicitly allows that there is no general way of giving the truth conditions of the various kinds of complex predications; in particular, he points out that even as simple a construction as the genitive case makes the standard truth conditions of identity predications inapplicable. "The donkey is Socrates's" is an affirmative predication. However, its truth requires, but it is not sufficient for it, that the subject and predicate supposit for different things ("donkey" for a domestic animal owned by a person, and "Socrates" for the owner of the animal). More generally, Ockham thought that mere identity predications are not sufficient to explain the expressive power of the actually used language. A richer variety of propositions had to be accounted for, but in fact they found no place in syllogistic logic. Thus, syllogistic logic was not a complete system covering all valid inferences.

After Ockham, Buridan took another approach. As he saw it, all categorical propositions can be reformulated as straightforward Aristotelian predications fitting the needs of the ordinary syllogism and having the rule of identity or nonidentity of supposition as the criterion of truth. For this purpose, he had to modify the traditional systems of combining different categorematic and syncategorematic elements so that they appear as geared toward building up terms whose suppositions can be decided. Perhaps most important, he saw that he could not assume that standard Aristotelian predications would be found as the end results of logicolinguistic analysis. Rather, he understood the building blocks of the syllogistic system—identity predications—to be more or less artificial constructions built from complex terms. For example, for the purposes of syllogistic logic the sentence "the donkey is Socrates's" must be read as "the donkey is Socrates's thing," although the predicate "Socrates's thing" clearly is not a simple term of the ideal mental language.

Buridan did not assume that all mental or spoken propositions would be identity predications. Rather, he assumed that for the purposes of syllogistic logic, any proposition could be transformed into an equivalent identity predication. By such means, syllogistic logic could serve as a complete system containing all inferences.

Buridan's strategy involved, therefore, a massive expansion of the syllogistic system toward incorporating increasingly complex terms. Whereas logicians up to Ockham had accepted that a wide variety of propositions are nonstandard from the viewpoint of syllogistic logic, Buridan builds rules on how the content of these nonstandard propositions can be expressed by standard structures involving very complex terms. Buridan provides elaborate rules concerning complex terms. The idea is to show how nouns and verbs interact with different syncategorematic expressions and produce terms that fit into standard Aristotelian predications. In Buridan's view, all propositions can be transformed so that the truth conditions can be expressed through the criteria of an identity predication. In affirmative sentences, the terms must supposit for the same thing, while in negative sentences, they must not supposit for the same thing.

To see the full strength of Buridan's new system, let us consider a somewhat more complicated example. Buridan analyzes "Each man's donkey runs" (*cuiuslibet hominis asinus currit*) in a new way. Traditionally, this Aristotelian sentence was understood as a universal affirmation consisting of the subject "man" in the genitive case, and a complex predicate. This analysis makes the subject supposit for men, and the predicate for running donkeys so that the assertion cannot be read as an identity predication. Thus, standard syllogistics are not applicable to a proposition like this. Most logicians up to and including Ockham seem to have been satisfied with the implied limitations of the syllogistic system. Buridan, however, analyzes the proposition as an indefinite affirmation. It has a complex subject "each man's donkey," which includes two simple categorematic terms ("man," "donkey"), a marker for the genitive case (the genitive ending "'s"), and a quantifier (the universal sign "each"). The quantifier does not make the proposition universal, because it has only a part of the subject in its scope and must therefore be understood as internal to the subject term. As a whole, the subject supposits for sets of donkeys such that each man owns at least one of the donkeys in the set. The predicate of the proposition is a simple term, "running." It supposits for sets of running things. Construed in this way, the predication can be evaluated with the standard criteria of truth, and standard syllogistics can be applied to it.

It seems clear that Buridan took seriously the programmatic idea that the Aristotelian syllogistic system should provide a universal logical tool which did not allow major exceptions to behave in nonstandard ways. But instead of analyzing complex propositions into combinations of predications with simple terms, Buridan provides elaborate rules concerning the ways in which complex terms are built.

4.8. Hypothetical Propositions

In the Middle Ages, not only conditionals but also conjunctions and disjunctions were called hypothetical (*hypothetica*) propositions. Otherwise the treatment of conditionals and disjunctions causes no surprises to a modern reader familiar with basic propositional logic. Walter Burley, for example, gives the following account. Conjunctions are propositions consisting of two further propositions that are joined with the conjunction "and" or something equivalent. Their truth conditions require that the propositions thus joined are true. Negating a conjunction makes reference to another type of hypothetical, namely disjunction, because denial of a conjunction requires only that one or the other of the conjuncts is denied. Disjunction, for its part, is defined in the inclusive manner: Its truth conditions require that one of the parts is or both of them are true. Denial of a disjunction produces a conjunction, and as Burley notes, denial of a disjunction of contradictories (e.g., "Socrates runs or Socrates does not run") produces a conjunction which includes contradictories.

Certain interesting issues are raised in more detailed discussions of conjunctive and disjunctive propositions. One such is the nature and exact content of conjunctive and disjunctive terms used in propositions like "every man runs or walks." Are they reducible to conjunctive and disjunctive propositions and why not exactly? How ought they be accounted for in inferential connections? Another, more philosophical issue was the question of whether the parts of conjunctions and disjunctions are strictly speaking propositions. As Buridan notes, the "force of the proposition" (*vis propositionis*) in a disjunction is in the connective, and thus not in either of the disjuncts. Hence, it is only the whole and not the parts that carry truth value in the composition. When someone utters a disjunctive proposition consisting of contradictories, he does not, according to Buridan, say anything false, although one of the parts would be false if uttered as a proposition. Thus, hypothetical propositions do not, strictly speaking, consist of categorical propositions but of linguistic structures exactly like categorical propositions.

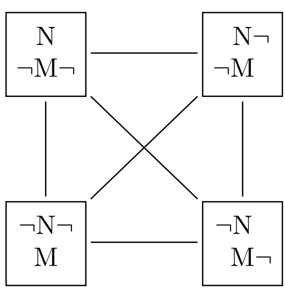
It seems that medieval logicians treated conjunctions and disjunctions in a straightforwardly truth-functional manner. It seems equally clear that their treatment of conditionals differs from the twentieth-century theory of material implication. Indeed, in the Middle Ages theory of conditionals was mainly developed in connection with a general theory of inference, under the label "consequences." Conditionals were taken to express claims concerning relationships of inferential type.

Medieval logicians also distinguished further types of hypothetical propositions. In Buridan's discussion (1.7), altogether six kinds of hypothetical propositions are accounted for, including conditional, conjunctive, disjunctive, causal ("because the sun shines above the Earth, it is daytime"), temporal ("Socrates runs when Plato disputes"), and local ones ("Socrates runs to where Plato disputes"). Buridan even vaguely suggests that perhaps other Aristotelian categories may also give rise to hypothetical propositions in a way similar to temporal and local hypotheticals. It is clear that this approach to hypothetical propositions relies on other ways of combining propositions than just the truth-functional ones. The connective may express something more than just a truth function.

4.9. Modal Operators

Logical issues connected to possibility and necessity, which in the twentieth century have been studied as alethic modal logic, were a central research topic in late medieval logic. These modal terms were usually discussed together with other modifiers operating in similar syntactic roles. For example, the twentieth-century fields of study known as deontic logic (dealing with permissibility and obligation) and epistemic logic (dealing with concepts of knowledge and belief) have their counterparts in the Middle Ages, where these issues were discussed together with possibility and necessity.

Most medieval logicians discussed altogether four modal operators crucial to modern alethic modal logic: possibility, impossibility, contingency, and necessity. These were defined in relation to each other so that the necessary was usually taken to be possible but not contingent, whereas the impossible was taken to be neither possible nor contingent. Like the square of opposition of simple predications, modal predications were often organized into a square of modal opposition following Aristotle's presentation in *De interpretatione* (ch. 13). It is particularly noteworthy that following Aristotle's model, the square of modal opposition typically contained just the modal operators, not complete sentences. In a somewhat schematized way, the basic square can be illustrated as follows:



In this square, the relations of contrariety, subcontrariety, subalternation, and contradiction were said to behave as they would in the basic square of simple predications.

Following Aristotle, medieval logicians made a distinction between two ways of understanding a modal predication to make sense of examples like the possibility that someone sitting walks. Understood *de dicto*, there is no such possibility. The sentence “someone sitting walks” is impossible. But understood *de re*, there is such a possibility, since the person who is sitting may be able to walk. Thus, if the modal predicate “can walk” is understood *de re*, or as concerning the person who is actually sitting now, the sentence might be true. All subsequent major logicians discussed this distinction in some form or other.

In the sections concerning modal propositions in Ockham’s *Summa logicae*, it is clear that the *de dicto* reading is given logical priority. Using another traditional terminology, Ockham prefers to call it the composite sense (*sensu composito*) and does not oppose it to a *de re* reading but to the roughly similar divided sense (*sensu diviso*). Ockham apparently thinks that modality is a property of propositions rather than terms, and aims at reducing readings *sensu diviso* to *sensu composito* through analyzing modal propositions in *sensu diviso* into propositions *sensu composito*.

There are three main models by which modern scholars have been able to account for the way in which medieval logicians understood what it means to say that something is possible: the statistical model, the potency model, and the consistency model. During the medieval period, the modal concepts used by particular logicians typically ought not to be explained through reference to a single model. Rather, these three different strands of thought have influenced to varying degrees the modal thinking of different medieval logicians.

The basic intuition explained by the statistical model is that all and only those things seem to be possible which sometimes occur. If something never happens, it means that it can’t happen. The potency model, for its part, explains the intuition that whether something is possible depends on whether it can be done. For something to be possible it is required that some agent has the potency to realize it, though it is not required that the thing is actually realized. However, because normally there are no generic potencies that remain eternally unrealized (why should we say that humans can laugh if no one ever did?), this model becomes clearly distinct from the statistical model only when God’s omnipotence is understood to reach wider than just actual reality. God could have created things or even kinds of things which he never did, and these things remain therefore eternally unrealized possibilities.

It seems that throughout the Middle Ages, God’s omnipotence was thought to be limited only by the law of noncontradiction. Contradictions are not real things, and therefore God’s power is not limited, although we can say that he cannot realize contradictions. This consideration seems to have been one of the reasons why logicians in the thirteenth century increasingly used the criterion of consistency to judge claims about possibility. But it seems that the development of syntactic logical techniques also made it natural to demarcate a class of propositions that are impossible in the traditional sense but nevertheless seem to involve no contradiction (e.g., that man is irrational, or that man is not an animal). Especially the traditional technique of laying down a false or even impossible thesis for an obligational disputation (see the following) seems to have encouraged consideration of consistent propositions or sets of propositions that are in some sense impossible. Some authors, like Boethius of Dacia, even use the special expression *compossibilitas* to refer to this kind of concept of consistency as distinct from possibility. As is well known, from Duns Scotus onward, several logicians made this kind of concept of consistency crucial for possibility in general.

The medieval discussion can be characterized as aiming at finding a way to account for these rather distinct intuitions of what it means to say that something is possible. A certain shift in emphasis is visible. Whereas earlier authors pay more attention to the statistical idea at the expense of consistency, later authors tend to neglect or argue against intuitions captured in the statistical model while emphasizing consistency as the criterion of possibility.

5. Classical Forms of Inference

5.1. Syllogisms

We must next turn to the "theory of inference." Ignoring probable inferences for now, we can say that this part of logic tries to describe how some propositions necessarily follow from others, from their premises. The propositions of a certain sequence have such properties that the last one must follow necessarily from its predecessors. An important type of inference is the *syllogism*—the inference on which Aristotle concentrated in his *Prior Analytics*.

The syllogism was the best-known and paradigmatic type of inference throughout the Middle Ages. In the thirteenth century, when logicians studied demonstrative inference, they were almost exclusively concerned with syllogistics; but afterward, when a more general inference theory developed, the policies of various authors differed widely. Thus Ockham still devotes the main part of his inference theory to a detailed analysis of syllogistics, and so does Buridan, whereas Burley regards it as a well-known special case of the more interesting subject of inference in general. The syllogism is probably the most famous item of "traditional" logic, but actually it has a not very dominant place in the works of medieval logicians. (For instance, in the *Logica magna* of Paul of Venice it is the subject of only one of 38 treatises.) However, it is systematically and historically so important that we must discuss it in relatively more detail.

All authors start by presenting or elaborating the highly condensed definitions in the beginning of the *Prior Analytics*. The often-quoted characterization in *An. Pr.* 24b19–20 says: "A syllogism is a discourse (*oratio*) in which, certain things having been supposed, something different results of necessity because these things are so." In a broad sense, any formally valid inference could be called a syllogism. But in the stricter sense, a syllogism has precisely two "things supposed," two premises. There has been quite a lot of discussion on whether Aristotelian syllogisms are better understood as conditionals ("if p and q then r") or as deductive inferences ("p, q; therefore r"). The latter interpretation is perhaps more popular nowadays, and apparently it is also the correct way to see medieval syllogistics, at least in its classical stage. (Obviously the two things have a systematic correspondence, the relation that we nowadays call the deduction theorem, and many medieval authors were fully aware of it.) This means that syllogisms are like natural deductions of present-day logic.

Though medieval syllogistics followed Aristotle closely, there were some formal differences. Thus Aristotle—for special reasons—had formulated his syllogistic propositions as "Y belongs to X," "mortal belongs to man." This manner was never adopted in Latin; the medieval logicians wrote just "X is Y," "man is mortal."

Aristotle himself had brought the theory of nonmodal syllogistics to such perfection that there was little room left for initiative or disagreement. However, medieval texts produced a more systematic form for the theory, obviously aiming at didactic clarity.

A syllogism consists of two premises and one conclusion; the first premise is called major, and the second premise minor. Each proposition has two terms, a subject and a predicate, connected by a copula. But the two premises have a term in common, the so-called medium, and the terms of the conclusion are identical with the two other terms of the premises. Syllogisms can then be classified according to their configuration Subj–Pred into four different *figures* as follows:

	I	II	III	IV
major	*M–B*	*B–M*	*M–B*	*B–M*
minor	*A–M*	*A–M*	*M–A*	*M–A*

Further, each syllogistic proposition belongs to its type *a*, *e*, *i*, or *o* because of its quality and quantity: They are affirmative or negative, universal or particular. *If* we proceed by defining that the conclusion must always have the structure *A–B*, then it is obvious that each figure includes $4^3 = 64$ alternative combinations, and the total number is 256. But this is not exactly the classical method, so let us have an overview of the syllogism as it was usually presented.

Medieval logicians have a full and standard apparatus for syllogistics as early as the first terminist phase. They list the same valid syllogisms, usually in the same order, and also call them by the same names. The textbooks of William of Sherwood and Peter of Spain supply these names, which stem from some unknown earlier source and even the famous mnemonic verse composed on them. The names have three syllables, one for each sentence, containing the logical vowels *a*, *e*, *i*, and *o*. (In the following list of syllogisms, we mention these names that have recurred in all later logic.) The valid syllogisms were known as *moods*.

The *first* figure includes the four famous syllogisms from which Aristotle starts:

every M is B, every A is M, therefore every A is B (*Barbara*)

no M is B, every A is M, therefore no A is B (*Celarent*)

every M is B, some A is M, therefore some A is B (*Darii*)

no M is B, some A is M, therefore some A is not B (*Ferio*)

The *second* figure has four moods:

no B is M, every A is M, therefore no A is B (*Cesare*)

every B is M, no A is M, therefore no A is B (*Camestres*)

no B is M, some A is M, therefore some A is not B (*Festino*)

every B is M, some A is not M, therefore some A is not B (*Baroco*)

Furthermore, the *third* figure contains six moods:

every M is B, every M is A, therefore some A is B (*Darapti*)

no M is B, every M is A, therefore some A is not B (*Felapton*)

some M is B, every M is A, therefore some A is B (*Disamis*)

every M is B, some M is A, therefore some A is B (*Datisi*)

some M is not B, every M is A, therefore some A is not B (*Bocardo*)

no M is B, some M is A, therefore some A is not B (*Ferison*)

After the Renaissance, logicians continue by giving the five moods of the fourth figure: *Bramantip*, *Camenes*, *Dimaris*, *Fesapo*, and *Fresison*. That, however, is not the orthodox Aristotelian way. Aristotle knew inferences like these but did not include a fourth figure in his theory. Instead, he wanted to place these syllogisms into the first figure. Following his remarks, Theophrastus developed a clear account of the matter, and it was well known in the Middle Ages through Boethius. In Theophrastus's account, the major term need not be the predicate in the conclusion, which can also have the inverted order B–A. This gives us the five so-called indirect moods of the first figure:

every M is B, every A is M, therefore some B is A (*Baralipton*)

no M is B, every A is M, therefore no B is A (*Celantes*)

every M is B, some A is M, therefore some B is A (*Dabitis*)

every M is B, no A is M, therefore some B is not A (*Fapesmo*)

some M is B, no A is M, therefore some B is not A (*Frisesomorum*)

This method can replace the fourth figure, though it does introduce a certain unsatisfactory asymmetry.

The problem of the missing figure has caused much scholarly debate that we cannot enter into here. Medieval logicians were quite aware of the problem since they had seen at least Averroes's comments on the fourth figure. Arguments were often given to refute "objections" questioning the sufficiency of three figures. Apparently medieval authors were unanimous in thinking that the fourth figure could be eliminated with the indirect moods of the first figure. They either said that there were only three figures, or more precisely, like Albert of Saxony, that the fourth is superfluous. It is noteworthy that they did not regard the order of premises as essential.

Thus there are 19 valid syllogistic moods. A small addition was obtained by allowing the five "subaltern" moods, which yield a particular conclusion though a universal one would be valid too. For example, *Barbari* instead of *Barbara* leads to "some A is B." This step would not be accepted in modern logic where universal implications have no existential import, and it indicates clearly that medieval syllogistics assumed that every term really had existential reference.

Aristotle had only implicit allusions to singular propositions in syllogisms, and it was a good achievement that medieval logicians constructed a full and systematic theory of singular syllogisms. Ockham was the most active worker here. He emphasizes that the singularity of terms makes no difference for the validity of inference. This amounts to a considerable reinterpretation of the whole notion of a propositional term. Moreover, he gives explicit cases of singular syllogisms in each figure, for example, the third figure "expository syllogisms" like "x is B, x is A, therefore some A is B." (For nominalists like him, the question had special epistemological relevance because of the basic status of truths about individuals.) Some later Ockhamists even drew a dichotomy across the whole syllogistics between expository syllogisms and those with general mediums.

5.2. Theory of Syllogistics

Syllogistics, undoubtedly, is just a small portion of logical inferences, but systematically it is extremely important. The unique thing in classical syllogistics is that it was a formal theory. Its results are not separate truths achieved by trial and error; instead, they are derived in a deductive manner. This had largely been achieved already in the *Prior Analytics* and continued by ancient commentators. Medieval logicians were very interested in this project.

The most important tool here is *conversion*. It is a completely general method that pertains to all propositions of the S–P form, but it finds good use in syllogistic theory. Briefly, in a conversion the subject and the predicate change places, and conversion rules tell when such a transposition is legitimate. The following set of (nonmodal) conversion rules was universally accepted. First, in simple conversion AeB converts with BeA, and AiB converts with BiA. In other words, some A is B if and only if some B is A, and

no A is B if and only if no B is A.

Second, in conversion *per accidens* AaB implies BiA, and AeB implies BoA (this negative one is the only rule that was not in Aristotle):

if every A is B then some B is A, and

if no A is B, then some B is not A.

These are only *per accidens*, because they change the quantity and do not hold in the opposite direction. Third, ever since Boethius even contraposition was taken as a type of conversion. It preserves the quality and quantity but "changes the finite terms into infinite ones." For example, "if every A is B then every non-B is non-A." Fourteenth-century logicians noticed that contraposition need not be valid when any of the terms is empty—an existential assumption is required.

With conversions, some syllogistic moods can be derived from others. The idea is that if certain syllogisms are selected as basic, others can be derived from them by a clever use of fixed methods. Aristotelians called this process "reduction," present-day logicians would call it *proof*. Conversion was the most important method of reduction. The other method was *reductio ad impossibile*: A mood is valid because the negation of the conclusion leads to the negation of a premise. With these methods, all syllogistic moods could be reduced to the direct moods of the first figure—in fact even further, to *Barbara* and *Celarent*. This was basic stuff in all textbooks, and the consonants in the names of moods refer to the methods of reduction. (S: convert simply; P: convert *per accidens*; M: transpose the premises; C: reduce *ad impossibile*.)

These privileged syllogisms are cases of *dici de omni et nullo*, in which the conclusions can be seen as immediate corollaries of simple affirmation or negation. As Buridan explains, "*dici de omni* applies when nothing is taken under the subject of which the predicate is not predicated, as in 'Every man runs'. *Dici de nullo* applies when nothing is taken under the subject of which the predicate is not denied." So direct first figure syllogisms are immediately self-evident, and medieval logicians, like Aristotle, called them "perfect." Others are imperfect in the sense that their validity needs to be shown.

The growth of syllogistic theory naturally leads to the philosophical question of its foundations. Such a problem can arise from two perspectives: One may wonder about the status of syllogistics in the totality of logic, or one may ask if particular syllogistic inferences depend on some other principles.

a. The question about the general status of syllogistics became current when the theory of consequence developed in the beginning of the fourteenth century (see section 6). Aristotle had started from syllogisms and proceeded to a brief discussion of other inferences; now logicians took the opposite direction. In the thirteenth century, some logicians' attitude seems to be that all strict demonstrative logic is syllogistical, but the more people were concerned with logical research, the clearer it became that other inferences are valid, too; and this was then explicated by means of the concept of consequence. However, the relation between syllogistics and consequences is not very clear. Syllogistics is a part of consequence theory, in the sense that one particular type of consequences are "syllogistic consequences." (This is especially clearly said by Buridan, whereas Ockham prefers to keep the titles unconnected.) And syllogisms hold because they are good or solid consequences, in our words, logically *valid* ones. But does syllogistics depend somehow on other parts of the theory? It seems that medieval logicians did not think so. They were aware of the importance of propositional logic—after all, the Stoic heritage had survived—but they did not work in the present-day fashion and build first a propositional calculus, then a predicate logic on it. Burley is an interesting case here: He really starts from the simple consequences of propositional logic. But he had no followers in this respect, and contrary to what has been suggested, even he does not apparently aim at any stratification of logics here.

b. More concretely, one might ask if the validity of a particular syllogistic mood is based on some principles, or if a syllogism involves the use of other logical laws. This problem does not appear in terminist manuals, but it is discussed in the 1240s by Robert Kilwardby. He insists that the necessity of *dici de omni et nullo* is of such a self-evident nature that it cannot be regarded as a genuine inference step. Many logicians agreed with him. Kilwardby even asks if syllogisms presuppose separate inferences of conversion, and argues that it is not so. Suppose that no B is A; just add "every A is A" as the second premise, and you get the converted sentence, "no A is B," by *Cesare*. Similarly in other cases, we see that conversion reduces to syllogism. This idea was not generally accepted, but conversion was occasionally considered so immediate a transformation that it could not be called an inference at all.

Soon, however, an alternative view was articulated. About 1270, Peter of Auvergne refers to *loci*, the governed steps of argumentation theory, and says that "every syllogism holds because of a *locus* from a more extensive whole to its part." Simon of Faversham, Radulphus Brito, and others then developed this thought that a syllogism must involve a "principle of consequence." The conclusion is somehow included in the premises. But the remarks are brief and obscure. In any case, they anticipate the fourteenth-century view of logically necessary consequence relation that is not peculiar to syllogisms.

5.3. Modal Syllogisms

Aristotle devotes a large part of his *Prior Analytics* to modal syllogisms. But unlike nonmodal syllogistics, this area remains in a very unsatisfactory state. The modalities he there studies are necessity, impossibility, and contingency. He wishes to produce a complete set of syllogisms in which some propositions have such modalities; further, he tries to systematize these syllogisms like the nonmodal ones. Here he needs conversion, *reductio*, and a third method, *ekthesis*, based on defining new predicates. Medieval logicians replaced *ekthesis* with a more elegant method of expository syllogisms.

The main problem is that Aristotle's theory looks incoherent. His set of accepted syllogisms might be the outcome if all modal propositions were read *de re* only, as concerning the modal properties of individuals. But then the conversion rules do not hold: Obviously "every A is something necessarily B" does not convert to "some B is something necessarily A." Moreover, his choice of valid syllogisms contains some oddities.

Ancient commentators struggled with these puzzles, and medieval Aristotelians could not avoid them. Peter of Spain's *Summulae* does not really discuss modal logic, but Kilwardby, Lambert, and Albert the Great try to save Aristotle's doctrine. They resort to a very strong interpretation of necessity, proposed by Averroes, which concerns only necessities which hold per se because of essences. Even this technique demands some arbitrary decisions, and in any case it amounts to a severe restriction of modal syllogistics.

A similar approach seems to have continued through the thirteenth century. The first known work that introduces new methods is the commentary by Richard of Campsall, written about 1308. Campsall's own theory is conservative, since he wants to maintain the Aristotelian syllogisms and conversions by means of a strict and somewhat confused *de re* reading. But the novelty is that he makes a systematic distinction between divided and composite readings. It is connected to the idea, initiated by Duns Scotus, of simultaneous alternative states of affairs.

This new semantics of modal notions made possible a new and different approach to modal logic. From this point of view, modal logic was seen to be much wider than the part that Aristotle had developed, and the relations of modes could be systematized in a new way. The basic notions were now necessity and possibility, which could be understood as realization in all and some alternatives respectively. The first exact presentation of the resulting syllogistics was the very thorough account in Ockham's *Summa logicae*. In Paris, the orthodox Aristotelian model survived much longer, but Buridan's *Tractatus de consequentiis* (1335) provides a modern theory, which is almost as full as Ockham's. A third and more concise classical text is in Pseudo-Scotus's commentary on *Prior Analytics* (c. 1340).

The new modal logic gave plenty of room for the notion of contingency, and it caused some disagreement, but for simplicity we bypass this and concentrate on the syllogistics of possibility and necessity. The composite and divided readings of them were strictly distinguished. The composite readings are easier, and accordingly they were less discussed. They were indeed *de dicto* in the sense that strictly speaking they only make a singular nonmodal claim about a *dictum*; for example, "it is necessary that some A is B" is interpreted as "the *dictum* 'some A is B' is necessary." The syllogistic for such propositions follows from the general consequence theory. Ockham and Buridan agree that in every mood, if both premises are prefixed with necessity N, the conclusion is necessary too. On the other hand, a syllogism MMM, with all the three propositions modalized as possible, does not hold because the premises need not be compatible. Ockham also remarks on NMM and MNM.

Much more problematic were divided premises, that is, propositions with genuinely modalized copulas. The main device for dealing with them was ampliation (see section 3.5), which extends the subject term to refer to supposita that occur in alternative nonactual states of affairs; thus "every A is possibly B" will be read "everything which is or possibly is A is possibly B." But ampliation may be blocked by adding *quod est* A, "what (actually) is" A. Now it is striking that ampliation was understood in two different ways. Ockham assumed that ampliation is good for possibilities (and contingencies)—but he did not accept it for necessities. In other words, only actual things could be said to have necessary properties. The reasons for this are not clear; perhaps he thought that necessities always involve some existence postulate. Buridan, in his turn, said clearly that all modalities amplify the subject in the same way, and this became the common view, that is, if the subject of a modality is not explicitly restricted to what is, it is freely amplified. (We must therefore be cautious if we wish to use present-day possible world apparatus here.) Buridan drew an octagonal diagram of the propositions "Every/Some A is necessarily/possibly B/not B" and analyzed all the 56 logical relations between them. This made the map of modalities much clearer.

Combinations of syllogistic moods, modalities, and restrictions produce a huge number of cases, and logicians could not mention every case explicitly, although they did pursue a full theory of them. They also comment on cases where some propositions are nonmodal. We can only sketch some outlines now. In the direct first figure, everybody accepted MMM syllogisms as valid. Buridan and Pseudo-Scotus accept NNN, MNM, and NMN. The seemingly surprising NMN here shows the effect of ampliation. (For instance, every M is necessarily B, some A is possibly M, therefore some A is necessarily B.) Ockham accepts NNN only when restricted to actuals; for Buridan's school this is another valid syllogism, like several other moods resulting from a restriction of subjects of N or M. Buridan also accepts, for example, _NM with an assertoric major. In the second figure, Buridan mentions NNN, NMN, and MNM (and Pseudo-Scotus mistakenly adds MMM). These again have restricted versions (in the style of: if every actual B is necessarily M and every actual A is possibly not M, then no A is B). But Ockham allows no valid syllogisms here. In the third figure, all accept MMM. Buridan and Pseudo-Scotus accept NNN, NMN, and MNM, while Ockham accepts only restricted versions of these. Some of them, not precisely the same ones, are in Buridan.

Ockham's theory looks somewhat unfinished: His view of ampliation causes trouble, and he derives a great number of results by discussing individual examples one by one. Buridan, on the other hand, uses a very elegant deductive method with, for example, cleverly formulated conversion rules. His theory is the summit of medieval modal logic. His pupils Albert of Saxony and Marsilius of Inghen continued to give comprehensive accounts of modal syllogistics, with some usually unsuccessful innovations, but after them modal syllogisms seem to have fallen out of fashion.

5.4. Topics and Methodology

An important part of medieval logic was *topics*. The dialectics of the old *trivium* mostly belonged to it. The ultimate source was Aristotle's *Topics*, but a second and simpler authority that replaced it for a long time was Boethius's *De differentiis topicis*. The main subject in this inquiry was *loci*, *locus* being Latin for Aristotle's *topos* (literally "place," here something like "consideration"). Aristotle does not define his *topos*, whereas Boethius gives two meanings for *locus*. It can be a "maxim," a self-evident sentence that needs no further proof, but it can also be a logically relevant feature that distinguishes two sides. Confusingly, the distinction can be between sentences, like affirmative and negative, antecedent and consequent, or between concepts, like genus and species, part and whole. For example, the distinction between genus and species supports the maxim: What belongs to the genus belongs to the species.

Boethius's double notion of *loci* long guided medieval topics. On the other hand, Aristotle emphasized an aspect which was not so prominent in Boethius: Topics concerns dialectical argumentation, the finding, testing, and examining of plausible theses. Hence it is not restricted to methods of demonstrative scientific proof of necessary results.

Treatises as early as the eleventh century discuss topics, and this interest culminates with the Aristotelian revival of the thirteenth century. Thus, Peter of Spain gives a detailed list of various *loci* which follows Boethius closely. An important idea in such lists is that *loci* are supposed to guarantee the validity of an inference or argument that was not immediately valid because of its form. For instance, Peter's inference "The housebuilder is good, therefore the house is good" is surely not formally valid—and not even quantified—but it is "confirmed" by the *locus* of cause and effect: "That whose efficient cause is good, is itself also good."

We see that the result is still not conclusively proved, but the addition connects the argument to syllogistics. This need of support is characteristic of "enthymematic" arguments, demonstrative or not. Nowadays we are accustomed to think that they are valid because of some suppressed deductive premises, but medieval authors did not always see the matter so. Often they thought that the support came from a rule and not from an implicit premise. The terminists were inclined to think that all valid arguments are reducible to syllogisms; topics gives metalogical directions for finding suitable middle terms for the reduction.

After the early terminists, topics was still constantly discussed. After all, the *Topics* was a big book in the *Organon* and belonged to the obligatory courses, at least in part. But the heyday of topics was over when *logica moderna* was developed. It was no longer a really inspiring field, although it undoubtedly had some importance: Topics apparently influenced the growth of consequence theory (see section 6.1), and the doctrine of *loci* was also relevant in discussions concerning the foundation of syllogisms.

When the *consequentia* theory developed, both syllogisms and nonsyllogistic inferences could be seen as cases of the same general patterns. As a result, topics lost an important function. The arguments that were formerly studied in topics were, in the fourteenth century, normally included in consequences. Also, it is significant that topics was no longer connected to enthymemes but to dialectical arguments, that is, its special character was seen as epistemic. Usually, the leading logicians no longer treated topics as a separate subject at all—Ockham, for instance, studied topical arguments only as a relatively uninteresting special case. On the other hand, Buridan still painstakingly devoted a whole treatise to topics. Later the interest in topics diminished even more; Paul of Venice did not speak of it. However, commentaries on Aristotle's *Topics* were written throughout the fourteenth and fifteenth centuries, but no new ideas were presented.

The Aristotelian theory of science was highly abstract; while it had little contact to concrete problems, it did have a close connection to logic. The basic source for medieval discussion was *Posterior Analytics*, though direct commentaries on this difficult work were not very common. In the Aristotelian picture, developed for example by Aquinas, an ideal science consists of a system of demonstrative syllogisms. Their premises must be true, necessary, and certain. Premises can be derived by other syllogisms, but ultimately they rest on evident necessities. As Kilwardby says, "the demonstrator considers his middle term as necessary and essential, and as not possibly otherwise than it is; and so he acquires knowledge, which is certain cognition that cannot change." Science is thus a system of syllogisms about causes and essences; it can use logical principles, but logic itself is obviously not a science. Much of this grandiose view later had to be given up, when first Scotus problematized the notion of necessity and then Ockham problematized the notion of evidence.

6. New Approaches to Inferences

During the thirteenth century, four new domains of logical research broadly falling into the scope of propositional logic emerged: *consequences*, *obligations*, *insolubles*, and *sophisms*. In overall treatments of logic like Ockham's *Summa logicae* and Buridan's *Summulae dialectica*, these new branches of logic were discussed in the place traditionally occupied by treatments of dialectical topics in the sense in which they referred to what Aristotle discusses in his *Topics*. This is not to say that the traditional theory of dialectical topics, for which Cicero and Boethius had provided the classical texts, had disappeared altogether. Nor can we say that these new areas of logic had replaced the tradition of dialectical topics. Rather, the purposes aimed at by research in these new areas were seen to be approximately similar to those traditionally aimed at by the theory of dialectical topics, and consequently the new fields were taken to complement traditional discussions. In modern terms, we can say that the point of gravity was moving from the theory of argumentation toward formal logic.

Let us start with consequences, considering four different issues pertaining to this crucial area of logic. Late medieval discussion of consequences aimed at giving clear and specific determinations of (1) what is a consequence, (2) the definition of the validity of a consequence, (3) how they should be classified, and (4) rules concerning valid consequences.

6.1. What Is a Consequence?

In general, late medieval treatments of consequences understood them as inferences. That is, they were not called "true" (*vera*) or "false" (*falsa*), but rather were said to be "good" (*bona*), or simply "to be valid" (*valeo*) or "to hold" (*teneo*), or in the opposite case "to fail" (*fallo*). Despite an acknowledged close connection to conditional propositions, consequences were usually discussed separately as belonging to a different place in the overall structure of logic. Ockham, for example, discusses conditionals within his theory of propositions, and turns to consequences as a theory of nonsyllogistic inference in the beginning of III, 3 in *Summa logicae*: "After treating syllogism in general and demonstrative syllogism, we now have to turn to the arguments and consequences that do not apply the syllogistic form."

The genre of logical writings on theory of consequences seems to have arisen in the thirteenth century from recognition of the fact that a general theory of inferential validity can be formulated in addition to, or as an extension of, the traditional syllogistic system. As such, medieval logicians had been aware of the idea at least since Abelard's work, and Boethius had already composed a special treatise on what he called "Hypothetical syllogisms," that is, on propositional logic. Nevertheless, Walter Burley's *De puritate logicae* seems to have been the first overall presentation of logic to discuss the theory of inference systematically starting from general issues of consequences and moving toward more particular issues after that, allowing syllogistic only the minor position of a special case.

That most medieval logicians saw consequences as inferences and not as propositions is reflected in the fact that they aimed at formulating general rules (*regulae*) of valid inferences; traditional dialectical topics were also seen to belong to this set in addition to a number of more formal ones. The outstanding exception in this picture is John Buridan and his *Tractatus de consequentiis*. He explicitly defines consequences as hypothetical propositions consisting of two parts, the antecedent and the consequent, joined by a connective like "therefore" (*ergo*). Thus, Buridan's consequences amount to conditional propositions with specific content. He treats consequences as pieces of discourse that assert the validity of an inference from the antecedent to the consequent: "One follows from the other" (*una sequatur ad aliam*). Accordingly, Buridan does not discuss or lay down metalinguistic rules (*regulae*) of consequences in this treatise, but instead asserts "conclusions" (*conclusiones*) concerning what can be truly said about the kinds of sentences following from each other.

It seems clear that all prominent medieval logicians saw the distinction between the acceptability of performing an inferential step and the assertion that a valid inferential relation obtains. Whereas most logicians thought that consequences should be understood as inferences, Buridan made the opposite decision. For him, a consequence was a proposition, a conditional claim concerning an inferential relation between the antecedent and the consequent. He seemed to have had no followers in this opinion, but because of his prominent position in late medieval logic, his surprising stand has caused a number of misunderstandings concerning the issue both for medieval authors and for modern commentators.

6.2. Criteria of Validity

The simplest way to formulate the definition of inferential validity was to ground it on the idea that it is impossible for the antecedent to be true and the consequent false. Indeed, it seems that all late medieval definitions of validity can be seen as variously qualified or modified versions of this principle. In the first known treatise directly dedicated to consequences, Burley's *De consequentiis*, we find the definition that a consequence is valid if "the opposite of the consequent is repugnant to the antecedent." The problem with this definition is that it seems unclear in which sense we are to take the word "repugnant," since it is often used in a way that already contains reference to inferential connections. Indeed, Burley elsewhere opts for alternative definitions closer to the modal criterion.

In Buridan, we find the following list of three alternative descriptions concerning when some proposition "is an antecedent to another" or, in other words, a consequence is valid:

(a) "that is antecedent to something else which cannot be true while the other is not true" "*illa alia non existente vera*";

(b) "that proposition is antecedent to another proposition which cannot be true while the other is not true when they are formed simultaneously" "*illa alia non existente vera simul formatis*";

(c) "that proposition is antecedent to another which relates to the other so that it is impossible that howsoever it signifies, so is the case, unless howsoever the other signifies, so is the case, when they are formed simultaneously" "*sic habet ad illam quod impossibile est qualiterqumque ipsa significat sic esse quin qualiterqumque illa alia significat sic sit ipsis simul propositis.*"

Buridan finds each of these three descriptions problematic, but accepts the last, if it is understood in a suitably loose manner. The problem with the first definition is related to the standard medieval requirement that a proposition must be actually formulated to be true. This makes it clear that almost no consequence would be valid according to the first criterion, since the consequent need not be formulated when the antecedent is. The second aims at correcting this problem through the simple addition "when they are formed simultaneously," but falls prey to it as well. A consequence like "no proposition is negative, therefore no donkey runs" should be invalid, but turns out to be valid on criterion (b) as well as on (a), because the antecedent is never true when it is actually put forward. Thus, it cannot be true without the consequent being true even if they were simultaneously formed. With criterion (c), Buridan takes another strategy. He recognizes that the consequential relation should not be seen to obtain with the sentences themselves, not even between potentially formulated ones, but rather between their contents. However, Buridan did not believe that such sentential contents would exist (see the section about propositional significates, *complexa significabilia*), and therefore the formulation of the criterion (c) makes problematic ontological commitments. Apparently he could not find a formulation that would avoid them, and thus we are left without a satisfactory description of inferential validity.

It seems, nevertheless, that Buridan's strategy of transporting criteria of validity from the actual sentences to their significations or contents became a generally accepted one. In some interesting sense, which still puzzles modern scholars, Buridan's further discussions on the topic take a "mentalistic" turn in the conception of logical validity. He considered that logical validity depended on the mind in a more crucial sense than many of his predecessors. Some formulations by his followers made this mentalistic turn even more obvious in ways that we shall see in the next section.

6.3. Classifications of Valid Consequences

The most traditional medieval distinction among kinds of valid consequences was the distinction between those valid "as of now" (*ut nunc*) and those valid "simply" (*simpliciter*). Validity *ut nunc* was taken to mean something like validity given the way things now are: From "every animal is running," it follows *ut nunc* that Socrates is running, at the time in which Socrates exists as an animal. After his death, the consequence ceases to be valid. Simple validity, on the other hand, meant validity in all circumstances. In this sense, from "every animal is running," it follows that "every human is running." It is noteworthy that validity *ut nunc* also contains some kind of necessity, and thus it cannot be compared to twentieth-century material implication.

Late medieval logicians put their main interest in two other, philosophically more interesting distinctions. Somewhat confusingly the concepts "form" and "matter" were used in both distinctions, so that when we come to Paul of Venice, a consequence may be, for example, "formally formal" or "materially formal," since he combines the two distinctions into one systematic presentation. In both distinctions the issue was to separate a class of consequences that were valid in a privileged manner: not only valid, but "formally valid."

In one sense, formal validity meant a substitutional kind of validity, where a consequence is formally valid, if it "is valid for all terms" (*tenet in omnibus* *terminis*), and only materially valid if its validity is based on the special content of some of the terms used in the inference. In this sense, the paradigm examples of formally valid inferences were syllogisms in the Aristotelian figures, but also examples like *modus ponens* could be put forward. In the other sense, an inference was called formally valid only if the consequent was "formally included" (*includit formaliter*) in the antecedent or in the "understanding" (*intellectus*) of the antecedent; this kind of formal validity was often called "natural" or "essential" validity. It seems that the roots of both distinctions can be traced back to the early Middle Ages. At least Kilwardby gives ground for both distinctions. Nevertheless, the two distinctions seem to have had a somewhat different history. Furthermore, the concept of material validity remained in most treatments rather obscure. It seems, however, that especially as related to the latter definition of formal validity based on inclusion, material validity was often understood as having to do with certain properties of the propositions used. The paradigm cases of materially valid inferences followed the rules "from the impossible anything follows" and "the necessary follows from anything."

Let us first look at the latter kind of formal validity, the one based on the idea that the antecedent must "formally include" the consequent. The concept of "formal inclusion" seems to have been developed by late thirteenth-century theologians, such as Henry of Ghent, Godfrey of Fontaines, and Duns Scotus. In many texts the topic comes up in a discussion concerning the role of the third person in the divine Trinity, employing the special technique of obligations (see following). These discussions resulted in elaborated theories of what it means to say that a concept is included in another concept, or that an assertion conceptually includes and thus entails another claim. The primary examples studied by medieval logicians included inferences like "a human exists, therefore, an animal exists," and the explanation of their "formal" validity was based on the necessary conceptual or essential relation between the species "human" and the genus "animal." The concept "human" was said to "formally include" the concept "animal," and thus the inferences based on this relation were said to be "formally valid." In twentieth-century terms, we would rather describe them as analytically valid inferences.

William Ockham was aware of this discussion and aimed at bringing the results into the systematic context of logical theory. In the classification of *Summa logicae*, a consequence is formally valid if it is valid by general rules of a specific kind. They must concern the syntactic features of the propositions (*forma propositionis*) involved in the consequence. Also, the rules must be self-evident (*per se nota*). This part of the definition is in effect identical with, or at least comes very close to, the substitutional type of definition of formal validity. But on the same page Ockham also admits as formally valid consequences that are valid by something he calls an "intrinsic middle." His example is "Socrates does not run, therefore a man does not run," which is valid by the "intrinsic middle," "Socrates is a man." It seems that Ockham wanted to present this type of formal validity to allow also inferences based on something like conceptual inclusion within this group, the inclusion being expressible as an intrinsic middle.

Some 10 years later, Ockham's student Adam Wodeham explicitly distinguished between two different ways of understanding the concept of formal validity. One of them uses only the substitutional criterion, while the other accepts as formally valid all consequences based on truths known in themselves (*per se nota*). Insofar as Wodeham's *per se nota* refers to all analytic truths and not only conceptual inclusion, the definition is wider than that derived from the traditional slogan "formally includes," but it is clearly on the same track.

Material validity is defined by Ockham with reference to something he calls "general conditions of the propositions," and he gives the *ex impossibile quodlibet* rule as an example. In Ockham's case this is strange, since he clearly knew that from a contradiction it is possible to derive anything with rules which he allows to be formal. Do we, thus, have inferences that are both material and formal?

In his definition of formal validity, Buridan presents only the substitutional principle, without mentioning the idea of conceptual inclusion. His examples of inferences which are valid but not formally so, however, show that he was aware of the criterion but did not want to use it. He straightforwardly claims that those inferences, which are valid so that all substitutions of the categorematic terms with other terms are also valid, are formally valid. Among formally valid inferences, Buridan explicitly counts inferences from contradictions, though of course not from weaker impossibilities like from "a man is not an animal." These he classifies as material.

It seems that in the latter half of the fourteenth century, Buridan had few followers in his classification principles. Only Albert of Saxony seems to have accepted the substitution principle as the sole criterion of formal validity. The majority of logicians seem to have wanted to develop an idea which is closer to what was later in the twentieth century called analytic validity. The criterion of formal validity was, therefore, formal or conceptual inclusion of the conclusion in the premises. As an interesting special case, Paul of Venice presents a system that uses both concepts of formal validity, thus producing a very elaborate system.

6.4. Rules of Consequences

Usually medieval discussions of the theory of consequences also included a selection of rules warranting valid inferences. Instead of anything close to a complete listing of such rules, we must here satisfy ourselves with a look at the types of such rules presented in the medieval discussions.

We have already encountered two such rules: "from the impossible anything follows" and "the necessary follows from anything." These rules were practically never completely rejected in the later Middle Ages. However, their applicability in specific contexts was often limited, and as they were typically classified as materially valid, they were understood as belonging to a somehow inferior kind.

Medieval authors knew the main rules of so-called classical propositional logic. For example, detachment is put by Burley as concisely as possible with a term variable: “If A is, B is; but A is; therefore B is.” Transitivity rule is presented by Burley as the consequence “from start to finish” (*a primo ad ultimum*). He also discusses other examples of basic propositional logic of the kind, but when we turn to the later fourteenth century, the selection of rules of this type leaves nothing to be hoped for.

One type of rules of consequences that seems to have interested medieval authors quite widely is based on epistemic operators. These were often discussed by direct comparison with modal rules; if something is necessary, it is the case, and if something is known, it is the case. More interesting (and more disputable) examples of relevant inference schemes are more complicated. The rule “if the antecedent is known, the consequent is also known” was often held to be valid only on the further condition that the consequence itself is known.

The first rule of consequences in Ockham’s *Summa logicae* is that “there is a legitimate consequence from the superior distributed term to the inferior distributed term. For example, ‘Every animal is running; hence every man is running’.” All medieval logicians accepted this example as valid, though they often formulated the rule differently, and the explanation of the kind of validity varied. Buridan, who relied on the substitution principle, thought that the consequence is valid in a standard syllogistic mood with the help of a suppressed premise. But almost all other logicians thought that something like the rule given by Ockham suffices for showing the validity. The reference to the relation between a superior and an inferior term given in the rule was understood in terms of the criterion on conceptual containment. It seems that in twentieth-century terms, the rules of this type could be characterized as regulating analytic validity.

These types of rules already bring us close to Aristotle’s program in the dialectical topics presented in the *Topics*. This work was indeed much used in compilations of the listing of the rules for consequences. Also, in many works the lists contain rules that have more of the character of the theory of argumentation than of formal logic. The rules for consequences are indeed one of the places where the differences between modern and medieval conceptions of logic are most clearly visible.

6.5. Obligations

The genre of late medieval logical literature that has perhaps been the most surprising for modern commentators carries the title obligations (*obligationes*). The duties or obligations at issue in the treatises carrying this title were of a rather special kind. The basic idea was based on the Socratic question/answer game as described and regulated by Aristotle in the *Topics*. In the specific medieval variant of the game the opponent put forward propositions that had to be granted, denied, doubted, or distinguished by the respondent. In giving his answers, the respondent was expected to pay recognition to the truth, but especially to some special obligation given to him in the beginning of the exchange by the opponent. This duty was understood to override the general duty of following the truth, but not the general logical duty of respecting arguments and avoiding contradictions.

Here we cannot go into details concerning the different variants of the system, although a number of interesting logical issues arose through the study of the particular kinds of possible duties. The main type of an obligational disputation, as medieval authors knew it, was based on a *positum*, a sentence put forward by the opponent in the beginning as something that the respondent has to grant. This sentence was typically false, and often even impossible in some way not directly implying a contradiction (conceptually impossible, naturally impossible, etc.). Then the opponent put forward further propositions, and in answering them the respondent had to pay attention to inferential relations between the *positum* and these later *proposita*. Altogether four main alternative sets of exact rules of how the inferential connections ought to be recognized were developed in the Middle Ages.

According to one late thirteenth-century system, described by the Parisian logician Boethius of Dacia in his commentary on Aristotle's *Topics*, the respondent must grant everything that the opponent puts forward after the *positum*, with the sole exception of propositions that are inconsistent (*incompossibile*) with the *positum* or the set of *posita*, if there are several. Boethius divides propositions into "relevant" and "irrelevant" ones with the criterion of an inferential connection to the *positum*. Those inconsistent with the *positum* are called repugnant (*repugnans*), and those following from it are called sequent (*sequens*). The repugnant ones must be denied and the sequent ones must be granted. Others are irrelevant, and Boethius claims that the respondent must grant them, since this implies nothing for the *positum*.

In his discussion, Boethius relied on an already traditional terminology, but not all of the earlier authors would have agreed with his rules. The early fourteenth-century discussion took place mainly in England, and there a different set of rules came to be accepted as the traditional system. According to these rules, the respondent should of course grant the *positum* and anything following from it. Similarly, he should deny repugnant propositions. But he should grant true irrelevant propositions and deny false ones. After having granted or denied such propositions, he should take them into account in the reasoning. He should grant anything that follows from the *positum* together with propositions that have been granted earlier or whose negations have been denied. Thus, the respondent must keep the whole set of his answers consistent, but otherwise follow the truth.

Duns Scotus claimed that in an obligational disputation based on a false *positum* one need not deny the present instant, but one can understand the counterfactual possibility at issue in respect to the present instant. (Unlike many of his predecessors, Scotus denied the principle "what is, is necessary, when it is.") After Scotus, it became customary to think of the set of answers after the disputation as a description of some consistently describable situation. This brought obligational disputations close to counterfactual reasoning and thought experiments.

Richard Kilvington suggested in his sophismata an interesting revision of the rules apparently based on the idea that the disputation ought to describe the situation that would obtain if the false *positum* were true. He claimed that this principle ought to be taken as the rule guiding answers, giving the respondent a duty to grant what would be true and deny what would be false if the *positum* was true.

Kilvington's suggestion did not gain many followers. Most authors kept to the traditional rules, probably because Kilvington's rules seemed too vague. Formally valid inferential connections were taken to provide a better foundation for obligational disputations. But another revision was also suggested, and for some time it gained more followers. Roger Swineshed suggested that all answers ought to be decidable solely on the basis of the *positum* without recognition of any subsequent exchange. Swineshed's suggestion was that the respondent ought to grant the *positum* and anything following from it, and deny anything repugnant with it. Other propositions were to be taken as irrelevant, and they were not to be respected in the reasoning. This had the implication that irrelevant propositions would have to be kept separate from the mainline of the disputation, as a kind of second column in the bookkeeping. As Swineshed explicitly recognized, contradictions between the two columns could arise so that, for example, a conjunction may be denied when one of its conjuncts is granted as the *positum* and the other is granted as true and irrelevant.

The main logical topic studied in obligational disputations was logical coherence. The disputations were in essence structures allowing propositions to be collected together into a set, with evaluation of the coherence of the set as the crucial issue at each step. The different rules formulated the alternative exact structures for such a procedure.

6.6. Insolubles

Early treatises on obligations are often connected with treatises carrying the title "insolubles" (*insolubilia*). In these treatises something is laid down in a way similar to how the obligational *positum* is laid down, but the crux of the discussion is that the given propositions appear to describe a possible situation and yet they entail a contradiction. The case is thus paradoxical. As a common example from the obligational treatises themselves, we may mention the rule that the respondent ought not accept "the *positum* is false" as his *positum*. The case is, of course, closely analogous to what is nowadays known as the Liar Paradox.

It is not clear that obligational disputations were the original context of the genre of logic that came to be called *insolubilia*, since the first treatments of such paradoxes in their own right seem to be equally early and have other sources, too. But the way medieval logicians formulated their versions of the Liar Paradox comes with an obligational terminology and context.

If we turn to the mature treatises of the early fourteenth century, the paradigmatic insoluble is the proposition "Socrates is saying what is false," and the assumed situation is that Socrates utters this and only this proposition. Then it is shown that if the sentence is true it is false (because if it is true, what it signifies is the case), and if it is false it is true (because it signifies that it is false, and that was assumed to be the case). Because these results cannot stand together—every proposition is true or false but not both—a contradiction seems to follow from what is clearly possible, for the only assumption seems to be that Socrates makes a simple understandable claim.

Medieval logicians discussed a wide variety of carefully formulated analogous paradoxes, and it seems that some of them were formed to counter specific purported answers to the paradox. For example, if the paradox is claimed to result from direct self-reference, we may be asked to consider other examples. For example, medieval logicians considered cases where two or more people make assertions about the truth or falsity of each other's claims and thus produce a paradoxical circle. A paradox reminiscent of the Liar Paradox can be produced without any proposition referring to itself—the paradox is not dependent on direct self-reference. It is also interesting to note that some practical analogs of the paradox were considered. Assume, for instance, like Buridan, that Plato is guarding a bridge when Socrates wants to cross it. Then Plato says, "If you utter something false I will throw you into the river, and if you utter something true I will let you go." Socrates replies, "You will throw me into the river." Now, what should Plato do? Cervantes makes Sancho Panza face a similar problem when he is the fake governor of an island, and indeed, Cervantes probably got the paradox from some medieval treatment of logic.

The variety and the history of the different solutions of the insolubles is too wide and complicated to be even summarized here. Some main alternative solutions presented in the medieval discussion must suffice for now.

In the early discussions, the so-called nullifiers (*cassantes*) claimed that the one who utters a paradoxical sentence "says nothing." If Socrates says only the sentence "Socrates says what is false," he has not really uttered a proposition at all, and thus no truth value is needed. The problem, of course, is to explain precisely why the utterance fails to be a proposition. Some authors gave the reason that a part, like "false," cannot refer to its whole; but this thesis is too generalized.

In his *Sophistical Refutations*, Aristotle mentions the case where somebody says something that is simultaneously both true and false. This remark occurs in connection to the fallacy of confusing truth in a certain respect and absolute truth (*secundum quid* and *simpliciter*). Thus applications of this fallacy were often tried in solving insolubles, but understandably the results were not very convincing. One related suggestion was that insolubles were to be treated not as cases of genuine self-reference but instead as cases where a certain shift of reference (*transcasus*) takes place. When Socrates says that he is lying, he simply cannot mean that very utterance itself, and therefore we must look for some other utterance in the immediate vicinity. In the assumed case, this approach makes the insoluble false simply because there is nothing else that Socrates says.

Fourteenth-century logicians found all these suggestions too simple-minded. In the early 1320s, Thomas Bradwardine used symbolic letters for propositions and assumed that every proposition a signifies, in addition to its ordinary signification, even "a is true." (Strictly speaking, this was formulated as a general doctrine only later.) Substituting $a =$ "a is false," we get "a is false and a is true," a contradiction that shows that a is false. A similar strategy is further refined by William Heytesbury (1335). He puts the issue within the framework of obligation theory, discussing cases where insolubles are pressed on the respondent. All insolubles turn out to be false, but he admits that there is no general solution; what is needed is a careful study of what exactly is extraordinary in the signification of each relevant sentence.

Some authors, like Swineshed round 1330, argued that an insoluble proposition "falsifies itself." This requires a new opinion about truth: For the truth of a proposition, it does not suffice that it signifies what is the case, but it also must not falsify itself. This fundamental novelty may have been one reason why the theory was not generally accepted—and, moreover, its applications soon lead to obscurities.

Later, Gregory of Rimini and Peter of Ailly tried to utilize the doctrine of mental language in this context. The complex theory that Peter developed (in the 1370s) argues that spoken insolubles correspond to two conflicting mental propositions, whereas a mental proposition cannot ever be insoluble. This idea became well known but did not gain general acceptance.

To sum up, we may say that the common view was that certain propositions were called insoluble not because of logical puzzles that could not be solved but because providing a solution "is difficult," as many authors remark. It was generally agreed that insolubles were false. Only a few authors took seriously the possibility that the paradox might be a genuine one, one that did not allow any satisfactory solution. But even they did not think of insolubles as a threat to the system of logic as a whole. Insolubles were not considered to undermine the foundations of logic but simply to be one interesting branch of logical studies. One might surmise that this can derive from the idea of looking at logic as an art dealing with the rational structures embedded in the mental basis of ordinary language, rather than as a calculating system based on special foundations.

6.7. Sophismata

Buridan's *Summulae de Dialectica* concludes with an almost 200-page section containing sophisms (*sophismata*), which are examples construed in a rather distinct way so that they make the need of logical distinction clearly visible. Buridan's work is no exception; different kinds of collections of such sophisms are commonly found in medieval logic manuscripts. It seems that they were used in medieval logic teaching as exercises to show how general logical systems could be applied in practical contexts. But often they also contain interesting material that is not discussed in systematic treatises.

Separate collections of sophisms circulated throughout the Middle Ages. Perhaps the most famous of the early examples of such aids of teaching was known as the *magister abstractionum*. Little is known of the person, and he may not have been a single person. It is possible that we simply have a collection of examples which circulated among teachers of logic, who would each add their own examples and drop out others. Later, many authors of logical textbooks compiled their own collections of sophisms. This is what we find for example in the case of Buridan.

In early fourteenth-century Oxford, such a textual genre gained new significance by assuming a relatively specific independent role not only in the university curriculum (where undergraduate students in their first years of university were called "sophists" [*sophistae*]) but also in logical study. The collections of sophisms composed by Heytesbury, Kilvington and some other members of the so-called group of Oxford calculators were an important locus of logicolinguistic and mathematical study providing important results that were later used by pioneers of early modern science.

A sophism in this sense of the word consists of (1) the *sophisma sentence*; (2) a *casus*, or a description of an assumed situation against which the sophisma sentence is evaluated; (3) a *proof* and a *disproof* of the sophisma sentence based on the casus; and (4) a resolution of the sophism telling how the sophisma sentence ought to be evaluated and how the arguments to the contrary should be countered.

In the discussion of sophisma 47 in his collection, Kilvington assumes that the procedure in solving a sophisma must abide with the rules of obligational disputations. That is, the casus is to be understood as having been posited in the obligational sense, and thus anything following it would have to be granted and anything repugnant to it would have to be denied. From this viewpoint, the proof and disproof can be articulated as obligational disputations. Although an explicit commitment to using obligational rules such as Kilvington's is rare in the collections of sophisms in general, obligational terminology is omnipresent.

In many sophisms, the problematic issue was to show how the sophisma sentence was to be exactly understood. For this reason, sophismata became an especially suitable place for determining exact rules of scope and the interpretation of words serving important logical roles. Indeed, this is the context where late medieval logicians developed the exactitude in regulating logical Latin that was ridiculed by such Renaissance humanists as Juan Luis Vives.

Heytesbury's *Rules for Solving Sophisms* (1335) is a good representative of the genre and can thus be used as an example here. It consists of six chapters. The first is on a topic we have already mentioned, so-called insolubles. The second discusses problems of epistemic logic with sophisms based on the words "to know" and "to doubt." The third tackles problems connected to the use of pronouns and their reference. In the remaining three parts, Heytesbury turns to problems that may be better characterized as natural philosophy rather than logical analysis of language. The fourth part considers a traditional topic, the verbs "to begin" and "to cease," and thereby issues connected to limit decision problems and temporal instants. The fifth part, on maxima and minima, continues on the same tract from a different viewpoint. The sixth and final part is dedicated to "three categories," referring to the Aristotelian categories of place, quantity, and quality. Especially this last part and its discussions of speed and acceleration proved very fruitful in the early development of modern science despite the fact that all the cases studied in it are purely imagined and lack any sense of experiment. For example, instead of real bodies in motion, medieval logicians considered imagined bodies in motion. In fact, this chapter and others of its kind show how the medieval *secundum imaginationem* method, relying only on logicolinguistic analysis, was able to provide results that have often been misguidedly attributed to experimental scientists working centuries later.

One of the specific techniques used in solving sophisms deserves treatment of its own in a history of logic. In early thirteenth-century texts, a sentence like "Socrates begins to be pale" was analyzed as something like "Socrates was not pale and Socrates will be pale." The analysis was accompanied with a discussion on which of the two conjuncts in the particular kind of change at issue should be given in the present tense, and how one should formulate the continuity requirement that Socrates, say, will be pale immediately after the present instant, even before any given determinate future instant. Such an analysis became a standard technique used in a large variety of cases and was called "exposition" (*expositio*). Without going into the particulars of the specific verb "to begin," it is worth pointing out here that the idea in such an analysis is to break down the sentence containing the problematic "exponible term" into a conjunction or a disjunction that is equivalent in its truth conditions. For fourteenth-century logicians, it was a commonly accepted doctrine that there is a large number of terms that admit, or in the contexts of a sophism, demand such an analysis. Furthermore, this kind of analysis was taken to be necessary for practically all philosophically central terms if there was a need to treat them in a logically exact manner.

7. The End of the Middle Ages

7.1. Later University Logic

Undoubtedly, the main plot of medieval Aristotelian logic lies in the development that began from the early terminists and led to the stage of Burley and Ockham, and then had its academic culmination in the systematic work of Buridan. But the discipline of logic survived after that, and some new special features appeared and new developments took place in the late Middle Ages.

It is probably true to say that logicians were no longer very original during this time. But here it is necessary to emphasize that logic was a widespread and multiform discipline; the volume of material is very great, and much of it is still unexamined.

A gradual change happened in philosophy in general during the fourteenth century, a change whose background is hard to explain. It has been pointed out that the whole cultural climate was no longer the same: The fourteenth century included great political upheavals; the Church had difficulties that led to the great schism; various protest movements appeared, and so on. All this contributed to the loss of the previous unity. It is customary to start the "autumn of the Middle Ages" from 1350, but this demarcation is largely symbolic; the only concrete thing that can support it is the Black Death, which killed many philosophers in 1349. After 1350, philosophy was still practiced in the old style, and logic has hardly ever been as prominent a part in philosophy as in the latter half of the fourteenth century. However, the overall authority of philosophy and logic started to diminish.

Let us try to sketch an overview of the historical development. Ockham, a political dissident, had never made an uncontested breakthrough—in fact, he was considered an extremist even among nominalists. In logic, however, his thought had a wide influence. Buridan, then, had more indisputable prestige, and as regards logic, his influence became dominant in Parisian philosophy during the 1340s. In this field he had two extremely competent pupils, Albert of Saxony (d. 1390) and Marsilius of Inghen (d. 1396). After the generation of Buridan's students, the position of Paris weakened, although it was still the most famous university.

England underwent a quite distinctive process. In the beginning of the fourteenth century the best logicians were English, and even after them there were original figures in Oxford, like Bradwardine, Heytesbury, and Billingham. Then, after 1350, logic turned to great technical sophistication but little essentially original appeared in the works of logicians such as Hopton, Lavenham, Strode, Feribrigge, and Huntman. Soon after 1400, a complete collapse took place in England, and only some elementary texts were produced during the fifteenth century.

But English logic was, however, very influential in the late Middle Ages on the Continent. English works of the fourteenth century were studied and commented on in Italy. Particularly Ralph Strode's logic achieved great fame. Paul of Venice had studied at Oxford, and he transmitted the comprehensive English tradition to the Italian logicians of the fifteenth century: Paul of Pergula, Gaetano of Thiene, and others.

Moreover, the fifteenth century is the era of the triumph of the university, which also involved a geographical expansion of philosophical studies. Hence we meet a number of new active centers of logic emerging in Central Europe, in universities like Prague, Cracow, and Erfurt.

A typical feature in fifteenth-century philosophy is a conscious turn toward old masters. Thus, philosophy formed into competing schools with their own clear-cut doctrines; this process was promoted by the commitment of religious orders to their official authorities and by the allotment of chairs in philosophy. These Thomist, neo-Albertist, Scotist, and nominalist currents were not very innovative in logic, though some of their leaders were first-rate logicians (like the Scotist Tartaretus).

The form of logical works changed gradually. Instead of voluminous commentaries, two other types of work became popular: shorter discussions of individual subjects, and more general *summulae* expositions. A far-reaching step was the innovation of printing, which led to the promotion of textbooks in particular. (The first printed logical book was the *Logica parva* by Paul of Venice, in 1472.) On the whole, we can say that logic was no longer very creative; there were few original results, and perhaps they were not even actively pursued. We can feel some signs of the later sentiment that the science of logic had already been completed. In statements like this, we must remember, though, that there has been particularly little historical research on fifteenth-century logic.

Attention was often concentrated on earlier results; thus there was much interest in all kinds of special cases and counterexamples, which we cannot discuss here. Generalizing crudely, we might say that the exponibilia, the sophismata, and the insolubilia became especially popular themes, whereas the fundamental questions of terms, propositions, and inferences were less debated. Modal logic seems to disappear, though it has a surprising revival at the end of the fifteenth century (with Erfurtians like Trutvetter). At the same time there is also a revival of philosophical logic in Paris (e.g., Scotsmen around John Mair).

The strictly formal part of older logic, such as syllogistics, was still taught everywhere, and occasionally even cultivated in so far as there was an opportunity to develop it. A famous example is the innovation of the so-called *pons asinorum.* Fifteenth-century authors formulated clearly this virtually mechanical method for finding a suitable minor premise by means of which a given conclusion can be syllogistically inferred from a given major.

In less formal matters, we encounter an interesting line by examining the widely read *Speculum puerorum* (1350s) by Richard Billingham. He discusses the *probatio*, literally "proof" but also meaning "trial," of propositions and concludes that it is only possible by a further *probatio* of its terms. "Immediate" terms are simple, but others can be submitted to some of the three forms of such a treatment. First, "exponible" terms can be replaced by several occurrences of simpler terms in a conjunction of simpler propositions which is equivalent to the original one. Thus "only a man" is exponible in the proposition "only a man runs," and its exposition leads to the equivalent "a man runs and nothing but a man runs." "Resoluble" terms are replaceable by simple terms, leading (not to equivalents but) to truth grounds; thus "a man" is resoluble to "this" since "a man runs" has a truth ground "this runs and this is a man." And for "official" terms, it can be shown that they hold an office together with *dicta*, like the modal and attitudinal operators do. By means of the *probatio* of terms, the proposition ought to acquire a logically elementary form, and problems arising from difficult constructions can then be handled. Finally Billingham gives some grammatical rules for advancing without error in the *probatio*.

Similar ideas can be found in a number of later English and Italian authors who discuss such basically non-Aristotelian themes. The attention turned to *logical grammar*. Logic courses often started from *logica vetus*, continued with material from the Aristotelian *Prior* and *Posterior Analytics*, and then concentrated on the new themes. This feature can be seen as a mark of a shift from logic in the strict sense toward *conceptual analysis* of logically difficult items: problematic concepts, ambiguous linguistic constructions, and so on. Accordingly, much attention was awarded to questions of grammatical deep structure and its accurate expression by means of variants in lexical forms and word order. The serious nature of these problems can now be appreciated again, in the light of present-day grammatical theory, but it is of course true that fifteenth-century authors did not have a sufficient technical apparatus for mastering their Latin sentences. It is also easy to understand that these undertakings seemed useless and annoying to many critics.

7.2. Reactions

It is common to speak about "medieval logic," and one easily thinks of it as a monolithic totality. Perhaps we have managed to say that the truth is much more complex. But all the authors we have discussed so far had a solid Aristotelian background. There were, however, even other tendencies, which started to grow during the fifteenth century. We might prepare the way for the novelties by mentioning earlier dissidents.

The Aristotelian methodology in science was rather restrictive, and for a long time repeated attempts had been made to find a place for something more innovative. Bacon is perhaps the most famous among these authors: He showed great curiosity in matters of empirical science and made initiatives in the philosophy of science. But his logic seems to follow well-known Aristotelian lines. A much more perplexing case is Raimundus Lullus (Ramón Llull, c. 1235–1316). Having no academic training, he did not care about *logica moderna*; instead, he sought to create an original way of argumentation that would undeniably prove Christian dogmas to infidels. This so-called *Ars magna*, to which he gave several formulations, uses various basically neo-Platonic sources. As basic concepts, he chooses some central divine attributes and cross-tabulates them with certain logicometaphysical aspects. This ought to produce, in the way of multiplication tables, a scheme of interesting manifestations. Lullus also suggested that concepts should be written on concentric circles and arguments performed by rotation of the circles. In fact, Lullus never achieved any logical results, and his program rests heavily on theological premises. But he introduced the idea of purely combinatorial procedure (with symbolic letters), and this was something that fascinated many later authors. "Lullists" reappeared during the fifteenth century, and even Leibniz was interested in Lullus.

Late medieval university logic acquired a respected opponent when Italian humanists began to propagate their new ideals. In the middle of the fourteenth century, Petrarch had violently attacked scholasticism and particularly logic, making it clear that professional logic was a corrupt and useless discipline that could not benefit a literary civilization. His leading followers, such as Bruni and Bracciolini, were more detailed in their criticisms. According to them, what is sensible in logic is delivered through the studies of language and dialectics, whereas university logic is mostly incomprehensible sophistry. They also pointed out, correctly, that medieval logic consisted of additions made by barbarians to the classical heritage.

The early humanists mainly expressed nothing but their discontent, but a more substantial alternative logic was developed by the famous philologist Lorenzo Valla (1407–1457) in his *Dialecticae disputationes*. He argued that a lot of the scholastic problems were actually illusory and resulted from obscure and abstract misinterpretation of questions that were essentially linguistic. Valla admitted that a small kernel of elementary logic was needed, as ancient Romans had already admitted, but for him formal validity was not as interesting as the informal convincing power of arguments. Thus he focused on the dialectical theory of reasoning and discussion, emphasizing matters of grammar and style. His work anticipates the revival of topics in a new form.

A similar nonscholastic development was continued by many other authors. Gradually the humanist influence extended outside Italy to the whole of Europe, and there grew a conscious effort to form a simple logic free of tradition. In this process, the new logic also found a place in the academic environment and much logical literature turned to dialectical issues, new ancient sources became known, and logic definitely entered the era of printed books. All this amounts to a basic transformation, and the next part can well start with it.

Selected Further Readings

Ashworth, E. J. 1974. *Language and Logic in the Post-Medieval Period* (Synthese Historical Library 13). Dordrecht: Reidel.

Ashworth, E. J. 1978. *The Tradition of Medieval Logic and Speculative Grammar from Anselm to the End of the Seventeenth Century: A Bibliography from 1836 Onwards* (Subsidia Mediaevalia 9). Toronto: Pontifical Institute of Mediaeval Studies.

Bäck, Allan. 1996. *On Reduplication: Logical Theories of Qualification* (Studien und Texte zur Geistesgeschichte des Mittelalters 49). Leiden: Brill.

Biard, Joël. 1989. *Logique et théorie du signe au XIV siècle*. Paris: Vrin.

Boh, Ivan. 1993. *Epistemic Logic in the Later Middle Ages*. London: Routledge.

Braakhuis, Henk A. G., C. H. Kneepkens, and L. M. de Rijk, eds. 1981. *English Logic and Semantics from the End of the Twelfth Century to the Time of Ockham and Burleigh* (Acts of the Fourth European Symposium of Medieval Logic and Semantics). Leiden: Ingenium Publishers.

Ebbesen, Steen, ed. 1995. *Sprachtheorien in Spätantike und Mittelalter* (Geschichte der Sprachtheorie, Teil 3). Tübingen: Narr.

Green-Pedersen, Niels J. 1984. *The Tradition of the Topics in the Middle Ages*. München: Philosophia Verlag.

Jacobi, Klaus, ed. 1993. *Argumentationstheorie: Scholastische Forschungen zu den logischen und semantischen Regeln korrekten Folgerns* (Studien und Texte zur Geistesgeschichte des Mittelalters 38). Leiden: Brill.

Knuuttila, Simo. 1993. *Modalities in Medieval Philosophy*. London: Routledge.

Kretzmann, Norman, Anthony Kenny, and Jan Pinborg, eds. 1982. *Cambridge History of Later Medieval Philosophy*. Cambridge: Cambridge University Press.

Kretzmann, Norman, and Eleonore Stump, eds. 1988. *Cambridge Translations of Medieval Philosophical Texts*, vol. 1: *Logic and the Philosophy of Language*. Cambridge: Cambridge University Press.

Kretzmann, Norman, ed. 1988. *Meaning and Inference in Medieval Philosophy*. Studies in Memory of Jan Pinborg. Dordrecht: Kluwer.

Lagerlund, Henrik. 2000. *Modal Syllogistics in the Middle Ages* (Studien und Texte zur Geistesgeschichte des Mittelalters 70). Leiden: Brill.

Nuchelmans, Gabriel. 1973. *Theories of the Proposition*. Amsterdam: North Holland.

Nuchelmans, Gabriel. 1980. *Late Scholastic and Humanist Theories of the Proposition*. Amsterdam: North Holland.

Panaccio, Claude. 1999. *Le discours intérieur de Platon à Guillaume d'Ockham*. Paris: Seuil.

Perler, Dominik. 1992. *Der propositionale Wahrheitsbegriff im 14. Jahrhundert* (Quellen und Studien zur Philosophie 33). Berlin: Walter de Gruyter.

Pinborg, Jan. 1972. *Logik und Semantik im Mittelalter*. Stuttgart: Frommann-Holzboog.

Pironet, Fabienne. 1997. *The Tradition of Medieval Logic and Speculative Grammar: A Bibliography (1977–1994)*. Turnhout: Brepols.

Read, Stephen, ed. 1993. *Sophisms in Medieval Logic and Grammar* (Nijhoff International Philosophy Series 48). Dordrecht: Kluwer.

de Rijk, Lambertus Marie. 1982. *Through Language to Reality*. Studies in Medieval Semantics and Metaphysics. Northampton: Variorum Reprints.

Rosier-Catach, Irène. 1983. *La grammaire spéculative des modistes*. Lille.

Spade, Paul Vincent. 1988. *Lies, Language and Logic in the Late Middle Ages*. London: Variorum Reprints.

Spade, Paul Vincent. 1996. *Thoughts, Words and Things: An Introduction to Late Mediaeval Logic and Semantic Theory*, available at http://pvspade.com/Logic/noframes/index.shtml.

Stump, Eleonore. 1989. *Dialectic and its Place in the Development of Medieval Logic*. Ithaca N.Y.: Cornell University Press.

Yrjönsuuri, Mikko, ed. 2001. *Medieval Formal Logic: Obligations, Insolubles and Consequences* (New Synthese Historical Library 49). Dordrecht: Kluwer.

3

Logic and Philosophy of Logic from Humanism to Kant

Mirella Capozzi and Gino Roncaglia

1. Humanist Criticisms of Scholastic Logic

The first impression of a reader who "crosses the border" between medieval and Renaissance logic may be that of leaving an explored and organized field for a relatively unexplored and much less ordered one. This impression is emphasized by the fact that while in the medieval period we can assume, despite relevant theoretical differences, some consensus about the nature and purpose of logic, such an assumption cannot be made with reference to the postmedieval and Renaissance period: The many "logics" coexisting and challenging each other were often characterized by deeply divergent assumptions, articulations, and purposes. As far as logic is concerned, we could almost be tempted to use this "explosion of entropy" as the very marker of the shift between the medieval and the Renaissance period.

The development of humanism, with its criticism of the late medieval logical tradition, is not the only factor contributing to this situation, but surely is a relevant one. Excessive and artificial subtlety, lack of practical utility, barbarous use of Latin: These are the main charges that humanist dialecticians made against scholastic logic. Such charges do not simply point out formal deficiencies that could be eliminated within a common logical framework, but call for a change of the logical paradigm itself. The effort to address such charges had a deep influence on the evolution of logic and resulted in a variety of solutions, many of which were based on contaminations between selected but traditional logical theories, on the one hand, and mainly rhetorical or

Though we decided on the general structure of this chapter together, sections 1–4 and 8 are by Gino Roncaglia, while sections 5–7 and 9–11 are by Mirella Capozzi.

pedagogical doctrines on the other. But the charges themselves were initially made outside the field of logic: One of the very first invectives against scholastic logic came from Francesco Petrarca (1304–1374), hardly to be considered a logician (Petrarca 1933–42, I, 7).

The central point at issue is the role of language. The late medieval scholastic tradition used language as a logical tool for argumentation, and favored the development of what J. Murdoch (1974) aptly called “analytical languages”: highly specialized collections of terms and rules which—once applied to specific and definite sets of problems—should help guarantee the formal precision of reasoning. In this tradition, the use of a simplified and partly artificial Latin could help the construction of sophisticated formal arguments. The humanists, on the contrary, privileged the mastery of classical Latin. For them, language—together with a few simple and “natural” arguments taken from ancient rhetoric—was a tool for an effective and well-organized social and pedagogical communication.

Besides the different theoretical standpoints, there is a social and cultural gap between two different intellectual figures. Scholastic-oriented teachers are usually university professors who tend to consider logic, philosophy, and theology as specialized fields. For them, knowledge is reached through a self-absorbing (and largely self-sufficient) intellectual activity, whose formal correctness is regulated by logic. Many humanist dialecticians, on the contrary, do not belong to and do not address themselves to the academic world: They consider logic a tool to be used whenever language is used with rhetorical or practical purposes, and regard a broad “classical” culture more important than a specialized and abstract one (see Jardine 1982, 1988).

One should be careful, however, in assessing the reasons for the privilege humanists accorded to rhetoric. For the humanists, logic—or rather dialectic, to use the term that, already present in the Ciceronian tradition and in the Middle Ages (see Maierù 1993), was preferred by most humanist and Renaissance authors—*has* to do with the use of arguments. But to be practically effective, such arguments have to be natural, aptly chosen, easily stated and grasped, expressed in good, classical Latin. And they don’t need to be demonstrative arguments: Probable arguments are also included within the scope of dialectic.

One should also be careful in considering humanism as a monolithic movement aimed at banishing all reminiscence of medieval logic. Humanism is not chronologically subsequent to scholasticism, and many humanists knew late scholastic logical texts fairly well, such as those by Paul of Venice. Some even praised them (Vasoli 1968, 20–23; Perreiah 1982, 3–22; Mack 1993, 14–15). Nevertheless, formally correct and truth-preserving arguments were considered as only some of the tools available to a good dialectician. The latter’s aim is to master the art of using language (*ars bene disserendi*), the Ciceronian *disserendi diligens ratio*, and this requires not only demonstrative skills but also the ability to persuade, to construct probable arguments, to obtain consensus.

The definition of dialectic provided by Rudolph Agricola (1444–1485)—one of the many Renaissance variations on Cicero’s own—is representative of this point of view. According to it, dialectic is the "ars probabiliter de qualibet re proposita disserendi" (art of speaking in a probable way about any proposed subject). The explication of "probabiliter" clarifies the broad scope of the term (see Mack 1993, 169–173, where "probabiliter" is translated as "convincingly"): "probable (*probabile*) in speaking is not only what is actually probable, that is, as Aristotle states, what is accepted by all, or by the most part, or by the learned. For us, probable is what can be said about the proposed subject in an apt and adequate way" (Agricola 1967, 192). This meaning of the concept is broad enough to include good old-fashioned demonstrative arguments in the field of dialectic (Risse 1964–70, I, 17–18), but they are no longer the only kind of arguments a dialectician should take into account.

A first introduction to sources, principles, and precepts of humanist-oriented logic is provided by the works of the prominent humanist dialectician Lorenzo Valla (1407–1457), who, significantly, received his cultural training mostly within the humanist circles of the papal *curia*. While some of the earlier humanists were content with a dismissal of scholastic logic—Petrarca's and Bruni's invectives against the *barbari britanni* being the most often quoted testimony of this attitude (Garin 1960, 181–195; Vasoli 1974, 142–154)—in his *Repastinatio dialecticae et philosophiae* (Valla 1982), Valla added to heavy criticism of traditional logical doctrines a complete and systematic reassessment of the nature and purpose of dialectic from a humanistic point of view.

According to Valla, dialectic deals with demonstrative arguments, while rhetoric deals with every kind of argument—demonstrative as well as plausible ones. Therefore dialectic is to be considered as a part of rhetoric, and rhetoric has to provide the widest spectrum of argumentative tools to all branches of learning. Moreover, dialectic should be simple and disregard all the questions that, though discussed by logicians with technical logical tools, actually pertain to Latin grammar. During the Middle Ages the relation between logic and grammar had been closely investigated by the so-called modist logicians. They worked at a sophisticated *speculative grammar*, based on an ontologically grounded correspondence between ways of being, ways of thinking, and ways of signifying. Valla's grammar, on the contrary, is based on the Latin of classical authors, and therefore on a historically determined *consuetudo* in the use of language. Valla thus carries out what has been described as a "deontologization" of language (Camporeale 1986; Waswo 1999).

Valla devotes the first of the three books of his *Repastinatio* to the foundations of dialectic and to a discussion of the Aristotelian doctrine of the categories. Here, too, Valla applies his general rule: simplification through reference to concrete uses of Latin, rather than to an abstract metaphysical system. The 10 Aristotelian categories are thus reduced to 3—substance, quality, and action—and examples are given to show how the remaining categories can be reduced to quality and action. Similarly, the transcendental terms, which according to the medieval tradition "transcend" the division among the 10 categories and are reciprocally convertible, are reduced to the only term "res." The reason why Valla prefers the term "res" to the traditional "ens" is that "ens" in classical Latin is not a noun but a participle that can be exposed as "that thing (*res*) which is." Therefore the term "res" is the true fundamental one. This example shows how Valla explains problematic terms or sentences by offering a reformulation considered more precise and easier to analyze. The practice of explanation through reformulation was familiar to medieval logicians under the name of *expositio*, but Valla uses *expositio* to reach linguistic, rather than logical clarification.

Valla's second book is devoted to proposition and addresses the question whether all propositions should be reduced to the basic tripartite form: subject–copula–predicate ("*A* est *B*"). This question was the object of a long debate, continued during the whole period we are dealing with (Roncaglia 1996), and had usually been investigated under the assumption that it was the *logical* structure of the proposition at issue. Valla, on the contrary, perceives the problem as related to the *grammatical* structure of the proposition, and accordingly offers a negative answer, since in the use of Latin the construction "est + participle" (*Plato est legens*) is not equivalent to the use of an indicative form of the verb (*Plato legit*). The Spanish humanist Juan Luis Vives (1492–1540) will share the same attitude (see Ashworth 1982, 70).

To support his contention, Valla considers propositions like *Luna illuminatur*, which—in Latin—can be transformed into a tripartite form only through a shift in meaning. A further argument is drawn from the idea that the participle form of the verb may be seen as somehow derivative with respect to the indicative form. Therefore—if something is to be reduced at all—it should be a participle like *legens*, to be reduced to *qui legit* (Valla 1982, 180). Logicians should not superimpose their logical analysis to the "good" use of language, but should rather learn from it. Language should be studied, described, and taught, rather than "corrected" from an external point of view.

Valla did not consider the study of modal propositions as pertaining to logic (hence his complete refusal of modal syllogistic). This refusal—common to most humanistic-influenced Renaissance philosophy—is once again defended on linguistic rather than purely logical grounds. Why should we attribute to terms like "possible" and "necessary" a different status from that of grammatically similar terms like "easy," "certain," "usual," "useful," and so on? (Valla 1982, 238; see Mack 1993, 90; Roncaglia 1996, 191–192.)

Valla's third book, devoted to argumentation, preserves the basic features of Aristotelian syllogistic, but dismisses the third figure and, as already noted, modal syllogisms. Owing to his desire to acknowledge not only demonstrative but also persuasive arguments, Valla pays great attention to hypothetical and imperfect syllogisms and to such nonsyllogistic forms of argument as *exemplum* and enthymemes.

The final section of Valla's work is devoted to sophistic argumentations. Medieval discussions of sophisms allowed logicians to construct interesting, complex, and borderline situations to test the applicability and the effectiveness of their logical and conceptual tools. Valla is fascinated by the persuasive and literary strength of "classical" problematic arguments, such as the *sorites* (a speech proceeding through small and apparently unavoidable steps from what seems an obvious truth to a problematic conclusion) or the *dilemma*, in which all the alternatives in a given situation are considered, only to show that each of them is problematic. Valla does distinguish "good" and "bad" uses of these kinds of "arguments," but his criterion is basically that of practical usefulness in persuasive rhetoric.

Valla's *Repastinatio* is also a typical example of the importance humanists assigned to the "invention" (*inventio*) of arguments, connected with topics. Renaissance dialecticians considered Aristotle's *Topica* as a systematic treatment of practical reasoning, and complemented it with Cicero's *Topica* and with the treatment of topics included in Quintilian's *Institutio oratoria*, which—rediscovered in 1416—had become one of the most popular textbooks on rhetoric by the end of the century, while Boethius's *De differentiis topicis*, widely used in the Middle Ages (Green-Pedersen 1984), had only few Renaissance editions (Mack 1993, 135). Both Cicero's and Quintilian's treatment of topics helped shift the focus from "formal" disputations to rhetorical and persuasive ones.

The most complete and influential Renaissance study of topics is contained in Agricola's *De inventione dialectica* (Agricola 1967, 1992). Agricola grounds his conception of topics on his realist conception of universals (Braakhuis 1988). In his opinion, things are connected by relations of agreement and disagreement, and topics are orderly collections of common marks, which help us organize and label relations, and find out what can or cannot be said about a given thing in an appropriate way. While being systematically arranged, topics, according to Agricola, are not a closed system: The very possibility of viewing things from different angles and perspectives, of relating them in new ways, not only enables us to draw or invent arguments but also allows us to find new common marks.

We have already considered Agricola's definition of dialectic. In his opinion, topics are the method of dialectical invention, while the discourse (*oratio*) is its context. There are, however, two different kinds of dialectical discourse: exposition (*expositio*) and argumentation (*argumentatio*). The former explains and clarifies, and is used when the audience doesn't need to be convinced, but only enabled to understand what it is said. The latter aims at "winning" assent, that is, at persuading. Although argumentation is connected with disputation, necessary arguments are not the only way to win a disputation: Plausible and even emotionally moving arguments should be considered as well. Agricola's concept of argumentation is thus connected with rhetoric, a connection strengthened by the fact that both use natural language. This explains why Agricola has no use for the kind of highly formalized, analytical language used by medieval and late medieval logic.

However negative Valla's and Agricola's attitude toward the logical tradition, it was never as negative as that of Petrus Ramus (Pierre de La Ramée, 1515–1572). According to his biographer Freigius, Ramus's doctoral dissertation (1536) defended the thesis: "everything that Aristotle said is misleading (*commentitium*).” This does not imply—as many assumed—that Ramus considers all Aristotelian theories to be false: In his opinion, Aristotle is guilty of having artificially complicated and corrupted the simple and “natural” logic which Aristotle’s predecessors—notably Plato—had devised before him (Risse 1964–70, I, 123–124). Scholastic logic is obviously seen by Ramus as a further step in the wrong direction.

Various versions of Ramus’s logic (including the 1555 *Dialectique*, in French: Ramus 1996; for a survey of the different editions of his works and of the stages marking the complex development of Ramus’s dialectic, see Bruyère 1984) were published between 1543 and 1573. After his conversion to Protestantism in 1561, his library was burned, and he had to flee from Paris. Ramus died on August 26, 1572, killed on the third day of the St. Bartholomew’s massacre. His being one of the Huguenot martyrs undoubtedly boosted the fortune of his already popular works in Calvinist circles.

Ramus’s concept of dialectic is based on three main principles: Dialectic should be *natural* (its foundations being the “eternal characters” which constitute, by God’s decree, the very essence of our reasoning), it should be *simple* (it deals with the correct way of reasoning, but disregards metaphysical, semantic, and grammatical problems as well as unnecessary subtleties), and it should be *systematically organized*, mainly by means of dichotomic divisions. Therefore, Ramus’s books extensively used diagrams, usually in the form of binary trees: A feature that may be connected—as argued by Ong (1958)—with the new graphical possibilities offered by printed books, and that will influence a huge number of sixteenth- and seventeenth-century logic textbooks, not only within the strict Ramist tradition.

The first and foremost division adopted by Ramus is Cicero’s division between invention (*inventio*) and judgment (*iudicium* or *dispositio*). They are the first two sections of logic. A third section, devoted to the practical and pedagogical exercise of dialectic (*exercitatio*), is present in the first editions of Ramus’s logical works but disappears after 1555.

The *inventio* deals with the ways arguments are to be found. Because arguments are to be found and classified by means of topics, according to Ramus, the treatment of topics should precede, rather than follow (like in Aristotle), that of judgment. Ramus’s table of topics, organized by means of subsequent dichotomic divisions, is strongly influenced by Agricola and by Johannes Sturm (1507–1589), who taught dialectic and rhetoric in Paris between 1529 and 1537 and greatly contributed to the popularity of Agricola in France.

Ramus’s treatment of judgment is also unconventional. While in traditional logic this section presupposes an extensive treatment of proposition, Ramus deals with this subject in a sketchy way and adds an independent (albeit short) section on the nature and structure of proposition only in the 1555 and successive editions of his work. In the last edition Ramus follows Cicero in using the term *axioma* to refer to a categorical proposition (having used earlier the term *enuntiatio* or *enuntiatum*), while he always gives the more specific meaning of major premise of a syllogism to the term *propositio*.

Syllogism and its various forms (including induction, example, and enthymeme) constitute the core of the "first judgment": the first of the three sections in which Ramus divides his treatment of judgment in the earlier editions of his dialectic. Ramus's explicit effort is that of simplifying Aristotelian syllogistic, but during the years between the 1543 edition of the *Dialecticae institutiones* and his death, his syllogistic underwent so many changes that it is impossible to give a faithful account of it in a few pages. Typical of Ramus's syllogistic is his use of the terms *propositio*, *assumptio*, and *complexio* to refer to the major premise, minor premise, and conclusion of a syllogism, and his tendency to favor a classification of syllogisms according to the quantity of the premises, considering as primary moods those with two universal premises. In the earlier editions of his dialectic, Ramus held that all moods with particular premises should be reduced to universal moods. He admitted some of them later on, but banned the *reductio ad impossibile* used to reduce second and third figure moods to the first figure. But Ramus's better known innovation in the field of syllogistic is the so-called Ramist moods: syllogisms in which both premises are singular, accepted on the ground that individuals could be seen as (lowest) species. The discussion about Ramist moods will keep logicians busy for most of the subsequent century.

The second section of Ramus's treatment of judgment (called "second judgment" in the earlier editions of his work) deals with the ways to connect and order arguments by means of general principles. Ramus attributes great importance to this "theory of method," which he further develops in the later editions of his logical works, and which in his opinion shapes the whole system of science (also offering the conceptual foundation for an extensive use of dichotomies). According to Ramus, the dialectical method (*methodus doctrinae*) goes from what is most general to what is most particular. This is done by means of divisions that, in turn, are drawn on the base of definitions expressing the essence of the concepts involved. Division and definition are thus the two main tools of method. The opposite route, going from particular instances to more general concepts (*methodus prudentiae*), might be used when either the lack of a more general conceptual framework or reasons of practical convenience force us to dwell on single or partial pieces of information. However, it cannot guarantee certainty; and is therefore mainly used in rhetorical discourse aiming at persuasion, rather than in demonstrative reasoning. In Ramus's opinion, however, the distinction between *methodus doctrinae* and *methodus prudentiae* does not imply that we have two methods: We have only one method—based on an ideal "knowledge space" organized by means of definitions and divisions—that, in given and concrete situations, also allows for tentative and partial bottom-up routes.

Thus conceived, the dialectical method is governed by three laws, which constitute the Ramist counterpart of the Aristotelian-Scholastic *de omni*, *per se* and *universaliter primum* principles. Ramus calls them the laws of *truth*, *justice* and *wisdom*: in the field of science every statement (i) should be valid in all its instances; (ii) should express a necessary (essential) connection of the concepts involved; (iii) should be based on subject and predicate that are proper and proportionate (allowing for simple conversion).

Ramus's logic was very influential in the second half of the sixteenth and in the first half of the seventeenth century (Feingold, Freedman, and Rother 2001). However, "pure" Ramist scholars—mostly active in the Calvinist areas of Germany, in Switzerland, in Holland, and in England—were to face an almost immediate opposition not only in Catholic but also in Lutheran universities, and saw their influence decrease after the beginning of the seventeenth century. Much more influential (and more interesting) were the many "eclectic" logicians who either tried to reconcile Ramus's and Melanchthon's logical views (Philippo-Ramists) or introduced some Ramist themes within more traditional (and even Aristotelian) contexts.

2. The Evolution of the Scholastic Tradition and the Influence of Renaissance Aristotelianism

Despite humanist criticisms, the tradition of scholastic logic not only survived during the sixteenth and seventeenth centuries but evolved in ways that are much more interesting and articulated than most modern scholars suspected until a few decades ago. Our knowledge of this evolution is still somehow fragmentary, but the scholarly work completed in recent years allows some definite conclusions. We can now say that in this evolution of the late scholastic logical tradition, six factors were particularly relevant: (i) the work of a group of Spaniards who studied in Paris at the end of the fifteenth and at the beginning of the sixteenth century and later taught in Spanish universities, influencing the development of logic in the Iberian peninsula; (ii) a renewed attention toward metaphysics, present in the Iberian second scholasticism and most notably in the works of Francisco Suárez (1548–1617), whose *Disputationes Metaphysicae* (Suárez 1965) influenced many authors all across Europe; (iii) the crucial role of the newly formed (1540) Society of Jesus, whose curriculum of studies (*Ratio Studiorum*) was to shape institutional teaching in all of Catholic Europe; (iv) the complex relations with humanism, and the influence of logicians like Agricola, whose doctrines, while taking as their starting point a humanist conception of logic, were nevertheless susceptible of somehow being absorbed or integrated within a more traditional framework; (v) the "new Aristotelianism" of authors like Jacopo Zabarella (1533–1589) and Bartholomaeus Keckermann (1572?–1609); and (vi) the renewed interest in scholastic logic, discernible in reformed Europe (and most notably in Germany) as a consequence of the doctrinal and theological conflicts with the catholic field and within the reformed field itself. In the following pages, we provide some details on this complex development.

At the end of the fifteenth century and in the first decades of the sixteenth, the Paris college of Montaigu became a center of logical research in which the late medieval logical (especially nominalist) tradition survived and to some extent flourished. A group of Spanish and Scottish logicians, lead by the Spaniard Jeronimo Pardo (d. 1505) and by the Scottish John Mair (1467/9–1550), debated themes such as the nature of supposition and signification, the distinction between categorematic and syncategorematic terms, the role of beings of reason (*entia rationis*), the nature of proposition (further developing the late medieval discussions on mental propositions), modality, and the theory of consequences. Somehow connected to this Paris group, or active there at the beginning of the sixteenth century, were the Spaniards Antonio Núñez Coronel (d. 1521), Fernando de Encinas (d. 1523), Luis Núñez Coronel (d. 1531), Juan de Celaya (1490–1558), Gaspar Lax (1487–1560), Juan Dolz (fl. 1510), the Frenchman Thomas Bricot (d. 1516), the Belgian Pierre Crockaert (Pierre of Brussels, d. 1514), and the Scot George Lokert (d. 1547).

Particularly interesting is their discussion about the nature of *complexe significabile* (propositional complex), a subject already debated by medieval logicians. The medieval defenders of this theory, associated with the name of Gregory of Rimini (c. 1300–1358), held that the object of science is not the proposition itself but what is signified by it (and determines its truth or falsity); such total and adequate meaning of the proposition is neither a physical nor a purely mental being and is not reducible to the meaning of its parts. It is rather similar to a state of affairs, which can be signified only by means of a complex (the proposition) and is therefore called *complexe significabile*. The discussion on the nature (and usefulness) of the *complexe significabile* was connected to the discussion on the role of the copula, since the copula was usually considered as the "formal" component of the proposition, "keeping together" subject and predicate. The copula was thus considered as a syncategorematic term: a term that does not possess an autonomous meaning but helps determine the meaning of the proposition as a whole. The defenders of a "strict" *complexe significabile* theory did not need a separate discussion of the mental copula, because in their opinion the *complexe significabile* is a unity and cannot be analyzed in terms of its parts. But many authors—among them John Buridan (c. 1295–1356)—assigned a much more relevant role to the copula, seen as the (syncategorematic) mental act that, in connecting subject and predicate, establishes the proposition. It is this very theory that was discussed by many of the above-mentioned late fifteenth- and early sixteenth-century Paris-based logicians (see Ashworth 1978, 1982; Muñoz Delgado 1970; Nuchelmans 1980; Pérez-Ilzarbe 1999). Pardo's position in this discussion was the most original. In his opinion, the copula is not purely syncategorematic: It is subordinate to a conceptual schema that represents something (i.e., the subject) as related in a certain way to something else (i.e., the predicate) or to itself (Nuchelmans 1980, 49). In this way the copula, while retaining its formal function, also signifies something (*aliquid*), that is, the subject, as considered in a given way (*aliqualiter*), namely as modified by the relation with its predicate. The idea of the copula signifying *aliquid aliqualiter*, and not simply *aliqualiter*, and the special relevance attributed to the subject in determining the meaning of the copula and of the proposition as a whole, were discussed, and generally criticized, by Pardo's successors. They especially investigated the role of impossible propositions, as well as propositions with a negative, privative, or impossible subject, and the problem of whether a quasi-syncategorematic nature could be attributed to the proposition as a whole.

The Iberian Peninsula was one of the strongholds of Catholicism. Moreover, as we have seen, it inherited many features (as well as textbooks and Paris-trained professors) from Parisian late scholasticism. This made the influence of the humanist movement—albeit discernible—less radical than elsewhere. Therefore, the Iberian Peninsula was the ideal context in which Catholic logicians—dwelling on the scholastic (chiefly Thomist) logical and philosophical tradition—could pursue the work of doctrinal and pedagogical systematization that was required by the struggle against the reformed field.

The Carmelite universities of Salamanca (*Salamanticenses*) and Alcalà (*Complutenses*) and the Jesuit university of Coimbra (*Conimbricenses*) each produced a complete philosophical course, including specific volumes devoted to logic. Of these the most influential was probably the Coimbra Logic, compiled by Sebastian Couto (1567–1639) but partially dependent on Pedro da Fonseca (1528–1599), who had been teacher at that university. Fonseca, the "Portuguese Aristotle," published the *Institutionum Dialecticarum Libri VIII* (Fonseca 1964) in 1564, a logical treatise built on the model of Peter of Spain and widely read throughout Europe. Fonseca's logic interprets the traditional emphasis on terms by giving a theoretical priority to the conceptual moment over the judicative one (truth and falsity are in concepts rather than in judgment) and among concepts, to singulars over abstracts and universals. To reconcile God's foreknowledge and human free will, and to handle the problem of future contingents—a theme of special interest for all Iberian philosophers—Fonseca developed, independently from Luis de Molina (1535–1600), a theory of the *scientia media*, or, as he says, of "conditioned futures," by which God foreknows all the consequences of any possible free decision.

Placing Fonseca's theories within a wider and more systematic treatment, the Coimbra logic offers a translation and a detailed commentary of Aristotle's *Organon*, which, in the form of questions, includes a discussion of most of the topics debated by sixteenth- and seventeenth-century logicians. The *Conimbricenses* reject the idea that beings of reason are the object of logic (in the scholastic tradition logical concepts such as "genus" and "species" were considered to be *entia rationis*, and the Thomist tradition considered them as the formal object of logic): Dwelling on the idea of logic as *ars disserendi*, they prefer to characterize it as a "practical science" dealing with the construction of correct arguments. Argumentation is, therefore, the first and main object of logical enquiry. Particularly interesting is the long section devoted to the nature of signs at the opening of the commentary on Aristotle's *De Interpretatione* (see Doyle 2001). The concept of sign is here taken in a broad meaning, as to include not only spoken, written, and mental "words," but also iconic languages and arithmetical signs. It is to be remarked that the influence of Coimbra logic was not limited to Europe: Jesuit missionaries used it in Latin America and even in China.

If the teaching of logic in Coimbra is connected to Fonseca, another important figure of Iberian logic and philosophy, Domingo de Soto (1494/5–1560), is connected to Alcalà and Salamanca, where he taught. Soto made important contributions to a plurality of fields, so much so that it was said *qui scit Sotum, scit totum* (who knows Soto, knows everything). Despite his endorsement of Thomism—testified by his defense of the theory that the object of logic are beings of reason—Soto was open to Scotist, nominalist, and even humanist influences, and his commentary on Aristotle's logic (Soto 1543) criticizes the "abstract sophistries" of the late scholastic logical tradition. This, however, did not prevent him from discussing and adopting many late scholastic logical theories, including large sections of medieval theories of terms. His *Summulae* (Soto 1980) are a commentary on one of the key works of medieval logic, Peter of Spain's *Tractatus* (best known as *Summulae Logicales*; see previous chapter), and include an ample discussion of signification, supposition, and consequences (see d'Ors 1981; Ashworth 1990; Di Liso 2000). Soto adopts an apparently Ciceronian definition of dialectic, considered as the art of discussing *probabiliter*. As remarked by Risse (1964–70, I, 330), however, this should not be considered a rhetorical attempt to establish apparent plausibility, but rather as an attempt to establish rational assertibility. Among the interesting points of the *Summulae* are the treatment of *induction* in terms of *ascensus* (the passage from a conjunction of singular propositions—or from a proposition with a copulative term as subject or predicate—to a universal proposition, or to a proposition with a general term as subject or predicate) and a complex square of modalities, which takes into account the quantity of the subject. Soto's discussion of second intentions offers what has been interpreted as a sophisticated theory of higher-level predicates (Hickman 1980).

One of Soto's students in Salamanca was Franciscus Toletus (1533–1596), who later taught both in Zaragoza and Rome, in the Jesuit Collegium Romanum, and was the first Jesuit to be appointed cardinal. Toletus wrote both an *Introduction* and a *Commentary* on Aristotle's logic (Toletus 1985). Like Soto, Toletus adopts some humanist theories—he takes the definition of logic as *ratio disserendi* from Boethius and divides it into invention and judgment—but his logic is actually a synthesis of Aristotelianism and Thomism, deeply influenced by the late medieval logical tradition. He considers beings of reason as formal objects of logic—thus partly endorsing the Thomist position—but maintains that logic's material object is constituted by our concepts of things and, ultimately, by things themselves, for logical beings of reason are only second intentions, based on first-order concepts—thus partly endorsing the position of Arab commentators of Aristotle (Ashworth 1985b, xli). Of special interest is his extensive use of physical and geometrical examples within the discussion of categories, and his long discussion of contingent futures within the commentary on *De Interpretatione*.

The most important Jesuit philosopher working in Spain at the end of the sixteenth century was Francisco Suárez (1548–1617). His *Disputationes Metaphysicae*, first published in 1597 (Suárez 1965), constituted a reference text and a model for further works both in the Catholic and in the Reformed fields. According to Suárez, metaphysics offers a general and unified theory of real being (*ens reale*) and of its divisions, whereas logic deals with the way of knowing and explaining such divisions. Though the *Disputationes Metaphysicae* is not a logic textbook, it discusses many issues relevant to the philosophy of logic. Suárez pays great attention to relations, subdivided into real relations (only conceptually and not really distinct from the things on which they are grounded, but nevertheless to be considered as a category of beings) and conceptual relations, which are only a product of the mind and as such do not have any ontological status. Suárez's detailed discussion of both kinds of relations helps to explain the special interest that many scholastic-oriented logicians devoted to this topic in the seventeenth century.

The last of the *Disputationes*—disputation LIV—is devoted to a subtle discussion about beings of reason (*entia rationis*) and relations of reason. According to Suárez, beings of reason are not "real" (actual or possible) beings and do not share a common concept with real beings; their only reality is that of being object of the understanding (they only have objective existence in the intellect). Therefore, they are not to be included within the proper and direct object of metaphysics. They can nevertheless be dealt with within the context of metaphysical research, given their nature of "shadows of being" (Suárez 1996, 57) and given their usefulness in many disciplines, especially logic and natural philosophy. Suarez's opinion on *entia rationis* is thus different both from that of those—like the Scotist Francis of Mayronnes (1280?–1327?)—who simply denied their existence, and from that of those—like many Thomists, including Cardinal Cajetanus (Tommaso de Vio, 1469–1534)—who thought that there is a concept common to them and to real beings. Suárez included impossible objects in the range of *entia rationis*: His discussion is thus especially relevant to the history of the logical and ontological status of impossible entities (Doyle 1987–88, 1995). The discussion on the nature of *entia rationis* was a lively one in sixteenth-century Spain and was bound to continue in Catholic Europe during most of the seventeenth century. An interesting example is that of the Polish Jesuit Martinus Smiglecius (1564–1618). In his opinion, the opposition between *ens reale* and *ens rationis* is not grounded on the fact that the *ens rationis* is not a form of being, but on the fact that it is by definition a being which is not, and cannot possibly be, an *ens reale*. A being of reason is thus, according to Smiglecius, one whose essence implies the impossibility of its real existence. The fact that *entia rationis* cannot have real existence is, according to Smiglecius, a logical and not just a physical impossibility. They, however, can have conceptual (and hence intentional) existence.

In the later Middle Ages, English logicians had been famous for their subtleties: The logical, physical, and epistemic sophisms discussed by the so-called *calculatores*, working at Merton College in Oxford, deeply influenced late fourteenth- and early fifteenth-century logic both in Paris and in Italy, and were exactly the kind of logical subtleties rejected by humanist logicians. During the fifteenth and in the first decades of the sixteenth century, however, the English logical tradition declined (see Giard 1985). This did not prevent a slow penetration of humanist ideas, testified by the 1535 statutes or the university of Cambridge, recommending the reading of Agricola and Melanchthon as substitutes for late medieval scholastic texts, and by the *Dialectica* published in 1545 by the Catholic John Seton (c. 1498–1567). The latter offers a drastically simplified treatment of traditional topics such as signification, supposition, categories, syllogism, but liberally uses nonformal arguments and literary examples, divides dialectic into invention and judgment, adopts Agricola's definition of dialectic as well as his classification of topics, and quotes, beside Cicero and Quintilian, modern humanists like Erasmus and Vives.

In the last decades of the sixteenth century, the debate on Ramism was to shake both English and continental universities. In England, Ramus found in William Temple (1555–1627) a learned defender and commentator, who, despite the strong opposition of his fellow Cambridge teacher and former master Everard Digby (1550–1592), managed to make of Cambridge, albeit for a short time, a stronghold of Ramism. The penetration of Ramism in Oxford was less substantial, and by the beginning of the new century the anti-Ramist positions were predominant in both universities. The defeat of Ramism was accompanied by the propagation of Aristotelianism—tempered by humanist-oriented attention toward classical literary examples rather than purely logical ones and toward rhetorical practices such as the *declamatio*—and by the circulation of the leading logic books published in the continent (among them Zabarella and Keckermann). The *Logicae Artis Compendium* by the Oxford professor Robert Sanderson (1587–1663; Sanderson 1985) is a good example of this new situation. Sanderson abandons the division of logic into invention and judgment, favoring a threefold division according to the three acts of the mind: The first, dealing with simple concepts, is associated with the treatment of simple terms; the second, dealing with composition and division, is associated with propositions; and the third, dealing with discourse, is associated with argumentation and method. Though this threefold division is present in the medieval and late medieval tradition and is discussed by the Conimbricenses, Zabarella, and Keckermann, Sanderson and other Oxford logicians seem to have been among the first to use it as the main division for logic textbooks (Ashworth 1985b, xli). In his logic, Sanderson includes medieval topics such as the theory of supposition and consequences, but their presentation is straightforward and not very elaborated. His discussion of method is more articulate and gives a foremost role to pedagogical concerns.

We have already mentioned the Padua professor Jacopo Zabarella, who, advocating a renewed, "pure," and philologically accurate Aristotelianism, absorbs both some humanist instances—visible in the pedagogical organization of his works and in the inclusion of Aristotle's *Rhetoric* and *Poetic* within a broad treatment of logic, on the ground of their dealing with probable arguments, such as rhetorical syllogisms and examples—and some features of the so-called Paduan Aristotelianism: a distinctive attention to the Arab interpretations of Aristotle (notably Averroes) and to Galen's concept of science. Zabarella wrote commentaries on Aristotle's logical works as well as autonomous logic tracts: Among the latter are the *De Natura Logicae* (in Zabarella 1966), the *De Methodis*, and the *De Regressu* (both in Zabarella 1985). According to Zabarella, logic deals with second intentions, that is, with the (meta)concepts produced by our intellect in reflecting on the first notions, those derived from and referring to real things. Because second intentions are products (and figments) of our intellect, logic is not a science but an instrument, or, to be more precise, an instrumental intellectual discipline, aimed at devising conceptual tools for correct reasoning and for discriminating truth and falsity. Because of its instrumental nature, logic is somehow similar to grammar: Just as grammar provides the tools needed to write and speak in an appropriate way, logic provides the tools needed to reason in an appropriate way.

According to Zabarella, order (*ordo*) and method (*methodus*) are among the main tools offered by logic: The first organizes the subject matter of a discipline and the knowledge we have acquired; the second gives the rules and procedures to be followed to acquire new knowledge, going from what we know to what we do not know (on the Renaissance concept of method, see Ong 1958; Gilbert 1960). In dealing with contemplative sciences, the *ordo* goes from the universal to the particular and to the singular ("compositive order"), while in dealing with practical and productive arts it goes from the desired effects to the principles that produce them ("resolutive order").

Although order concerns a discipline as a whole, method always has to do with the handling of specific problems, of specific "paths" going from what is known to what is unknown. Those paths are basically syllogistic demonstrations: The method is thus somehow a special case of syllogism. And since a syllogism can only go from cause to effect (compositive method or *demonstratio propter quid*) or from effect to cause (resolutive method or *demonstratio quia*), the same will hold for method. The resolutive method is used in the "hunt" for definitions, and is most needed in natural sciences; the compositive method is used in mathematics, where we start from already known, general principles and try to demonstrate all their consequences. Both methods presuppose necessary connections and are therefore only valid within contemplative sciences: Practical and productive arts, dealing with contingent truths, will have to content themselves with rhetorical and dialectical arguments, which, being only probable, are not subject to a rigorous application of method. However, even in contemplative sciences (especially in natural sciences), our knowledge of effects and of their causes is often far from clear, and we need a process of refinement, which Zabarella calls *regressus* and which involves the use of both compositive and resolutive methods: (1) We first use the resolutive method to go from a confused knowledge of the effect to a confused knowledge of its cause; (2) we then examine and clarify the knowledge of the cause (*examen*)—an activity that Zabarella connects to a specific ability of the human mind, interpreted by some modern scholars in terms of the construction of a model; and (3) we finally use the compositive method to go from a clear knowledge of the cause to a clear knowledge of the effect. This last stage is the highest sort of demonstration (*demonstratio potissima*), a notion already present in the Thomist tradition.

3. Logic in Reformed Europe: From Humanism to "Protestant Scholasticism"

It is unfortunate that most historical accounts of logic devote relatively little attention to Philipp Melanchthon (1497–1560), "Germany's teacher" (*praeceptor Germaniae*), prominent reformer and close collaborator of Luther. Actually, in the overall context of European logic in the mid-sixteenth century, the role played by Melanchthon is one of the highest significance. From 1520, when at the age of 23 he published his *Compendiaria dialectices ratio* (Melanchthon 1520), to 1560—the year of his death—there are records of more than 60 different editions of his logic works. The last version of Melanchthon's dialectics, the *Erotemata dialectics* (Melanchthon 1846), was to be the standard reference for protestant logic until the beginning of the seventeenth century.

What makes Melanchthon's logic interesting and explains its influence is above all the very evolution of his works. In the *Compendiaria Dialectices Ratio*, a young, strongly antischolastic Melanchthon offers a simplified and rhetorically oriented treatment of dialectic, purged of many "superfluous" scholastic subtleties. Like many humanist dialecticians, here Melanchthon rejects the third syllogistic figure (which he considers "remote from common sense") and the treatment of modality (the scholastic theories on modality are considered "tricky rather than true"). A few years later, however, Melanchthon's opinions on both matters (as well as on many others; see Roncaglia 1998) radically changed. In the *De Dialectica libri IV* (Melanchthon 1528) the third figure is accepted and discussed at length, and Melanchthon bitterly criticizes Valla for rejecting it, while in the *Erotemata* (Melanchthon 1846) the discussion of modal propositions is considered to be "true and perspicuous, useful in the judgment of many difficult questions." The evolution of Melanchthon's logic is thus marked by a progressive rejection of humanistic-rhetorical models and by a return to the Aristotelian and scholastic tradition.

Two further aspects of the evolution of Melanchthon's dialectic deserve attention: the gradual shift from a bipartite toward a tripartite conception of the structure of the proposition, and the growing interest in fallacies. In 1520, Melanchthon endorses the theory that every proposition has two main components: subject and predicate. In 1528, the question is seen from a grammatical perspective, and noun and verb are considered as being the two main components of the proposition. The verb, however, is further subdivided: It may be a proper verb or a construct made up of the substantive verb (the copula "est") and a noun. The copula acquires a fully autonomous role in the *Erotemata*, where every proposition is seen as having not two but three main parts: the subject, the predicate, and an (explicit or implicit) copula, seen as the formal sign of the connection between subject and predicate. Such a theory will play an important role in subsequent logic, because the copula will be considered to be not only the logical "glue" of the proposition but the actual bearer of its modal and quality modifications (see Nuchelmans 1980, 1983; Roncaglia 1996, 2003).

The discussion of fallacies also testifies Melanchthon's increasing use of scholastic doctrines. Absent in 1520, a short section on fallacies appears in 1528, accompanied, however, by the observation that anyone who has fully understood the precepts supplied for the construction of valid arguments doesn't need special rules to avoid paralogisms. But in the following editions of his dialectics, Melanchthon systematically adds new divisions and new examples. He distinguishes between fault of matter and fault of consequence, corresponding to the traditional division of fallacies *in dictione* and *extra dictionem*, and presents the principal fallacies of both sorts. In the *Erotemata*, fallacies undergo a still closer scrutiny within a systematic framework clearly derived from scholasticism.

Many elements indicate that there was one main and primary reason for this return to scholasticism: the perception that the Reformed field—engaged in the sharp debate with Catholic theologians, and in the equally sharp debate among different Reformed confessions—desperately needed effective logical tools. Rhetoric could be useful in winning popular support, but was much less effective in winning subtle theological debates. In the complex theological and political struggle that was under way in Europe, universities were to become crucially relevant players. Logic was to become a weapon in the theological struggle, and Melanchthon was probably the first to perceive that clearly.

Melanchton's works on dialectic, together with the *Dialectica* by Johannes Caesarius (1460–1550), another interesting and influential mixture of humanist and Aristotelian elements, had thus the ultimate effect of paving the way that was to be followed by Protestant logicians: endorsement of some humanist doctrine (first and foremost the pivotal role of topics and *inventio*), and great attention to the pedagogical organization of their work, but within a context that retained many tracts of traditional logic; and that—given the relevance of logic for the theological debate—was to devote a renewed attention even to some of the once deprecated scholastic subtleties.

It is therefore hardly surprising that the attempt, made by the so-called Philippo-Ramist logicians, to conjugate Ramus's drive for simplicity and for systematic, method-oriented classification, with Melanchthon's humanist-influenced but somehow more conservative treatment of logic, was not destined to have a long success. Given the renewed role of logic in the interconfessional theological debate, the two paths were bound to diverge and the Ramist component was to succumb: At the end of the sixteenth century, the anti-Ramist pamphlet was to become a well-established literary genus in the logic production of Protestant Germany. A fierce battle against Ramism was led by Cornelius Martini (1568–1621), among the founders of the so-called Protestant

Scholastic. Martini endorses Zabarella's definition of logic as mental *habitus* dealing with "second notions," that is, concepts used to represent and classify, rather than immediately derived from perception. Martini divides logic into formal and material, with formal logic seen as dealing with the pure form of (syllogistic) consequences.

Zabarella's influence is also apparent in the work of Bartholomaeus Keckermann (ca. 1571–1608). In his *Systema logicae* (1600, in Keckermann 1614), the effort to organize logic as a discipline (largely on the basis of topics) is clear from the very definition of logic, which can be considered as a human ability—and is then to be regarded as a mental *habitus*—but can also be considered as the corpus of doctrines resulting from the use of this ability (*ars externa*): that is, as a system. In this perspective, knowledge of the historical constitution of this doctrinal corpus becomes important: Therefore, it is not by chance that the short section on the history of logic, present in many sixteenth- and seventeenth-century treatises, acquires in Keckermann status, accuracy, and completeness. Keckermann's interest in the history of logic is also connected with the eclectic tendency of many early seventeenth-century logicians: Given that in a Zabarella-oriented perspective logic is a human activity (and is also the systematically arranged, historical product of this activity), it is natural to try to collect the "logical tools" developed by different logicians in different times and contexts. This eclectic tendency is usually implicit, and is not necessarily connected with the endorsement of Zabarella's positions (some aspects of which were actually criticized by many systematic-oriented or eclectic logicians), but it is clearly present in the encyclopedism of authors such as Johann Heinrich Alsted (1588–1638) or Franco Burgersdijk (1590-1636), whose *Institutionum logicarum libri duo* (1626) was the standard logic handbook in the Netherlands, and, like Keckermann, included a large section on the history of logic (see Bos and Krop 1993). A remarkable feature of this eclecticism is the tendency to reabsorb, within a context usually marked by Renaissance Aristotelianism, even some of Ramus's doctrines, notably the emphasis on the practical utility of logic and on the need of a well-arranged, easily graspable, and pedagogically oriented method.

In the first half of the seventeenth century, in reformed Europe, despite the terrible destruction of the Thirty Years War, the university system was expanding, and acquiring a political relevance that was bound to transform any doctrinal difference in the occasion of sharp conflicts (see Wollgast 1988b). This complex situation enhanced logical research and produced some new and interesting theories. In discussing the structure of the proposition, the Berlin-based Johannes Raue (1610–1679) proposed a new theory of the nature and role of the copula. In his opinion, the standard proposition of the form "S is P" should be analyzed as "that what is S is that what is P" (*id quod est S est id quod est P*), that is, as having three copulas. The role of the main copula (the middle one, which Raue calls "real copula") is then differentiated from that of the auxiliary ones: It can be used only in the present tense, while time and modal modifications are seen as operating on the auxiliary copulas.

The pronoun "id" stands for the "third common entity" (*tertium commune*) in which subject and predicate are joined, and it has been observed that "Raue delights a Fregean reader when he emphasizes that 'S', the subject, ... in his analysis is predicated of the *tertium commune* just as the predicate 'P' " (Angelelli 1990, 188). This "newest theory," of which Raue is very proud, was criticized by Johannes Scharf (1595–1660): a polemical exchange that was well known to Leibniz. Leibniz had the highest opinion of another famous logician of the time, Joachim Jungius (1587–1657). Jungius's *Logica Hamburgensis*, one of the most clear and complete logical works of the seventeenth century, deals at length with such relevant and "advanced" topics as the theory of relations and the use of nonsyllogistic consequences (Jungius 1957, 1977). Jungius is not the only one to deal with such theories, which were considered useful in theological disputations, but his treatment of them is always clear and insightful. This is especially true of his investigation of the *inversio relationis* (from "David is the father of Solomon" to "Solomon is the son of David") and of the consequence *a rectis ad obliqua* (from "the circle is a figure" to "he who draws a circle draws a figure"). Jungius's discussion of the latter—which he considers a simple consequence (the consequent is inferred from the antecedent without the need of a middle term)—was the subject of a detailed analysis in the correspondence between Leibniz and Jungius's editor Johannes Vagetius (1633–1691), who tried to offer a formal representation of its structure (see Mugnai 1992, 58–62 and 152–153).

4. Descartes and His Influence

> When I was younger I had studied, among the parts of philosophy, a little logic, and, among those of mathematics, a little geometrical analysis and algebra.... But, in examining them, I took note that, as for logic, its syllogisms and the greater part of its other teachings serve rather to explain to others the things that one knows, or even, like the art of Lull, to speak without judgment about those of which one is ignorant, than to learn them.... This was the reason why I thought that it was necessary to seek some other method, which, comprising the advantages of these three, were free from their defects. (Descartes 1994, 33–35)

This passage, from René Descartes (1596–1650) *Discours de la méthode* (1637), offers a good synthesis of Descartes's attitude toward traditional logic. Descartes's criticism of syllogism does not concern its validity but its power as a tool for scientific research, and is clearly expressed in his *Regulae ad directionem ingenii*: "dialecticians are unable to devise by their rules any syllogism which has a true conclusion, unless they already have the whole syllogism, i.e. unless they have already ascertained in advance the very truth which is deduced in that syllogism" (Descartes 1964–1976, X, 406). The core of the argument is a classic one, advanced in different forms at least since

Sextus Empiricus (see Gaukroger 1989, 6–25): Syllogism is a circular form of reasoning, since it only holds if both its premises are already known to be true, but if both premises are already known to be true, the conclusion is already known to be true, too. Therefore, to discover something new, we cannot depend on syllogism.

Pierre Gassendi (1592–1655) advanced a similar criticism. He observed that the evidence needed to accept one of the premises of a syllogism is provided or presupposed by its conclusion. Thus, in the *Barbara* syllogism "All m are p, all s are m, therefore all s are p," the truth of "all m are p" can only be established by generalization of the fact that all instances of m—including s—are p: The truth of the conclusion is presupposed by, rather than inferred from, the truth of the premises.

Descartes discusses a further argument against syllogism: The validity of a syllogism does not guarantee the truth of its conclusion, which depends on the truth of the premises. The syllogism alone—while giving us the false impression of dominating the concepts we are dealing with—cannot establish it. This argument too is fairly traditional; in the period we are dealing with, we find a similar one in Francis Bacon (1561–1626):

> We reject proofs by syllogism, because it operates in confusion and lets nature slip out of our hands. For although no one could doubt that things which agree in a middle term, agree also with each other (which has a kind of mathematical certainty), nevertheless there is a kind of underlying fraud here, in that a syllogism consists of propositions, and propositions consist of words, and words are counters and signs of notions. And therefore if the very notions of the mind (which are like the soul of words, and the basis of every such structure and fabric) are badly or carelessly abstracted from things, and are vague and not defined with sufficiently clear outlines, and thus deficient in many ways, everything falls to pieces. (Bacon 2000, 16)

While in Bacon this argument is used to advocate the need of "true induction" (progressive generalization accompanied by the use of his "tables of comparative instances"), in Descartes it is used to advocate the role of intuition. According to Descartes, the process of knowledge acquisition depends on (1) intellectual intuition, that is, the intellectual faculty that allows a clear, distinct, immediate, and indubitable grasp of simple truths; and (2) deduction, that is, the grasp of a connection or relation between a series of truths. According to Descartes, deduction is therefore not to be seen as an inferential process governed by logical rules, but rather as the exercise of an intellectual faculty that is ultimately based on intuition. The process of mastering a long or complex deduction is a sort of intellectual exercise, consisting in the recursive application of intuition over each of its steps. The aim of this process is certainty: The idea of degrees of certitude or probability is totally alien to Descartes's intuition-based conception. And since for Descartes intellectual intuition is a natural faculty, there are no abstract rules or inference patterns governing intuition or deduction: We can only give precepts—like the well-known four *regulae* given in the *Discours*—helping us in the better use of this faculty.

One aspect of Descartes's concept should be stressed: The combined use of intuition and deduction allows us to attain knowledge, but does not suffice by itself to guarantee that the knowledge we attain is true. If a proposition p is intuitively clear and evident for us, we are entitled to *claim* that it is true. But while this claim is justified, its correctness is not grounded on the fact that p is perceived by us as clear and evident, because a deceptive God could give us a clear and distinct intuition of something that is not true (i.e., something that does not correspond to reality; from this point of view, Descartes is now generally considered as holding a correspondence theory of truth; see Gaukroger 1989, 66). Therefore, the well-known *cogito* argument is needed to ensure God's external guarantee of our knowledge. According to Descartes, we only need (and we can only achieve) this external guarantee: God's knowledge is not a model for our knowledge, and there is no set of eternal truths binding God's knowledge and ours in the same way, since eternal truths themselves result from the joint action (or rather from the unified action) of God's will and understanding.

Descartes's resort to intellectual intuition as ultimate foundation of certainty is somehow at odds with his work in the field of algebra and geometry and with his discussion on the relevance of analysis. In Descartes's opinion, analysis is associated with the discovery of new truths (while synthesis has to do with presenting them in such a way as to compel assent), and its function is apparent in mathematics and in analytical geometry, when we use variables (general magnitudes) instead of particular values. Descartes, however, doesn't seem to perceive the possible connection of this method with deductive reasoning: on the contrary, he seems to associate deduction with the less imaginative, painstaking word of synthetically computing individual magnitudes.

The influence of scholasticism on Descartes's philosophy is greater than one might suspect at first sight (see already Gilson 1913). For instance, hints at a "facultative" concept of logic (see section 6) were present in authors (among them Toletus, Fonseca, and the Conimbricenses) he knew. But Descartes's concept of deduction differed very much from traditional logic. This did not prevent some Cartesian-Scholastic logicians to reconcile them. Johann Clauberg (1622–1665), in his aptly named *Logica vetus et nova* (Clauberg 1658), defended Descartes's methodical rules against the charge of being too general or useless, attributing them the same kind of rigor and strength of Aristotle's logical rules. Johann Christoph Sturm (1635–1703) made a similar attempt.

More articulate was the position of Arnold Geulincx (1624–1669), who published a *Logica Fundamentis suis restituta* (1662), and a logic *more geometrico demonstrata*, the *Methodus inveniendi argumenta* (1663; both in Geulincx 1891–1893). Like most Dutch logicians, Geulincx was deeply influenced by the eclectic Aristotelianism of Burgersdijk (see section 3). He thus merged late scholastic, Aristotelian, and Cartesian themes in a logic that, with some hyperbole, he labelled "geometric." Its treatment includes the so-called De Morgan rules (well known in medieval scholastic logic, but less frequently dealt with by Renaissance logicians). He also devised a "logical cube" whose faces represented all the axioms and argument forms of his logical system.

Descartes's influence is also evident in the *Port Royal Logic*, which we will discuss in the next section. Before dealing with it, however, there are two authors—somehow difficult to classify by means of traditional historiographic labels—which are worth mentioning: the French Jesuit Honoré Fabri (1607–1688) and the Italian Jesuit Gerolamo Saccheri (1667–1733). Neither of them was an "academic" logician, and they both had wide-ranging interests. Fabri corresponded with most of the major philosophers and scientists of the time (including Descartes, Gassendi, and Leibniz); was interested in philosophy, mathematics, astronomy (he discovered the Andromeda nebula), physics, and biology; and wrote on calculus and probabilism (his book on this subject was condemned by the Church). His *Philosophia* (1646) is influenced by Descartes, but the section on logic is pretty original: He developed a combinatorial calculus which allowed him to classify 576 syllogistic moods in all the four figures; he also used a three-valued logic (based on truth, falsity, and partial falsity) which he applied to the premises and conclusions of syllogisms, and used disjunctions to express hypothetical judgments.

Wide-ranging were also the interests of Saccheri, who, besides working on logic, also wrote on mathematics and geometry. In trying to prove the parallel lines postulate, in the *Euclides ab Omni Naevo Vindicatus* (1733) he hints—against his will—to non-Euclidean geometries. Both in logic and in geometry he makes use of the *consequentia mirabilis* (well known to the mathematicians of the time): If p can be deduced from non-p, then p is true. In his *Logica demonstrativa* (published anonymously in 1697)—a treatise on logic organized "more geometrico"—he applies the *consequentia mirabilis* to syllogistic. One of his proofs refers to the rejection of AEE syllogism in the first figure. Saccheri shows that this very rejection (stated in E-form: "no AEE syllogism in first figure is valid") can be the conclusion of a first figure AEE syllogism with true premises. If such a syllogism is not valid, then it constitutes a counterexample to the universal validity of AEE syllogisms (which are thus to be rejected). If it is valid, the truth of its premises implies the truth of its conclusion. Such an elegant demonstration has been correctly seen as the mark of an argumentation strategy based on the skillful use of confutations and dilemmas (see Nuchelmans 1991, 133–137).

5. The *Port-Royal Logic*

A mixture of ancient and new doctrines characterizes the *Logique ou l'Art de penser* published anonymously in 1662 but written by Antoine Arnauld (1612–1694) and Pierre Nicole (1625–1695). The authors belonged to the Jansenist movement of Port-Royal, hence the current denomination of their work as the *Port-Royal Logic* (Arnauld and Nicole [1683]). According to widespread opinion, the authors endorse Descartes's philosophy. This is in many respects true, especially as regards the origin of ideas and the account of the scientific method. Indeed, in the 1664 edition of the book, the authors declare that the section on the analytic and synthetic method is based on the manuscript of Descartes's *Regulae ad directionem ingenii*. However, the *Port-Royal Logic* is not a straightforward Cartesian logic because it relies on many sources. Apart from the influence of Augustine and Pascal (1623–1662), the authors, though condemning scholastic subtleties, acknowledge the utility of some scholastic precepts and are not always adverse to Aristotle. True, they reject the Aristotelian categories and topics, but describe these doctrines and make them known to their readers.

The *Port-Royal Logic* is also different from a humanistic *ars disserendi*, and even more from an *ars bene disserendi*, as Ramus would have it, for it is intended to be an *ars cogitandi*, an art for thinking. The authors maintain that, since "common sense is not so common a quality as people think" (First Discourse 17, trans. 6), people ought to educate themselves to be just, fair, and judicious in their speech and practical conduct. Such an education should be offered by logic, but traditional logic pays too much attention to inference, whereas it should concentrate on judgment because it is in judging that we are liable to make errors compromising our rational and practical conduct. So, because judgment is a comparison of ideas, a detailed study of ideas must precede it.

In the *Port-Royal Logic*, "idea" is an undefined term: "The word "idea" is one of those that are so clear that they cannot be explained by others, because none is more clear and simple" (I, i, 40, trans. 25). "Idea," therefore, is a primitive term that can only be described negatively. Accordingly, the authors maintain that ideas are neither visual images nor mere names, and are not derived from the senses, because, although the senses may give occasion to forming ideas, it is only our spirit that produces them. Once ideas are produced, logic investigates their possible relations and the operations one can perform on them. Such relations and operations are founded on a basic property of universal or common ideas (as different from singular ideas): the property to have a comprehension and an extension.

The *comprehension* of an idea consists of "the attributes that it contains in itself, and that cannot be removed without destroying the idea. For example, the comprehension of the idea of a triangle contains extension, shape, three lines, three angles, and the equality of these three angles to two right angles, etc." (I, vi, 59, trans. 39). The fact that the comprehension of "triangle" contains not only three lines but also the property proved by the theorem that the sum of its angles is equal to two right angles, gives way to speculations as to what extent humans dominate the comprehensions of their own ideas (Pariente 1985, 248ff). The *extension* of an idea "are the subjects to which the idea applies. These are also called the inferiors of a general term, which is superior with respect to them" (I, vi, 59, trans. 40). The notion of extension is ambiguous. The subjects to which an idea applies can be intended either as the class of individuals of whom that idea can be predicated, or as the ideas in whose comprehension that idea is contained. In the latter sense, extension is clearly defined in terms of comprehension; therefore, comprehension is considered a primitive notion in comparison with extension. The ambiguity of the notion of extension, already noticed by some interpreters (Kneale and Kneale 1962, 318–319), is somehow intended, for it serves, as we will see, to define different properties of the operations that can be performed on ideas, as well as to solve classical problems of quantification in the doctrine of judgment and reasoning.

If we subtract an attribute from the comprehension of an idea, by definition we obtain a different idea, in particular a more abstract one: Given the idea "man," whose comprehension certainly contains "animal, rational," by subtracting "rational" we destroy the idea "man" and obtain a different and more abstract idea, which is the residual idea with respect to the original comprehension of "man." The operation just described is *abstraction* that, if reiterated, produces an ascending hierarchy of increasingly abstract ideas.

The operation of abstraction is the means by which the *Port-Royal Logic* introduces an *inverse relation* between comprehension and extension, often called the "Port-Royal Law" in subsequent literature. For the authors maintain that in abstractions "it is clear that the lower degree includes the higher degree along with some particular determination, just as the I includes that which thinks, the equilateral triangle includes the triangle, and the triangle the straight-lined figure. But since the higher degree is less determinate, it can represent more things" (I, v, 57, trans. 38). This means that the smaller the comprehension, the larger the extension, and vice versa.

To move from a higher to a lower idea in the hierarchy we have to *restrict* the higher one. Restriction can be of two kinds. The first kind of restriction is obtained by adding a different and determined idea to a given one: If to the idea A (animal) we add the idea C (rational), so as to have a new idea composed by the joint comprehensions of A and C, and if we call B (man) the idea thus composed, then B is a restriction of A and C and is subordinated to them in the hierarchy of ideas. This restriction cannot be obtained by adding to an idea some idea it already contains in itself, for the alleged restriction would be a mere *explication* of the given idea: If B contains A, then by adding A to B we obtain B, that is $BA = B$. The second kind of restriction consists in adding to a given idea "an indistinct and indeterminate idea of part, as when I say 'some triangle.' In that case, the common term is said to become particular because it now extends only to a part of the subjects to which it formerly extended, without, however, the part to which it is narrowed being determined" (I, vi, 59, trans. 40). The possibility of two kinds of restriction shows that comprehension and extension do not enjoy the same properties. While the first restriction modifies the comprehension of the restricted idea so that we get a different idea having a richer comprehension and a smaller extension, the second restriction concerns only the extension of the restricted idea with no modification of its comprehension: It remains the same idea. But this depends on the fact that the notion of extension of the *Port-Royal Logic* is, as already noticed, ambiguous. In this second case, the extension is obviously constituted by the individuals to whom the idea applies.

The *Port-Royal Logic*, also due to Augustine's influence, is very attentive to the linguistic expression of ideas. The authors maintain that

> if reflections on our thought never concerned anyone but ourselves, it would be enough to examine them in themselves, unclothed in words or other signs. But we can make our thoughts known to others only by associating them to external signs, and since this habit is so strong that even in solitary thought things are presented to the mind by means of the words we use in speaking to others, logic must examine how ideas are joined to words and words to ideas. (untitled preface 38, trans. 23–24)

This means that thought is prior to language and that a single thought can underlie different linguistic forms. This view of the relation between thought and language is one of the guidelines behind the project of a universal grammar contained in the *Grammaire générale et raisonnée*, published in 1660 by Arnauld in collaboration with Claude Lancelot (1616–1695) (Arnauld and Lancelot [1676]). This view, which has received great attention since Noam Chomsky's (1966) much-debated claim that it prefigures transformational generative grammar, is also relevant to logic (see Dominicy 1984).

Given that logic studies the properties of ideas, their mutual relations, and the operations that can be performed on them, and given that ideas are designated by words which can be equivocal, the authors establish the convention that, at least in logic, they will treat only general or universal ideas (as different from singular ideas) and univocal words (I, vi, 58, trans. 39). Such are the words associated to ideas by way of a *nominal definition*, meant as the imposition of a name to an idea by way of a free, public, and binding baptismal ceremony, of the kind used in mathematics and whose model is found in Pascal ([1658 or 1659], 242ff). Nominal definitions make it possible to use words (particularly *substantives*) of ordinary language as if they were the signs of a formal language in which everything is explicit:

> The best way to avoid the confusion in words encountered in ordinary language is to create a new language and new words that are connected only to the ideas we want them to represent. But in order to do that it is not necessary to create new sounds, because we can avail ourselves of those already in use, viewing them as if they had no meaning. Then we can give them the meaning we want them to have, designating the idea we want them to express by other simple words that are not at all equivocal. (I, xii, 86, trans. 60)

We can not only make abstractions and compositions of ideas but also *compare* them and, "finding that some belong together and others do not, we unite or separate them. This is called *affirming* or *denying*, and in general *judging*" (II, iii, 113, trans. 82). Since judging is a comparison of ideas, and since the activity of *making syllogistic inferences* can be considered as a comparison of two ideas through a third one, syllogistic inferences lose much of their importance, while judging is established as the most important of our logical activities (Nuchelmans 1983, 70–87). This does not mean that the *Port-Royal Logic* neglects syllogisms. On the contrary, it contains an articulate doctrine of syllogism based on the *fundamental principle* that, given two propositions as premises, "one of the two propositions must contain the conclusion, and the other must show that it contains it" (III, xi, 214, trans. 165). It also contains a nontrivial treatment of syllogistic moods that was to be implemented by the young Leibniz in his *De Arte Combinatoria* (see section 8). Such a treatment is based on "the law of combinations," applied to "four terms (such as A.E.I.O.)" giving 64 possible moods, and on a set of rules that make it possible to select the well-formed ones, so that, given rules for the valid moods in each figure, one can dispense with the doctrine of the reduction of other figures to the first: Each mood of any figure is proved valid by itself (III, iv, 88–89, trans. 143–144). Indeed, though the first edition of the book contained a chapter on the reduction to the first figure, that chapter is left out of all subsequent editions. Nevertheless, though the authors seem competent in pointing out a frequent confusion between the fourth figure and the first figure with transposed premises (III, viii, 202, trans. 155), they are very traditional in other respects. For instance, they reduce the *consequentiae asyllogisticae*, called complex and composed syllogisms, to the classical categorical moods, thus provoking a reproach from Vagetius in the preface to the second edition of Jungius's *Logica Hamburgensis* (see section 3).

The *Port-Royal Logic* attributes great importance to method, corresponding to the operation of the spirit called "ordering." Method is divided into two major sections: The first, devoted to demonstration and science, follows Descartes's methodical rules, thus giving an outline of the methods of analysis and synthesis; the second, devoted to opinion and belief, contains interesting observations about epistemic modalities and probability, as well as the outlines of Pascal's wager on the existence of God. By introducing the question of probability, the *Port-Royal Logic* breaks away from one of the major tenets of Descartes's philosophy, and opens new perspectives for a probability not limited to games of chance but extended also to events valuable on the basis of frequencies (Hacking 1975b).

The *Port-Royal Logic* was highly successful, as can be gathered from its numerous editions (Auroux 1993, 87). Its influence, also thanks to Latin, English, and Spanish translations (Risse 1964–70, II 79), is apparent in most of the subsequent European logical literature. Obviously some scholars still preferred the Aristotelian model. For instance, John Wallis (1616–1703), one of the best mathematicians of the time, though the *Port-Royal Logic* agreed with the argument he had already produced in 1638 that in syllogisms singular propositions must be considered as universal, rejects the Ramist syllogistic moods (Wallis 1687) on Aristotle's authority alone. A direct reaction to the *Port-Royal Logic* is presented by Henry Aldrich (1648–1710), whose *Artis Logicae Compendium* (Aldrich [1691]), published anonymously, reprinted many times and still widely used in the Victorian era, preserves scholastic doctrines and an account of the syllogism which "is the best available" (Ashworth 1974, 237). In some concluding remarks, Aldrich criticizes the fundamental principle of syllogism of the *Port-Royal Logic*, which he considers as a disguised version of the *dictum de omni et nullo*. Yet the very existence of such a criticism implicitly proves the fame the *Port-Royal Logic* had achieved (Howell 1971, 54–56). As for Port-Royal's seminal theory of probability, the young Leibniz already acknowledged its merits in 1667 (Leibniz 1923–, VI, i, 281n), while Jakob Bernoulli wrote his *Ars conjectandi*, published posthumously (Bernoulli [1713]), as a development of that theory and as a complement to the *Ars cogitandi*, the Latin version of the *Art de penser* (Hacking 1975b, 78; Daston 1988, 49).

6. The Emergence of a Logic of Cognitive Faculties

Toward the end of the seventeenth century, many logicians developed an interest in the analysis of cognitive faculties. Descartes moved in that direction when he focused his attention on the operations of intuition and deduction, but also the *Port-Royal Logic* considered the reflection on the nature of the mind as the means for a better use of reason and for avoiding errors. The study of cognitive faculties was not simply meant to provide an expositive framework for logical doctrines. As a matter of fact, dealing with the nature and object of logic and with the justification of the traditional partitions of logical treaties through a reference to mental operations had been a well-established practice since Aristotle's *Organon*: The operations of simple apprehension, judgment, and reasoning had been mentioned as mental counterparts to the logical doctrines of concepts, judgments, and inferences (for a similar approach see Sanderson, section 2). The novelty of what has been called the "facultative logic" of the late seventeenth and eighteenth centuries (Buickerood 1985) is that the cognitive operations involved in the formation and use of ideas become a central concern of logicians.

The most important author working on a logic of cognitive faculties is John Locke (1632–1704). The *Essay concerning Human Understanding* (Locke 1690) is often quoted as a primary example of indifference, if not contempt, for logic. This is not true if it is intended to describe Locke's attitude to logic in general, rather than his attitude toward the doctrine of syllogism. Locke, who had been provided at Oxford with a sound scholastic logical education (Ashworth 1980), asks logicians to give up their claim that they teach humans, who are born rational, how to reason by means of syllogisms (*Essay*, IV, xvii, 4). Logical research should rather investigate the way we form, designate, combine and, in general, use ideas.

This concept of logic attributes fundamental importance to language. If logic is the study of the faculties that produce and work with ideas, then it becomes impossible to ignore that we can work with ideas only in so much as we connect them to linguistic signs. This very connection of ideas to linguistic signs, although very useful, is a main source of our errors; therefore logic must provide a detailed treatment of the use and of the abuse of words. Indeed Locke compares logic to semiotic: "σημειωτική, or *the Doctrine of signs*, the most usual whereof being Words, ... is aptly enough termed also λογική, Logick" (*Essay*, IV, xxii, 4). Locke does not intend to privilege the study of words over that of ideas, for he considers both ideas and words as signs, with the difference that ideas are signs of things and words are signs of ideas (see Ashworth 1984). Nevertheless it cannot be doubted that the success of Locke's philosophy contributed to the extraordinary importance language was to have for a large section of later philosophy and logic (see Hacking 1975a).

One could say that a logic centered on the study of faculties that produce ideas includes a good deal of epistemological, psychological, and linguistic research (see Hatfield 1997). This is true, but one should consider that this is a consequence of a double effort. On the one hand, Locke wanted to continue the old battle against the ontological basis of the Aristotelian logical tradition by eliminating all talk about natural essences. On the other hand, he wanted to win the new battle against Cartesian innatism (accepted by the *Port-Royal Logic*) by investigating the empirical origin of our thought. A study of such questions and of the correct use of our ideas and their signs could provide a guide to man's intellectual conduct in the exercise of judgment, especially in an age characterized by a strong skeptical movement. But providing men with a guide in making judgments was seen as the purpose of logic: In this respect, Locke had many predecessors, notably the authors of the *Port-Royal Logic*. It must also be considered that it was still left to logic, once the observation of cognitive operations was completed and a careful reflection on them was performed, to establish *norms* for the correct use of those very operations.

Defenders of the Aristotelian tradition, such as John Sergeant (1622–1707), criticized Locke's approach to logic (see Howell 1971, 61–71). But many more were Locke's admirers, and the impact of his views was impressive. Particularly successful was his refusal of innate ideas (*Essay*, I, ii), and his emphasis on probable knowledge, though he was far from considering probability from a quantitative point of view (see Hacking 1975b, 86–87). What interests us is that from a very early stage, Locke's doctrines were included in logic textbooks, frequently in association with direct or indirect references to the *Port-Royal Logic*. This seems strange if one considers that Locke and the Port-Royalists were so far apart on the question of the origin of ideas. But this important difference was overlooked by taking into account what Locke and the *Port-Royal Logic* had in common: the purpose of logic as the guide for correct judgment, the idea that logicians should reflect on human understanding, the importance of the linguistic expression of our thoughts. Moreover, the *Port-Royal Logic*, whose intended readers included the students of the *petites écoles* of Port-Royal, who had to be prepared for the curricula of European universities, was a good source of information about traditional logical topics (such as syllogisms, fallacies, method) that the *Essay* lacked and that the Port-Royalists had treated without deference toward Aristotle.

Perhaps the first of such logic handbooks was the *Logica sive ars ratiocinandi* (Le Clerc 1692) written by the Swiss Jean Le Clerc (1657–1736), who adopted Locke's doctrine of judgment, his classification of ideas, and his philosophy of language, and proposed a mixture of Locke's and *Port-Royal Logic*'s theses as concerns the question of probability. A similar approach to logic is to be found in another Swiss scholar, Jean-Pierre Crousaz (1663–1750), who published a series of books on logic, the latest of which was *Logicae systema* (Crousaz 1724). As it can be expected, a number of British authors published Locke-oriented logic handbooks, most of which had several editions in Britain and elsewhere. Isaac Watts (1674–1748), in his *Logick* (Watts 1725) and in a popular supplement to it (Watts 1741), followed Crousaz and Le Clerc in offering a Lockean analysis of human nature (especially perception) with a preference for judgment and proposition, rather than syllogism. William Duncan (1717–1760), in his *Elements of Logick* (Duncan 1748), stated that the object of logic is the study of the faculties of the human understanding and of cognitive procedures, and divided logic into four parts according to the model of the *Port-Royal Logic* (see Yolton 1986). Francis Hutcheson (1694–1746) too, in his posthumous *Logicae compendium* (Hutcheson 1756), considered logic as the study of cognitive faculties, but also offered a kind of axiomatic presentation of syllogistic.

The *Essay* was promptly translated into Latin and French. By 1770, professors of Prussian universities were officially asked to follow Locke in their lectures on metaphysics (von Harnack [1900], I, i, 373). But much before that date Georg Friedrich Meier (1718–1777), the author of the text adopted by Kant for his logic courses (see section 11), already used the *Essay* in his lectures. The reception of the *Essay* as a book of logical content was made easier by the inclusion of Locke's doctrines in logic handbooks such as those we have mentioned. But it was left to Locke's posthumous *Of the Conduct of the Understanding* (Locke [1706] 1993), originally intended as an additional chapter to the fourth edition of the *Essay*, to enter directly into the field of logical literature. For in this small book, Locke explicitly declares his views to be an improvement over the standard logic of his time (Buickerood 1985, 183). *Of the Conduct of the Understanding* was widely read not only in Locke's own country (Howell 1971, 275ff.) but also in Germany. In the second decade of the eighteenth century, the Thomasian philosopher Johann Jakob Syrbius (1674–1738) used it as a guide for his lectures, and Wolff referred to it in his *German Logic* (Wolff [1713], Preface). Later on Georg David Kypke translated it into German (Locke [1755]), but at the time of this translation Locke had already been favorably discussed in the incipient German literature on the history of logic (Budde 1731).

In France, Locke's views were well received by authors who also supported Port-Royal's doctrines. This is the case of Jean Baptiste d'Argens (1704–1771) who, in the section devoted to logic of a philosophical treatise addressed to ladies, repeated the Lockean argument that all ideas originate from the senses, at the same time referring to the *Port-Royal Logic* (d'Argens 1737, Log. §§3, 1) in very positive terms. Among French scholars, Étienne Bonnot de Condillac (1714–1780), the influential promoter of a radically empiricist philosophy usually referred to as "sensationism," deserves a special mention. In works of logical content written in the later part of his life—*La Logique* (Condillac [1780]) and the posthumous *La langue des calculs* (Condillac [1798])—he developed a concept of logic that owes much to Locke, although he proudly maintains that it is similar to no one else's. Condillac describes logic as consisting of an analysis of experience by which we study both the origin of ideas and the origin of our own faculties (Condillac [1780], Preface). For, differently from Locke, who admitted sensation and reflection as sources of our ideas, Condillac admits sensation only and maintains that not only ideas, but also all our faculties originate from the use of the senses. This is possible because our senses are complemented by our fundamental linguistic capacity: We would not have complex ideas nor would we be capable of operating with them if we had no language. Consequently, Condillac maintains that any science is a well-made language and such is also the art of reasoning, a conviction that made him reverse the priority order of grammar and logic (Auroux 1993, 93). He adds that, to build a well-made language, we need a method because language, despite its role in the acquisition of our cognitive faculties, has also been used to produce a jargon for false philosophies. The method Condillac recommends for the construction of a well-made language in any science is analysis, in particular analysis as it is used in mathematics, that he considers the paradigm of a well-made science whose language is algebra.

7. Logic and Mathematics in the Late Seventeenth Century

At the end of the seventeenth century, another image of the discipline began to emerge. It was borne out of a comparison of logic with mathematics, a comparison intended to prove the superiority of mathematics over logic.

Some authors ascribed the superiority of mathematics to its axiomatic-deductive method. This conviction had enjoyed considerable success (see De Angelis 1964), and was enhanced by an interpretation of Descartes's rules of method as recommendations to begin with a few simple and already known notions (axioms) and then proceed to unknown notions (theorems). Those who endorsed this interpretation seemed to draw the conclusion that the old logic centered on syllogism should be replaced by a new logic intended as a method to find and order truths according to the model of the *mos geometricus*.

Some authors attributed the superiority of mathematics to the problem-solving and inventive techniques of algebra. In this perspective, the search for equations relating unknown to given elements, exemplified in Descartes's

Géométrie (1637), was interpreted as the true Cartesian logic and was absorbed into the tradition that viewed mathematics as a universal science of invention. In the seventeenth century, algebra was still a new technique, independent of logic, that many considered a rediscovery, but also an improvement of ancient mathematical doctrines. A reference to Descartes was somehow inevitable in this field, too, since Descartes appreciated algebra as an intermediate step toward his more abstract *mathesis universalis*. (On the origins of algebraic thought in the seventeenth century, see Mahoney 1980.)

Naturally enough, some authors suggested that algebra could be a useful model for logic. This is the case of Ehrenfried Walter von Tschirnhaus (1651–1708). He left syllogisms and other traditional parts of logical treatises out of his logic, but, as the title of his major work declares, he held that logic must provide a *Medicina mentis*, a remedy against the illness represented by our errors, and an aid for the healthy art of invention (Tschirnhaus [1695]). In particular, he claimed that his method of invention would have, in all fields of knowledge, the same function of algebra in the mathematical sciences. What he actually did, however, was to give an exposition of the methods of analysis and synthesis and a comment on Descartes's rules of method, albeit with a new attention to empirical sciences for which he envisaged a mixed method of a priori and a posteriori elements (Wollgast 1988a).

The assessment of the positive role that algebra could have for logic outlived the idea that logic should imitate, or even be substituted by, the axiomatic-deductive method. The latter was either reduced to a mere synthetic (top-down) order of exposition, or was declared inadmissible outside mathematics, either because of intrinsic differences between mathematics and other sciences (and philosophy), or because it was held responsible for the degeneration of Cartesianism into Spinozism. But also the algebraic model underwent profound changes. For the algebraic model, followed by many logicians from the late seventeenth century up to almost the end of the eighteenth century, is not the same as the algebraic model used in problem solving. Algebra is no longer considered as a methodical paradigm to be followed analogically by logic to restore the latter's function as an intellectual guide but as a tool for logic. Many logicians now try to apply algebraic techniques directly to *logical objects*, that is, to ideas and propositions. In other words, they try to build a logical calculus based on a symbolic representation of logical objects and on rules for manipulating signs, on the assumption that an adequate symbolism has been used.

From this point of view, doctrines of ideas such as those of the *Port-Royal Logic* and of the emergent logic of cognitive faculties, usually considered extraneous to the development of mathematically oriented logic, instead acted as stimuli and provided a field of application for the first tentative logical calculi. On the one hand, as it has been pointed out (Auroux 1993, 94), a calculus of ideas needs a theory of ideas. On the other hand, scholars who had welcomed a logic of cognitive faculties professed the highest esteem for algebra. We have mentioned Condillac's positive reference to algebra as the language of mathematics, but decades earlier Nicolas Malebranche (1638–1715) had claimed that "algebra is the true logic" (Malebranche [1674] 1962, VI, i, v). Also Locke, who belittled syllogistic and the axiomatic method, did not hide his admiration for algebra (*Essay*, IV, xii, 15). Locke, however, did not even think of applying the powerful tool of algebra to ideas. Nor did Thomas Hobbes (1588–1679), although in his *De corpore*, published in 1655, whose first part is significantly entitled *Computatio sive Logica*, he maintained that reasoning is computation, where computing means adding several things or subtracting one thing from another in order to know the rest (Hobbes [1839] 1961, I, i, §2). But other scholars were ready to attempt the actual construction of logical calculi.

We examine some of such attempts, but first consider a declared failure to establish a parallelism between logical and algebraic reasoning, that is, the *Parallelismus ratiocinii logici et algebraici* (Bernoulli [1685] 1969) of the above-mentioned Jakob Bernoulli (1655–1705). This is an academic dissertation in which Bernoulli was the *Praeses* (and therefore the real author) and his younger brother Johann (1667–1748) was the *Respondens*, a circumstance that has often brought to the attribution of the work to the "Bernoulli brothers." The parallelism concerns the objects, the signs, and the operations of both logic and mathematics.

The objects of logic are *ideas of things*, while the objects of mathematics are *ideas of quantity*. Likewise, the signs of ideas of things are words ("man," "horse"), while the signs of quantity are letters of the alphabet: a, b, c for known quantities, and x, y, z for unknown quantities. Bernoulli does not use literal symbols for ideas of things because, on a par with the *Port-Royal Logic*, he assumes that every idea of thing is (at least in logic) univocally designated by a word, so that every idea of thing has its own sign.

Bernoulli then introduces the operations we perform on both kinds of ideas: (1) to put together, (2) to take away, (3) to compare.

1. Ideas of quantity are put together by the sign "+", as in "$a + b$." Ideas of things are put together by the connective "and", as in "virtue and erudition."

2. From an idea of quantity one can take away a smaller quantity, thus obtaining the difference: Given a and b, where a is greater than b, the taking away of b from a is denoted by "$a - b$." Similarly, from a complex idea of thing, one can take away one of the less complex ideas it contains, thus obtaining the difference: From the complex idea "man" one can take away "animal" and the difference is "rational."

3. Given two ideas of quantity, if the mind perceives an equality between them, it unites them by the sign "=", as in "$a = b$." If the mind perceives an inequality between them, it uses the signs ">" and "<", as in "$a > b$," "$a < b$." Given two ideas of things, the mind can find (1) agreement or identity between them, and in this case it will affirm one of the other; (2) disagreement or diversity between them, and in this case it will deny one of the other. Affirmation and negation take place thanks to an enunciation (*enunciatio*), and are expressed by "it is" (*est*) and "it is not" (*non est*), as in "man is animal," "man is not brute."

While there is a parallelism between algebra and logic with respect to the operations of putting together and taking away, the parallelism breaks down in the case of comparison. Bernoulli first considers the case of agreement. Two ideas of quantity agree when a common measure, applied to them the same number of times, exhausts both, that is, when they are equal. Two ideas of things agree, so that it is possible to affirm a predicate of a subject, provided that a third idea, common to both, exhausts at least the predicate: It is possible to affirm "theft is sin," provided that some common idea is found in "theft" (no matter if "theft" contains some other ideas besides) and exhausts "sin."

It is clear that Bernoulli intends the comparison of ideas of things as the comparison of their comprehensions. This is confirmed by the fact that "theft is sin" is an indefinite proposition, that is, a proposition whose subject is not quantified. Now, Bernoulli explains, in an indefinite proposition the predicate is found *in the nature* of the subject, which means that it cannot be taken away from the idea of the subject in which it is contained without destroying it, according to the *Port-Royal Logic* definition of comprehension.

We have seen that for Bernoulli the agreement or identity of subject and predicate subsists even if it is incomplete, that is, if the third idea common to both exhausts the predicate without exhausting the subject. Here a problem arises: While the equality of quantities is mutual (if $a = b$, then $b = a$), an affirmative indefinite proposition expressing an incomplete agreement is not convertible: "Man is rational" is true, but "Rational is man" is false. Moreover, what happens if the predicate is not found in the nature of the subject but is constituted by some accidental attribute? The answer is that it would be impossible to establish even a partial agreement and, strictly speaking, it would be impossible to affirm that predicate of the subject. Bernoulli decides to overcome these problems by quantifying over the subjects, that is, by taking extensions into account. In this way it becomes possible to form true affirmative propositions such as "All men are sinners" and "Some men are learned," that is, propositions that in the indefinite form ("Man is sinner" and "Man is learned") are false. The proposition "All men are sinners" is particularly interesting because it is a true universal proposition although "sinner" is not found in the nature of "man" (Jesus is [also] man, but is not sinner). Therefore, Bernoulli states that the subject of universal and particular propositions are "the species or the individuals that are contained under that [subject]" (Bernoulli [1685] 1969, §11, trans. 176).

Is it practical to consider which are the essential attributes of the subject and which are the accidental ones and, in the first case, be allowed to make indefinite judgments while, in the second case, resort to quantified ones? And how to overcome the problem of the impossibility of converting true indefinite propositions, which are exactly those that most resemble algebraic equations? Bernoulli suggests that one should always quantify all affirmative propositions, including true indefinite ones. Consequently, one will be allowed to say "Some men are learned," which can be converted *simpliciter* into "Some learned beings are men," as well as "All men are sinners" and "All men are rational," which can be both converted *per accidens* into “Some sinners are men” and “Some rational beings are men.”

Differences between algebra and logic also appear when a comparison of ideas shows their disagreement. In algebra, the disagreement of two ideas of quantity means that between them there is an inequality, a relation designated by the sign “$<$” that is perfectly convertible: If $a < b$, then $b > a$. In logic, the disagreement of two ideas of things is expressed by a negative indefinite proposition. But the subject and the predicate of negative indefinite propositions can be converted only if the disagreement depends on the opposition of the ideas considered, as in the case of “man is not beast.” This means, as Bernoulli explained in later essays (see Capozzi 1994), that by converting “Animal is not a man” one obtains “Man is not animal,” which is false, because man and animal are not opposite ideas. Also in this case, logic has to resort to the quantified propositions of old syllogistic, but this means that there are no real logical equations between the ideas themselves. Not so in algebra, as it can be proved by the fact that every inequality is perfectly convertible. The conclusion is that no complete parallelism exists between algebra and logic. As a substitute, Bernoulli recommends the direct use of algebra in science by arguing that in science everything can be quantified and all that can be quantified can undergo algebraic treatment. His pioneering mathematical treatment of probability goes in that direction.

Bernoulli’s case is instructive. It shows that this is not a lethargic period of logic, as some historians have maintained (Blanché 1996, 169–178), but it also makes one wonder what made Bernoulli fail where other logicians—at the same time or a few decades later—made progresses. In our opinion, the main reason for Bernoulli’s failure was the doctrine of ideas he choose. We have already pointed out that Bernoulli depends on Port-Royal’s view that every idea of thing can be univocally designated by a word (at least in logic), and that every idea of thing is endowed with an indestructible comprehension, conveyed by the word. In the case of affirmation, this makes Bernoulli consider only the relation of containment of the predicate in the subject as basic. Consequently, he is unable to deal with possible predicates that do not disagree with the content of the subject but are not contained in its comprehension.

To build a calculus it is not enough to have a rudimentary algebra and a doctrine of ideas. One has to choose a suitable doctrine of ideas.

8. Leibniz

The German logician and philosopher Gottfried Wilhelm Leibniz (1646–1716) has a foremost role in the history of formal logic. However, it is almost impossible (and probably misleading) to represent his contributions to logic as a single and coherent set of theories nicely inserted within a linear path of development. There are at least three reasons that rule out such a reassuring view. In the first place, Leibniz contributed ideas—often through scattered and incomplete fragments rather than through structured and polished writings—to a plurality of logically relevant subjects: from the arithmetization of syllogistic to the theory of relations, from modal logic (and semantic of modal logic) to logical grammar—and the list could be easily extended. Moreover, relevant contributions are often found in works, fragments, or letters not explicitly devoted to the field of logic.

In the second place, most of Leibniz's writings testify to a work in progress in the deepest meaning of the expression: Different and sometimes incompatible strategies are explored in fragments dating back to the same years or even to the same months, corrections and additions may substantially modify the import of a passage, promising and detailed analysis remain uncompleted or are mingled with sketchy hints. Nevertheless, there is a method in Leibniz's passionate and uninterrupted research: a deeper unity that is given by a set of recurring problems and by the wider theoretical framework in which they are dealt with. Last but not least, it should always be kept in mind that Leibniz's works known by his contemporaries and immediate successors constitute a very limited subset of his actual production. It is only during the twentieth century that Leibniz's role in the history of logic came to be fully appreciated, and this appreciation is connected to at least two different moments: the publication by Louis Couturat, in 1903, of the *Opuscules et fragments inédits de Leibniz* (Leibniz 1966; see also Couturat 1901), many of which were devoted to logic, and the progress made during the second half of the century in the publication of the complete and critical edition of Leibniz's texts (Leibniz 1923–). Almost three centuries after his death, this edition (the so-called *Akademie-Ausgabe*) is, however, still to be completed.

Leibniz's interest in logic, and the amplitude of his logical background, is already evident in his youthful *Dissertatio de Arte Combinatoria* (Leibniz 1923–, VI, i, 163–230). This work is subdivided into 12 *problemata* (problems), mainly devoted to the theory of combinations and permutations, accompanied by a discussion of some of their *usus* (applications). Of greatest logical relevance are the combinatorial approach to syllogistic and the discussion of a symbolic language (*characteristica*) based on a numerical representation of concepts.

In dealing with syllogistic, Leibniz takes the work done by Hospinianus (Johannes Wirth, 1515–1575) as his starting point. Like him, Leibniz considers four different quantities—singular (S) and indefinite (I) propositions are added to universal (U) and particular (P) ones—and the two traditional qualities given by affirmative (A) and negative (N) propositions. Given that a syllogism consists of three propositions (the two premises and the conclusion), we have 4^3 possible combinations of the four different quantities and 2^3 possible combinations of the two different qualities. The number of possible different simple moods of the syllogism, valid and invalid, is therefore according to Leibniz $(4^3 \times 2^3) = 512$: the same result obtained by Hospinianus. If we take into account the four different syllogistic figures (Leibniz includes and explicitly defends the fourth figure, which was rejected by Hospinianus), we get a total of $(512 \times 4) = 2048$ "moods in figure." It is still through a combinatorial method, based on the exclusion of the syllogisms conflicting with four classic rules (nothing follows from pure particulars, no conclusion can be of stronger quantity than the weaker premise, nothing follows from pure negatives, and the conclusion follows the quality of the weaker premise) and of eight further moods conflicting with the rules given for the four syllogistic figures, that Leibniz gets the number of 88 valid syllogistic moods.

Whereas Hospinianus assimilated singular propositions to particular ones, Leibniz considers them similar to universal propositions, while indefinite propositions are connected to particular ones. In this way the number of valid moods in figure can be reduced to 24 (6 in each of the four figures): the 19 "classical" ones, plus 5 new ones that are actually the result of applying subalternation to the conclusions of the 5 "classic" moods with a universal conclusion.

The interest of Leibniz's treatment of syllogistic in the *De Arte Combinatoria* is not to be found in radical innovations concerning the number of moods of valid syllogisms, but rather in the fact that they are obtained through the systematic use of a combinatorial calculus, used as a sort of deductive device. The syllogistic "deduction" of the rules of conversion is also part of this attempt, based—as in Ramus and in a number of sixteenth- and seventeenth-century German logicians, including Leibniz's former teacher Jakob Thomasius (1622–1684)—on the use of identical propositions. In a later fragment, the *De formis syllogismorum mathematice definiendis* (Leibniz 1966, 410–416), identical propositions are used to obtain a syllogistic demonstration not only of conversion but also of subalternation.

Since in this demonstration a first figure syllogism is used, Leibniz can "deduce" all valid moods of the second and third figure using only the first four moods of the first figure, together with subalternation and the rule according to which if the conclusion of a valid syllogism is false and one of its premises is true, the second premise should be false, and its contradictory proposition should therefore be true (*methodus regressus*). The valid moods of the fourth figure can be deduced using conversion (the syllogistic proof of which only required moods taken from the first three figures). In this way, Leibniz will complete his construction of syllogistic as a sort of "self-sufficient" deductive system.

The second result of the *De Arte Combinatoria* worth mentioning is the construction of a symbolic language in which numbers are used to represent simple or primitive concepts, and their combinations (subdivided in classes according to the number of primitive concepts involved) are used to represent complex or derivate concepts. Fractions are used to simplify the representation of complex concepts, with the numerator indicating the position of the corresponding term within its class and the denominator indicating the number of the class, that is, the number of primitive concepts involved. In Leibniz's opinion, such a language would offer a solution to the main problems of the *logica inventiva* (logic of invention): finding all the possible predicates for a given subject, all the possible subjects for a given predicate, and all the possible middle terms existing between a given subject and a given predicate. This would also allow a mathematical verification of the truth of propositions and of the correctness of syllogistic reasoning. As we shall see, this project was to find a more developed and logically satisfactory form a few years later in 1679.

The idea of constructing a symbolic language in which numbers represent concepts is not new. Such an idea was already present in a number of attempts to construct a "universal language," attempts that were often influenced by the combinatorial works of Ramon Llull. Leibniz himself makes reference to the works by Johann Joachim Becher (1635–1682?; *Character pro notitia linguarum universali*, 1661) and Athanasius Kircher (1602–1680; *Polygraphia nova et universalis ex combinatoria arte detecta*, 1663); similar attempts were made by Cave Beck (1623–1706?; *The Universal Character*, 1657) and others, and are described by Kaspar Schott (1608–1666; *Technica Curiosa VII—Mirabilia graphica, sive nova aut rariora scribendi artificial*, 1664). In the same period, the Spanish Jesuit Sebastian Izquierdo (1601–1681; *Pharus Scientiarum*, 1659), aiming at a sort of "mathematization" of the *ars lulliana*, substituted numerical combinations for the alphabetical ones used by Llull, and something similar to a "numerical alphabet"—highly praised by Leibniz—was developed by George Dalgarno (c. 1626–1687) in his *Ars Signorum* (1961).

In Leibniz, however, the construction of a numerical *characteristica* is not only a handy representational device; it is strictly connected with the idea of the inherence of the predicate in the subject in every true affirmative proposition (*predicate-in-subject* or *predicate-in-notion* principle). This idea was already present in the scholastic tradition: In his *Commentary* on Peter of Spain's *Tractatus*, Simon of Faversham (c. 1240–1306) writes that propositions "are called complex because they are founded on the inherence of the predicate in the subject, or else because they are caused by a second operation of the intellect, namely the composition and division of simples" (Simon of Faversham 1969). During the Middle Ages, however, the inherence theory of the proposition was confronted with the idea according to which "in every true affirmative proposition the predicate and the subject signify in some way the same thing in reality, and different things in the idea" (Thomas Aquinas 1888–1889, I, xiii, 12).

The predicate-in-notion principle was to become a cornerstone of Leibniz's logical work, and Leibniz was to apply it not only to analytical but also to contingent propositions: "always, in every true affirmative proposition, necessary or contingent, universal or singular, the concept of the predicate is included in some way or other in that of the subject" (Leibniz 1973, 63). In the *De Arte Combinatoria*, however, Leibniz only deals with propositions made of general terms, and—as it has been already mentioned—the principle is mainly used for the discovery of subjects and predicates within the context of the *logica inventiva*. Its explicit use as a method for checking the truth of a proposition given its subject and its predicate is to be found only in the 1679 essays (see Roncaglia 1988), where Leibniz chooses to represent simple or primitive terms by means of prime numbers.

The advantages of this notation were already stressed in a fragment, dated February 1678, known as *Lingua generalis*: "The best way to simplify the notation is to represent things using multiplied numbers, in such a way that the constituting parts of a character are all its possible divisors. ... Simple elements may be prime or indivisible numbers" (Leibniz 1923–, VI, iv, 66). Just as compound (reducible) terms can be traced back—by means of definitions—to the simple, irreducible terms constituting them, the "characteristic numbers" of compound terms will be obtainable from the multiplication of the characteristic numbers of the simple terms constituting them, so that the characteristic number of a compound term can always be univocally broken down into those of the simple (relative) terms composing it.

Even at the time of the *De Arte Combinatoria*, Leibniz was conscious of the difficulty in finding terms that are really simple, and he had to be satisfied with simple terms of a relative and provisional nature. This difficulty was gradually to become, for Leibniz, an actual theoretical impossibility dependent on the limits of human understanding. While we can cope with abstract systems that are the result of human stipulation (and in which simple terms are established by us), only God can handle the much more complex calculus representing the infinite complexity of the actual world (and, as we shall see, of the infinite number of possible worlds among which the actual one has been chosen).

To give an example of his notation, Leibniz uses the definition of "homo" as "animal rationale," to which the following characteristic numbers are assigned:

$$\begin{aligned}&\text{animal} = 2\\&\text{rationale} = 3\\&\text{homo} = (\text{animal rationale} = 2 \times 3) = 6\end{aligned}$$

To verify the truth of a proposition, one has just to check whether the prime factors of the characteristic number of the predicate are or not all included among those of the characteristic number of the subject. The proposition "Homo est animal" is thus true, since the characteristic number of "animal" (i.e., 2) is a prime factor of the characteristic number of "homo" (i.e., 6).

A network of relations is thus established between the field of logic and its numerical "model," allowing an actual logical "interpretation" to be assigned to the numbers and arithmetical operations employed:

Number	Term
Prime number	Simple term
Prime factorization of number	Analysis of term
Number expressed in factorial notation	"Real" definition of term by means of its component simple terms
Multiplication (calculation of the least common multiple)	Conceptual composition
Exact divisibility of a by b (where a and b are the characteristic numbers of the terms A and B)	Verification of the truth of the proposition "A is B"

This calculus, however, presents some difficulties if it is used as a device by which to verify all forms of syllogistic reasoning, as Leibniz intended to do. While it permits adequate representation of UA and PN propositions, it is clearly unsuitable—despite Leibniz's repeated attempts (including the use of fractions and square roots) to get around the problem—to represent UN and PA propositions.

The difficulties Leibniz encounters here are connected to the representation of the incompatibility between terms: Though it is always possible to find the least common multiple of the characteristic numbers of two terms, it should not always be possible to construct a term that includes any two given predicates. Some predicates are simply incompatible. It is then hardly surprising that without a way to aptly "restrict" the combinations of the terms's characteristic numbers (i.e., to restrict conceptual composition), Leibniz finds it "too easy" to verify PA propositions and "too difficult" to verify UN propositions.

To solve this problem, Leibniz modifies his notation, making it more complex but much more powerful. Instead of using only one characteristic number for each term, he uses a pair of numbers—one positive and one negative—with no common prime factors. The use of "compound" (i.e., with a positive and a negative component) characteristic numbers allows for the following correspondences:

"Compound" characteristic number	Term
Positive component of characteristic number	"Affirmative" component of term
Negative component of characteristic number	"Negative" component of term
Prime factorization of "compound" characteristic number	Analysis of term
"Compound" characteristic number expressed in factorial notation, where no common prime factor is present in its positive and negative components	"Real" definition of term, demonstrating its possibility by the absence of contradictions within its definition
Presence of a common prime factor in the positive and negative components of a characteristic number	Logical impossibility of the corresponding term

Now, assuming that $a(+)$ and $a(-)$ represent the positive and negative components of the compound characteristic number assigned to the term A, that $b(+)$ and $b(-)$ have the same function with respect to the term B, and that A and B are possible terms (i.e., that no same primitive factor is present in either $a(+)$ and $a(-)$ or in $b(+)$ and $b(-)$), the following rules for the verification of propositions can be stated:

Presence in $a(+)$ and in $b(-)$ or in $a(-)$ and in $b(+)$ of at least one common prime factor	Incompatibility between A and B: verification of the proposition "no A is B" (UN)
Absence in $a(+)$ and in $b(-)$ or in $a(-)$ and in $b(+)$ of common prime factors	Compatibility of A and B: verification of the proposition "some A is B" (PA)
Exact divisibility of $a(+)$ by $b(+)$ and of $a(-)$ by $b(-)$	Verification of the proposition "every A is B" (UA)
Non exact divisibility of $a(+)$ by $b(+)$ and of $a(-)$ by $b(-)$	Verification of the proposition "some A is not B" (PN)

This more elaborate attempt is not without flaws, the most relevant being the problem of representing conceptual negation. Leibniz's proposal to obtain the characteristic number of a negative term like "non-A" by simply changing the sign of the two components of the characteristic number of the corresponding positive term "A" is ill founded and leads to inconsistencies in the calculus. The question raised here—the nature of conceptual negation—has always raised problems for Leibniz (see Lenzen 1986), and is also connected to the more "philosophical" problem of establishing the nature of incompossibility, a problem clearly stated in a famous passage of the fragment known as *De veritatibus primis*: "This however is still unknown to men: from where incompossibility originates, or what can make different essences conflict with each other, given the fact that all the purely positive terms seem to be compatible the one with the other" (Leibniz 1875–1890, VII, 195).

Nevertheless, Leibniz's logical essays of April 1679 represent one of the most interesting and complete attempts of arithmetization of the syllogistic, and offers a well-developed—albeit not fully satisfactory—account of traditional logic by means of an intensional calculus. In a sense, they also represent a turning point in Leibniz's logical works. The unsolved difficulties in finding a numerical model for his still mostly combinatorial calculus, and the problems associated with the representation of negation, probably led him to a twofold shift in his strategies. On the one hand, despite the interest that notational systems will have for him during all his life, Leibniz became increasingly aware that the research of an apt notation should be accompanied by a closer investigation of the logical laws and principles that should constitute the structure of the calculus. On the other hand, semantic acquires a deeper role: Leibniz perceives that the rules governing conceptual composition cannot be reduced to a sort of "arithmetic of concepts," and are much more complex. Negation, modality (with special emphasis on compossibility and incompossibility among concepts), relations (with special emphasis on identity), complete concepts of individual substances—much of Leibniz's logical and philosophical work in the following years will deal with these subjects.

The first tendency—a closer investigation of the logical laws and principles to be used in the calculus—is already clear in the *Specimen calculi universalis* and in its *Addenda* (both probably dating around 1680). Here, Leibniz abandons the exposition "by examples," favoring the much more powerful algebraic notation, which uses letters to represent concepts. The perspective is still intensional, and the inclusion of the predicate in the subject remains the cornerstone of the system. According to the *Specimen*, the general form of a proposition is "a is b," and a per se true proposition is of one the three following forms: "ab is a," "ab is b," or "a is a." A per se valid conclusion is of the form "if a is b and b is c, then a is c" (principle of syllogism), and according to Leibniz all true propositions can be derived from (or rewritten as) per se true propositions. The *Specimen* also contains the first clear formulation of one of Leibniz's key principles, that of substitution *salva veritate*: "Those are 'the same' if one can be substituted for the other without loss of truth" (Leibniz 1973, 34; for a discussion of this principle, see Ishiguro 1991, 17–43). Among the principles used are those according to which in conceptual composition the order of terms and the repetition of a term are irrelevant ("$ab = ba$," and "$aa = a$"). The *Addenda* offer a short discussion of negation, which was not considered in the *Specimen* itself, and add some further proofs, including the theorem "if a is b, and d is c, then ad will be bc." Leibniz calls it "praeclarum theorema" and proves it in this way: "a is b, therefore ad is bd (by what precedes); d is c, therefore bd is bc (again by what precedes), ad is bd, and bd is bc, therefore ad is bc" (Leibniz 1973, 41).

In the following years Leibniz will often use the signs "+" (or "⊕") and "−" to indicate logical composition and logical subtraction, stressing that the rules governing operations with concepts are different from those of arithmetical addition and subtraction: Whereas in arithmetic "$a + a = 2a$," in the case of conceptual composition "$a+a = a$." Moreover, Leibniz will carefully distinguish between conceptual subtraction and logical negation: While in an abstract conceptual calculus it is always possible to "subtract" from the concept of man that of rationality, seen as one of its intensional components, the result of *denying* it ("men non-rational") is a simple impossibility, given the fact that rationality is an essential part of the concept of man (*Non inelegans specimen demonstrandi in abstractis*, Leibniz 1923–, VI, iv, n. 178). These principles are among the ones governing the so-called plus-minus calculus that Leibniz developed in a number of fragments dating around 1687. The basic assumption of the plus-minus calculus is that "$A+B = L$" is to be interpreted as "A (or B) is included in L," where the relation of inclusion is—as usual—the intensional inclusion between concepts. "$L - A = N$" is to be interpreted as conceptual subtraction: N is the intensional content of the concept L that is not included in the concept A. If "$A+B = L$," the terms A (or B) and L are said to be *subalterns*; two terms, neither of which is included in the other, are said to be *disparate*, and two terms that have a common component are said to be *communicating*. It should be stressed that, here as elsewhere, Leibniz uses "term" to designate not a linguistic entity but a concept: His calculus is thus directly an "algebra of concepts." The calculus uses "nihil" to represent a term with empty intensional content, and the rules "$A+\text{nihil} = A$" and "$A-A = \text{nihil}$" are introduced. Leibniz also uses "nihil" as a way to obtain "privative" concepts: if "$E = L - M$," "$L = \text{nihil}$," and M is a nonempty concept, E will be a privative concept. This assumption has been criticized due to the fact that it introduces inconsistencies in the calculus (Lenzen 2004), but can be seen as a further indication of the relevance that Leibniz attributed to the representation of negative or privative concepts within his logic, and of the difficulties connected with the difference between conceptual subtraction, arithmetical subtraction, and logical negation.

The plus-minus calculus has recently been the subject of much interpretive work, also due to the publication of the long-awaited vol. VI, iv of the *Akademie Ausgabe*, which offers the first critical and complete edition of the relevant texts (Leibniz 1923–, VI, iv, vols. 1–3). Among the problems debated (see Lenzen 2000, 2003, 2004; Schupp 2000) are the possibility of a set-theoretical representation of the calculus, and the relation between its intensional and extensional interpretations. As we have seen, Leibniz's approach is—in most of his logical writings—clearly intensional. However, Leibniz himself was well aware of the difference between intensional and extensional approaches, and considers the one as the reversal of the other:

> the method based on concepts is the contrary of that based on individuals. So, if all men are part of all animals, or if all men are included in all animals, it is true that the notion of animal is included in the notion of man. And if there are animals that are not men, we need to add something to the idea of animal to get the idea of man. Since when the number of conditions grows, the number of individuals decreases. (Leibniz 1966, 235)

This thesis may be (and has been) criticized, since the number of actually existing individuals falling under a given concept is usually contingent: From a Quinean point of view, "it might just happen that all cyclists are mathematicians, so that the extension of the concept *being a cyclist* is a subset of the extension of the concept *being a mathematician*. But few philosophers would conclude that the concept *being a mathematician* is in any sense included in the concept *being a cyclist*" (Swoyer 1995, 103). Nevertheless, as Lenzen (2003) correctly observes, this criticism cannot be applied (or at least not in such a naive form) to Leibniz's logic: The (extensional) domain of Leibniz's logic is consistently characterized by Leibniz himself as one of *possible* rather than of *actual* individuals. The (possible) contingent coincidence of the sets of actually existing cyclists and mathematicians would by no means imply, from a Leibnizian point of view, that the two concepts have the same extension and therefore should have the same intensional content. In a later fragment, known as *Difficultates quaedam logicae* (Leibniz 1875–1890, VII, 211–217), Leibniz will even use the idea that the domain of his logic is one of possible rather than of actual individuals to justify the conversion per accidens of UA propositions (from "All A are B" to "Some B is A"), thus avoiding the problem of the existential import of PA propositions.

Despite the fact that they probably precede most of the texts on the plus-minus calculus, the 1686 *Generales Inquisitiones* (Leibniz 1982) are generally considered Leibniz's most developed and satisfactory attempt of logical calculus; Leibniz himself considered them a "remarkable progress" over his earlier works. The main feature of the *Generales Inquisitiones* is the attempt to offer a unified framework for a calculus of terms and a calculus of propositions. As far as terms (or concepts) are concerned, Leibniz distinguishes between *integral terms* (terms that can be the subject or the predicate of a proposition: the *categorematic* terms of scholasticism) and *partial terms* (terms like "same" or "similar," which are to be used only in conjunction with one or more integral terms, and specify or modify an integral term or a relation among integral terms: the *syncategorematic* terms of scholasticism). The introduction of partial terms and the discussion of oblique cases clearly testify to the new interest Leibniz devoted to relations. Being discussed at the term level (and therefore at the level of concepts), relations and oblique cases are clearly not considered by Leibniz as mere linguistic accidents. The problem of the possible "reduction" of partial or relational terms and of relational propositions to nonrelational ones is therefore not one of simple "surface-structure" reformulation of a spoken or written sentence, but rather one of logical analysis of the proposition and of its constituent terms. Leibniz was to devote much effort and a large number of texts and fragments to this analysis, clearly influenced by the late scholastic discussion on relations and on the connection between the relation itself and its *fundamenta*, that is, the concepts or the things among which the relation is established. Starting with Russell (1900), who attributed to Leibniz a straightforward and uniform reductionistic approach with respect to relations, criticizing it, Leibniz's treatment of relations has been the subject of much interpretive work. It is now clear that Leibniz offered different accounts for different kinds of relations and that, while he consistently denied relations an extramental reality independent from that of the related concepts, he thought that at least some relations (among those involving different individuals) are not reducible in a straightforward way to simple and nonrelational monadic predicates. However, this does not imply, according to Leibniz, the need of propositions which are not in subject-predicate form, but rather the need (1) to consider within the properties pertaining to a given subject also those expressing relational accidents, and (2) to recognize the logical role of reduplicative terms, which can be used in connecting propositions referring to the different *fundamenta* of a same relation. Thus, the proposition "Paris loves Helen" can be reduced, according to Leibniz, to "Paris is a lover, and *eo ipso* (for this very reason) Helen is a loved one," rather than to the simple and independent propositions "Paris is a lover" and "Helen is a loved one" (Mugnai 1992). Reduplicative terms like *quatenus*, *eo ipso*, and so on, which were already studied by scholastic and late-scholastic logicians, in this way acquire a special role within Leibniz's logic.

In the *Generales Inquisitiones*, integral terms are further subdivided into *simple*, *complex*, and *derivative*. The discussion of simple terms shows a clear shift when compared to the earlier combinatorial attempts: While general abstract terms like "ens," derived from the scholastic tradition and from the discussion on transcendental terms, are still present, Leibniz adds to the class of simple terms also terms connected to individuals and perceptions, like "Ego" ("I") or the names of colors; a passage that somehow anticipates the discussion about simple and innate ideas that will be at the core of the much later *Nouveaux Essais* (Leibniz 1923–, VI, vi). However, as in the earlier attempts, the choice of simple terms remains provisional and is strongly influenced by the limits—both necessary and contingent—of our knowledge. A special case is that of the privative term *non-ens*: Like *nihil* in the plus-minus calculus, *non-ens* corresponds here to a term with empty intensional content, and plays an important role in the axiomatization of the calculus.

Complex terms are obtained by composition from simple terms, while derivative terms are obtained from partial terms "completed" by integral terms, that is, from integral terms modified or connected by means of syncategorematic and relational terms or by the use of oblique cases.

In the *Generales Inquisitiones*, for the first time, Leibniz includes a discussion of complex terms referring to individuals (Leibniz 1982, 58–62) in his logical calculus. According to Leibniz, they are based on *complete concepts*, that is, concepts that include all which can be said of that individual. Their complexity, however, is such that only God can carry out their complete analysis: Men can only rely on experience to assert the possibility of a given complete concept (i.e., the absence of contradictions within its intension) and the inclusion of a given contingent predicate within a given complete concept. Complete concepts (corresponding to individual substances) are another theoretical cornerstone of Leibniz's philosophy, and it is no coincidence that in the very same year in which he was working at the *Generales Inquisitiones* Leibniz also wrote the *Discours de métaphysique* (Leibniz 1923–, VI, iv B, 1529–1588), the text that offers for the first time and in a structured way the philosophical framework in which the theory of complete concepts is to be placed. Much of the interpretive work done on Leibniz's philosophy and philosophy of logic in the last three decades deals in one way or another with the discussion of complete concepts[1]: from the possibility of distinguishing within them a "core set" of essential properties, which could also allow for transworld identification of individuals across possible worlds (each complete concept, if considered in its integrity, is bound to a given possible world, and possible worlds themselves are seen by many interpreters as maximal sets of mutually compossible complete concepts), to the presence within complete concepts of relational predicates; from the discussion of Leibniz's conception of contingency and of individual freedom to that of preestablished harmony.

Shifting from terms to propositions, Leibniz states in the *Generales Inquisitiones* that "'A is B' is the same as 'A is coincident with some B' or $A = BY$" (Leibniz 1973, 56), where B is part of the intensional content of A:

a formulation close to the ones we have already discussed, which, however, can be of interest if we consider the role attributed here to "Y," seen as a sort of existential quantifier applied to the predicate. Leibniz offers a wider discussion of predicate quantification in an undated fragment known as *Mathesis Rationis* (Leibniz 1966, 193–206; see Lenzen 1990), and in some of his (many) attempts of graphical representation of the basic notion of conceptual inclusion and of the four forms of categorical propositions, mainly based on the use of lines or circles (the fragment known as *De formae logicae comprobatione per linearum ductus* probably being the most notable among them; Leibniz 1966, 292–321).

In the *Generales Inquisitiones*, the connection between the treatment of propositions and that of terms is, if possible, even stronger than in the preceding essays, since Leibniz observes that the four traditional forms of categorical propositions can be rewritten in the following way (Leibniz 1982, 112; his thesis is clearly indebted to the late-scholastic treatment of the passage from proposition "tertii adjecti" to propositions "secundi adjecti"):

(PA)	Some A are B	AB est res
(PN)	Some A are not B	A(non-B) est res
(UA)	All A are B	A(non-B) non est res
(UN)	No A is B	AB non est res

Here—if the proposition is not one about contingent existence—"est res" is to be interpreted as "is possible" (Leibniz 1982, 110), and possibility is in turn to be interpreted as absence of contradiction within the intension of the composed term. A similar conception is to be found in the *Primaria Calculi Logici Fundamenta*, dating to August 1690 (Leibniz 1903, 232–273).

This treatment of propositions leads to a term-oriented treatment both of syllogistic inferences and of hypothetical propositions. According to Leibniz, just like in a categorical proposition the subject includes the predicate, in a hypothetical proposition the antecedent includes the consequent. Therefore, an implication of the form "If p, then q" is to be interpreted as "If (A is B) then (C is D)," which in turn can be rewritten as "(A is B) is (C is D)," or "(A includes B) includes (C includes D)." This idea, already present in a fragment known as *Notationes Generales* probably written between 1683 and 1685 (Leibniz 1923–, VI, iv, 550–557), will return in many later texts and leads Leibniz to hold that the forms and modes of hypothetical syllogisms are the same as those of categorical syllogisms.

Leibniz never devoted a detailed analysis to the logic of propositions, but in many of his works and fragments refers to propositional rules derived from the medieval tradition of consequences and from the late-scholastic discussion on topical rules; clearly, in his opinion, an "algebra of propositions" can only be grounded on the algebra of concepts.

As already noted, most of the logical texts and many of the most remarkable achievements of Leibniz's logic were not known to his contemporaries and his immediate successors. Nevertheless, Leibniz's logic cannot be considered simply an isolated product of a genial mind: He had a deep knowledge of the late-scholastic logical tradition, from which he derives not only many topics he deals with but often also the approaches adopted in dealing with them. He also had a wide net of relationships—both through letters and by personal acquaintance—with many of the most prominent figures of the European learned world (among them Arnauld, Tschirnhaus, Jakob Bernoulli, and, as we will see, Wolff, to name just some scholars mentioned in this chapter). His logical and philosophical theses are also the result of those interactions, and probably some of them circulated even without the support of publication. Despite the great interest of Leibniz's logic from a contemporary point of view, Leibniz was a seventeenth-century logician, not a twentieth-century logician in disguise.

9. Logic in Germany in the First Half of the Eighteenth Century

In the period we are considering, German logic deserves special attention. Since logic was a subject included in most academic curricula, it became a privileged field of study and a great number of logical texts were published (see Risse 1965). Many German logicians enter the debate on Cartesianism, are fully aware of Bacon's exhortation to work at a logic of empirical sciences, pay attention to the notion of probability, examine the relationship between logic and mathematics, and seem open to the suggestions of facultative logic. If one had to name a single author who takes a stand on all these questions, one should mention Leibniz. But, as already said, in this period Leibniz's logic enters marginally in the official picture of German logic, not only because most of his strictly logical production was unknown at that time but also because he did not belong to the academic world. What was known of Leibniz's philosophy and logic influenced a number of German logicians of that time, but the logical scene of the first two generations of the German Enlightenment was dominated by Christian Thomasius and Christian Wolff.

Christian Thomasius (1655–1728) was the son of Jakob Thomasius, Leibniz's teacher. However, he does not share Leibniz's view of logic, in as much as he agrees with the humanists in criticizing schoolmen for having instructed generations of students in the making of useless subtleties. At the same time he advocates the study of logic. This is less paradoxical than it sounds. While the *Port-Royal Logic* recommended a logical instruction because common sense is not so common as people believe, Thomasius offers a moral and religious justification. He maintains that because of the original sin, mankind has darkened its natural light (*lumen naturale*) and has to achieve a healthy reason through a process of purification, so as to avoid the errors it usually makes. This cathartic process is entrusted by Thomasius to logic because logic can teach how to counteract errors in judgment and their main source, namely, prejudice (in particular, the prejudice of authority and the prejudice of self love). In this way Thomasius agrees with Tschirnhaus and the *Port-Royal Logic* that logic is a medicine and its primary aim is the correctness of judgments (Thomasius [1691b], dedication). From the *Port-Royal Logic*—translated into Latin (Arnauld and Nicole 1704) by one of his followers—Thomasius also borrows arguments for rejecting Aristotelian categories (Thomasius 1702, VII §25).

Such a concept of logic requires that technicalities should be abolished: Only an "easy" logic can dispel prejudices and teach how to profit from the few precepts needed for the rational conduct of common human beings—and not only learned scholars—in the search of truth and in the practical exercise of prudence. Thomasius's precepts consist of two basic rules of method: (1) to proceed from what is easy and known to what is more difficult and unknown, (2) to connect remote conclusions to principles only through near (*propinquae*) conclusions. In these two rules one can find an echo of Descartes. But Thomasius is not a Cartesian because he opposes the doctrine of innate ideas, convinced that in the intellect there is nothing that has not been in the senses (a thesis he argues for independently of Locke). He also rejects the tradition of the *mos geometricus* because he holds it responsible for the degenerate Spinozistic version of Cartesianism so that, from this point of view, he differs also from Tschirnhaus. Thomasius is rather an eclectic (Beck [1969], 247–256) who encourages the study of other philosophers's ideas, thus promoting studies in the history of philosophy and also in the history of logic: From 1697 to the first decades of the eighteenth century it is possible to register a number of essays on the latter subject (Risse 1964–70 II, 507). Thomasius's eclecticism can be easily appreciated if one considers that, on the one hand, he adds precepts derived from the tradition of humanistic dialectic to his apparently Cartesian rules of method and, on the other hand, he insists that logic should concentrate on the problem of certainty in empirical knowledge. Like many others in this period, Thomasius believes that, although in empirical matters complete certainty is not attainable, it is still possible to avoid skepticism by working on the notion of probability. However, Thomasius's interest for probability is not to be overestimated, since he inclines to a notion of probability still strongly connected with Aristotelian dialectic and with the doctrine of topical syllogism.

Thomasius's ideas were well received by the incipient age of the Enlightenment that looked favorably to a logic meant for ordinary people (and this favored the proliferation of textbooks) and, from a more theoretical point of view, approved of his antiskeptic battle regarding empirical knowledge. Nevertheless, Thomasius not only advocated a rigid separation between mathematics and philosophy, but also opposed any formalism in logic and deplored the enormous growth of syllogistic, convinced that the first figure is sufficient, though incapable of guaranteeing the truth of the conclusions (Thomasius 1702, IX, §12; [1691a], XII, §§19–21). Many of his followers agreed on these matters, with a notable exception. In Halle, where a Thomasian circle was flourishing, Andreas Rüdiger tried to reconcile Thomasius's views on logic and philosophy with his own research on the nature and scope of logic.

Rüdiger (1673–1731) agrees with Thomasius that, due to the original sin, mankind no longer participates in God's archetypal logic and is prey of prejudices, thus making it necessary to conquer a *recta ratio* through the study of logic (Rüdiger 1722, I, i, 1). Rüdiger also agrees with Thomasius that logic should deal with probability as a response to skepticism, and to this effect he produces an articulated doctrine of probability that obtained remarkable diffusion through the first edition of the *Philosophisches Lexicon* (Walch [1726]) published by Johann Georg Walch (1693–1775). But what seems to interest Rüdiger most is that logic should be recognized as a legitimate means for finding truths and should be proved capable of attaining this purpose with a procedure as different as possible from the procedure of mathematics.

Rüdiger believes that Spinozism rests on two pillars: the doctrine of innate ideas and the illegitimate application of the mathematical method outside mathematics. Therefore, he rejects both. He is perfectly aware that mathematics is inventive, but he makes this depend on the fact that mathematical proofs can resort to sensibility. Rüdiger does not ascribe the sensibility of mathematical proofs to the use of "visual" aids, such as geometric figures, but to the demonstrative procedures based on numeration. In his opinion: (1) mathematics is the science of quantity, (2) all quantities are measurable, (3) we can measure only in so far as we can numerate, (4) "all numeration is of individuals, in so much as their terms are perceived by the senses." His conclusion is: "Therefore all numeration is sensible: but the entire way of mathematical reasoning is numeration, then this entire way [of reasoning] is sensible Q.E.D." (Rüdiger 1722, II, iv, 1a). The possibility to avail themselves of this kind of sensible reasoning (*ratiocinatio sensualis*) enables mathematicians to refer to sensible data that would escape their attention if they could rely on intellect alone, whereas sensible data offer them the basis for the discovery of unknown (mathematical) truths (Rüdiger 1722, II, iv, 3c; see Cassirer 1922, 525–527).

The heuristic capacity of *ratiocinatio sensualis* rests on Rüdiger's theory of truth. According to Rüdiger, the truth of our judgments (which he calls logical truth) consists in the agreement of our knowledge with our sensation ("convenientia cognitionis nostrae cum sensione"; Rüdiger 1722, I, i, 12). But we can trust the agreement of our knowledge with sensation only because there is a metaphysical truth that consists in the agreement of sensation with its objects ("convenientia ipsius sensionis cum illo accidente, quod sentitur"). This means that the metaphysical truth, which makes us trust our logical truth, presupposes that our senses are not fallible (Rüdiger 1722, I, i, 11). Because our senses are not fallible (under God's guarantee), the certainty and inventive power of mathematics can rest on sensible reasoning.

Outside mathematics, however, and in particular in the field of philosophy, we work only with ideas and cannot resort to the *ratiocinatio sensualis*. It is nevertheless possible to use a logical way of reasoning meant for ideas (*ratiocinatio idealis*) and as inventive as the *ratiocinatio sensualis*: the syllogism. By claiming that syllogism is inventive, Rüdiger openly challenges a long series of scholars, including Thomasius, who had criticized syllogism for being sterile. He argues that the inventive function of syllogism has not been appreciated because syllogism has been used for finding the premises of a given conclusion. This means that syllogism has been used analytically, whereas syllogism can be inventive if it is used synthetically, as a means for searching an unknown conclusion beginning from a given premise (Rüdiger 1722, II, vi, 1).

To show how a *synthetic syllogistic* is possible, Rüdiger assumes that every proposition (of the four kinds that can enter into syllogisms, A, E, I, O) expresses a precise relation between the subject and the predicate, a relation belonging to a set Rüdiger carefully classifies: subordination, opposition, partial diversity (he considers identical ideas as the same idea; see Rüdiger 1722, I, xii, 2, 3). On this assumption he maintains that we make a synthetic syllogism beginning with a premise whose two terms stand in one of the admissible idea-relations. We then obtain a conclusion by connecting one of the terms of the premise with *any* unknown idea that stands in a definite relation (included in the set of classified idea-relations) with the other term. For instance, given a universal affirmative proposition "All A are B" (where A is subordinated to B) as premise, we can connect the predicate B with any idea C of which we only know the relation it entertains with the subject A. Let such a relation be that of subordination: We can validly conclude that the unknown idea C, being subordinated to A is also subordinated to B, so that, since the relation of subordination can be expressed by a universal affirmative proposition, we obtain the "new" conclusion "All C are B" (Rüdiger 1722, II, vi, 55 1–4). This single example makes it clear that Rüdiger's synthetic syllogistic is founded on the old technique of the *pons asinorum* (Thom 1981, 72–75), traditionally used for finding premises, a technique that Rüdiger could find in works of the Peripatetic tradition he knew well, *pace* Thomasius who had ridiculed it (Thomasius [1691a], XII, §11). Rüdiger simply reverses the *pons asinorum* in the search for a conclusion, as it can be appreciated from the graphical representations he gives of his synthetic syllogisms (reunited in a single representation by Schepers 1959, 99).

Rüdiger expressed his views on *ratiocinatio idealis* in a logical environment in which they must have been unpopular. It is not surprising, therefore, that his direct followers, who approved of his separation of mathematics from logic, were no longer interested in his reason for separating them, that is, his passionate defense of the inventive capacity of syllogism. Adolph Friedrich Hoffmann (1703–1741) and Christian August Crusius (1715–1775) choose to refer to Rüdiger as the philosopher who did not simply denounce, like Thomasius, the ill effects of the application of the *mos geometricus* to philosophy, but had argued that mathematics and philosophy should never be mixed because they have different objects and different methods of reasoning (Tonelli 1959). It may seem paradoxical, therefore, that Rüdiger's classification of idea-relations, supporting his doctrine of syllogism (as well as his doctrine of conversion and other nonsyllogistic inferences) and connected to his thesis of the independence of logical from mathematical reasoning, was to stimulate later logicians to once again take up projects of an algebraic calculus of ideas. But such projects were not resumed until the Thomasian school, including its peculiar Rüdigerian variant, was no longer dominant, due to the emergence of Christian Wolff and his school.

Wolff (1679–1754) has much in common with the first generation of the German Enlightenment. Like the Thomasians, he believes that philosophy should improve human life and give due importance to empirical knowledge. He also shares some of their religious motivations and their appreciation for Locke (see section 6). What divides Wolff from Thomasius on logical matters is, first, a very different evaluation of mathematics and of the tradition of the *mos geometricus*: Wolff had received a mathematical education and was professor of mathematics in Halle. Second, Wolff acknowledges the intellectual influence exerted on him by Leibniz, with whom he exchanged a correspondence that ended only with Leibniz's death. In one of his letters Leibniz had recommended to Wolff to pay due attention to syllogism: "I absolutely never dared to say that the syllogism is not a means for finding truths" (Leibniz [1860], 18).

Consequently Wolff, far from contrasting syllogistic and the mathematical method, makes a double revolution with respect to the Thomasian school. He (a) assumes mathematical reasoning as an example to be followed in any research field, (b) claims that the allegedly empty and useless syllogism is the inner fabric of any reasoning, including the exemplary mathematical one. In different places—but especially in his logical works, the so-called *German Logic* of 1713 and the so-called *Latin Logic* first published in 1728—Wolff considers geometrical demonstrations as chains of common syllogisms in the first figure (Wolff [1713], IV, §§20–25, [1740] §551), and concludes that, where understanding and reason are concerned, there is a single rational procedure, a single method, a single logic valid for mathematics and philosophy alike: "both philosophy and mathematics derive their method from logic" (Wolff [1740], Preliminary Discourse §139 note).

The importance of syllogism is justified by Wolff by the argument that syllogism mirrors the natural way of reasoning. But he does not ground logic on empirical psychology alone (better: on the empirical psychology of a privileged set of men, the mathematicians, see Engfer 1982, 225). He declares that logic has a solid foundation also in ontology (Wolff [1740], Preliminary Discourse §89).

Both pillars of this foundation of logic conspire in favoring syllogism: (1) Ontology, as it was established in the scholastic tradition, justifies the *dictum de omni et nullo* that presides over syllogism in the first figure (Wolff [1740], §380); (2) empirical psychology ensures that the model of natural inference is the simplest syllogistic inference, that is, syllogism in the first figure. Wolff's foundation of syllogism—and indeed of logic—on ontology and empirical psychology grants an absolute privilege to syllogisms in the first figure. To this effect, Wolff maintains that figures different from the first (he does not admit the fourth figure) are not simply reducible to it but are already syllogisms in the first figure in disguise: They are cryptic first figure syllogisms (Wolff [1740], §§382–399; see Capozzi 1982, 109–121). He also maintains that noncategorical syllogisms, *consequentiae immediatae* and any other kind of inference are reducible to syllogisms in the first figure. In this way, Wolff contrasts anti-syllogistic conceptions, but makes no concessions to Rüdiger: The syllogisms he refers to are absolutely standard and in no way synthetic. In a letter to Leibniz, he scorns Rüdiger's synthetic syllogistic and refers to the idea of a mathematical *ratiocinatio sensualis* as to one of the "paradoxa" of Rüdigerian logic (Leibniz [1860], 117).

Despite his meager syllogistic, Wolff offers a deeply "logicist" philosophy of logic, for not only does he bring logic and mathematics together, he also considers logic prior to mathematics. This makes one wonder why he did not try a mathematical calculus of ideas. This problem is clearly related to the possibility of a heuristic. Wolff knows that mathematicians use heuristic devices not reducible to syllogisms. In his *German Logic* he denies that "the whole algebraic calculus ... takes place only according to syllogisms in form" (Wolff [1713], IV, §24). A similar statement is to be found in his *Latin Logic*: "logic [i.e., logic centered on syllogism] has a notable and famous use in the art of discovery, but nevertheless it does not exhaust it" (Wolff [1740], §563). Elsewhere he explains that, to discover hidden truths, it is sometimes necessary to resort to heuristic artifices such as, in the a priori invention, the artifices of the *ars characteristica* (see Arndt 1965, 1971). For, he says, this art helps separate geometric and arithmetic truths from images, so as to obtain truths from the data by means of a calculus. He grants that such an art is the most perfect science, but he believes that we only have a few examples of that art in algebra, yet none outside it (Wolff [1713] IV, §22). The latter statement is revealing: To Wolff the establishment of a suitable alphabet of thoughts as a prerequisite of the *ars characteristica combinatoria* must have appeared too great an obstacle. We know that this was a problem for Leibniz, and we know how he dealt with it. But apparently Wolff was convinced that a calculus *in logicis* is utopian and considered it a mere *desideratum*.

As to Leibniz's interest in a logic of probability to be used when deliberating about political, military, medical, and juridical matters (Leibniz [1860] 1971, 110), Wolff seems to doubt that a mathematical *ars conjectandi* could be of practical use regarding such matters (Leibniz [1860] 1971, 109). Wolff's doubt was not so strong as to make him exclude that Leibniz's wish could ever come true, but was strong enough to make him exclude that it could come true in the foreseeable future. Nevertheless, Wolff includes probability in the practical part of his logic, paying attention to the features of probable propositions, in particular to the ratio of sufficient and insufficient reasons that make it possible to consider probability a measurable degree of certainty. In this way, he definitely abandons not only Thomasius's obsolete treatment of probability reminiscent of the old topic but also Rüdiger's nonmathematical analysis of probable knowledge.

Wolff made a great impact on German philosophy, and some of his doctrines were well received in Europe at large, as can be appreciated from the evident Wolffian imprint of some entries in the French *Encyclopédie* (Carboncini 1991, 188ff.), where Locke and *Port-Royal Logic* had already found much space (Risse 1964–70, II 528). However, at first Wolff met with harsh criticism, albeit not for his logical but for his metaphysical views. In 1723, the Halle circle succeeded in convincing King Friedrich Wilhelm to banish Wolff from the city because his alleged determinism, inherited from Leibniz, was a peril to religion (Wundt [1945], 234–244; Beck [1969], 258). Wolff went to Marburg, where he wrote a series of works in Latin, beginning with the *Latin Logic*. Wolff's Latin works increased the number of his followers, so much so that a second anti-Wolffian offensive, launched in 1734, ended in a defeat when, in 1740, Wolff was readmitted to Halle with great honors.

The Wolffian era, as concerns logic, was very positive. Despite literature that considers him an exponent of the dark ages of logic, it is difficult not to give him credit for proposing a positive image of the discipline resting not only on its function as a guide in making judgments and in avoiding errors but also on the power of its inferences. But above all, it is impossible to ignore that his revaluation of syllogism differed from Rüdiger's because he used it to counteract the idea of a gap between logic and mathematics. This explains why Wolff's success promoted a revival of logic that did not simply contribute to the production of new logical textbooks—given the importance he accorded to logic in academic curricula—but spurred logical investigations. It must be stressed that those who were encouraged by Wolff's philosophy to engage in logical research did not usually follow the details of his logic. Independent authors, as well as a number of Wolffians, referred to old logical literature, including sixteenth- and seventeenth-century Aristotelian and scholastic treatises, and even to the works of anti-Wolffians, especially those of Rüdiger, undoubtedly the most distinguished of them. In this sense one must agree with Risse that if one excludes the very first generation of Wolff's followers, it is difficult to draw precise boundaries between the Wolffian school and its opponents (Risse 1964–70, II, 615).

An example of the new post-Wolffian logicians is Johann Peter Reusch (1691–1758). In his fortunate *Systema logicum* (Reusch [1734]), though in many respects faithful to Wolff, Reusch admits the influence of Aristotle, Jungius, the *Port-Royal Logic*, Johann Christian Lange (1669–1756), and Rüdiger (Reusch [1734], preface). As to the question of syllogistic, Reusch informs his readers about traditional doctrines and about the combinatory of syllogistic moods with a reference to Leibniz's *De Arte Combinatoria* (Reusch [1734], §530). Nevertheless, he also proposes a syllogistic that, he maintains, opens the gates of the syllogistic moods (*modorum cancelli*) (Reusch [1734], §543), being founded on a single rule to which all syllogisms of any figure must conform: "The entire business of reasoning is done by substitution of ideas in the place of the subject or of the predicate of the fundamental proposition, that some call equation of thoughts" (Reusch [1734], §510). In other words, Reusch conceives of syllogisms as consisting of a single premise (*propositio fundamentalis*) and a conclusion obtained by assuming a new idea and substituting it for either the subject or the predicate of the *propositio fundamentalis* by a substitution rule. Such a substitution rule—which he means as a version of the old *dictum de omni et nullo*—is governed by the principle of contradiction and presupposes a network of relations among ideas.

The description of the admissible idea-relations is so important for Reusch's syllogistic, and indeed for his whole logic, that two chapters of his *Systema logicum—De convenientia et diversitate idearum* and *De subordinatione idearum*—are devoted to it. This study is clearly influenced by Rüdiger (Capozzi 1990, lxvi), but Reusch did not refer to Rüdiger as the defender of a synthetic inventive syllogistic but as the author of a syllogistic based on a definite set of idea-relations. That this was the outstanding feature of Rüdiger's logic was clear to the historian of logic von Eberstein who, though unfavorable to Rüdigerian philosophy, in 1794 stated that Rüdiger had been the first to determine "syllogistic figures according to the relations of concepts and not according to the position of the middle term" (von Eberstein [1794–1799], I, 112–113). No wonder the independent Wolffian Reusch was attracted to this approach to syllogistic so as to prefer it to Wolff's idolatry for the first figure and to Leibniz's combinatory of moods in the *De Arte Combinatoria*.

In Reusch, however, there is no hint of a separation between the *ratiocinatio sensualis* of mathematics and the nonmathematical *ratiocinatio idealis* advocated by Rüdiger as an argument in favor of the inventive power of his synthetic syllogisms. This is true not only of Reusch. After Wolff, logic is acknowledged as the only argumentative structure used in every field of knowledge, from mathematics to philosophy. This is why a few logicians felt entitled to take a further step: If, according to Wolff, there is no longer a gap between logic and mathematics, nothing prevents one from disregarding Wolff's restriction of logic to an outdated syllogistic. These logicians felt entitled to apply mathematical tools to ideas according to the study of idea-relations made by the adversaries of Wolffian logic or by independent Wolffians.

10. Logical Calculi in the Eighteenth Century

In Jena, where Reusch was professor of logic and metaphysics since 1738, his attitude toward logic was not an exception, as can be seen in the logical work of Joachim Georg Darjes (1714–1791). But the outstanding work written in this logical context is an essay that Reusch recommended in the 1741 edition of his *Systema logicum* to his more specialized readers. This work is the *Specimen Logicae universaliter demonstratae* (Segner [1740] 1990) written by the mathematician and scientist Johann Andreas Segner (1704–1777) with the explicit aim of treating syllogistic by way of a calculus (*per calculum*) based on the example of algebra.

To this end, Segner builds an axiomatic system consisting of 16 definitions, 3 postulates, and 2 axioms. The definitions introduce ideas, their relations, their arrangement in a hierarchy of genera and species, and the operations for forming ideas. Segner defines *idea* as a mental representation of something. If the idea is simple, its contents are obscure ideas and the simple idea is confuse for us; if the idea is composite, its contents are clear ideas and the composite idea is distinct for us. Consequently, by definition, every idea contains some idea within itself. In this way, Segner can presuppose the relation of *containment* (viewed from an intensional perspective) as the basic relation between two ideas. But it must be clear that Segner does not identify the content of an idea with its comprehension in the sense of the *Port-Royal Logic*. He simply says that given two ideas A and B, A is contained (or involved) in B if, whenever B is posited, A is also posited.

The notion of containment is used to define all the relations between two ideas that are relevant for the construction of a calculus. Segner designates such relations by special algebraic symbols "$-$", "$=$", "$>$", "$<$", "$\times$", and defines them as follows:

I. A is opposed to B, if A contains $-B$ and B contains $-A$.

II. A is identical to B ($A = B$), if A contains B and B contains A.

III. A is superior to B ($A > B$), if A does not contain B and B contains A.

IV. A is inferior to B ($A < B$), if A contains B and B does not contain A.

V. A is coordinated to B ($A \times B$), if A does not contain B and B does not contain A.

As can be seen from this list, Segner does not have a symbol for the relation of opposition, but expresses the opposition between A and B by saying that A contains $-B$ and B contains $-A$. By the expression $-A$ Segner refers to the idea that is *infinitely* opposite to A, and defines A as infinitely opposite to $-A$ if A contains $--A$, and $--A$ contains A. It must be stressed that Segner does not intend an idea designated by $-A$ as a "negative" idea opposed to a "positive" one: If by A we indicate "nontriangle," by $-A$ we indicate "triangle."

Ideas that can be put in a hierarchy of subordination, can also be submitted to the operations of composition and abstraction, whose possibility Segner guarantees by two special postulates. A further postulate guarantees that if A is an idea, then $-A$ is an idea. Segner finally states two axioms that express conditions satisfied by some of the operations:

Axiom I: If A and B are opposite ideas, then there is no idea C such that $C = AB$.

Axiom II: If A contains B, then $AB = A$.

In this simple system (somewhat simplified here) Segner derives a number of propositions (either problems or theorems) strictly connected to his calculus. Among the most important are the following:

1. Given an idea A, by abstracting some of its contents, find an idea B such that $B > A$.
2. If, given two ideas A and B, there is an idea AB. different from either A or B, then $A \times B$.
3. Given a universal idea A and its coordinate idea B, $AB < A$ and $AB < B$.
4. No infinite idea can be inferior or identical to a finite idea.
5. The defined relations among ideas are exhaustive and reciprocally exclusive: They are not reducible one to the other.

A number of theorems establish how the relation (and therefore the sign) between two ideas changes if one of them is replaced by its opposite:

6. If $A = B$, then $-A = -B$.
7. If $A < B$, then $-A > -B$.
8. If $A > B$, then $A \times -B$.
9. If $A \times B$, then $-A \times B$; if $-A \times B$, then either $-A \times B$ or $A > B$.

Further theorems give an exhaustive list of valid syllogisms:

10. If $A = B$ and $C = B$, then $C = A$.
11. If $A = B$ and $C > B$, then $C > A$.
12. If $A = B$ and $C < B$, then $C < A$.
13. If $A = B$ and $C \times B$, then $C \times A$.
14. If $A > B$ and $C < B$, then $C < A$.
15. If $A > B$ and $C \times B$, then either $C < A$ or $C \times A$.
16. If $A > B$ and $C > B$, then C is consentient with A (i.e., is not opposite).

Segner also proves a theorem that singles out all invalid syllogisms. Then he pays attention to some nonsyllogistic inferences and claims that "they shine of their own light," whether or not we can give them syllogistic form. In particular, he proves the following inferences by composition:

17. If $A = B$ and $C = D$, then $AC = BD$.
18. If $A = B$ and $C < D$, then $AC < BD$.
19. If $A < B$ and $C < D$, then $AC < BD$.

Segner then proves the following theorem concerning inferences by abstraction:

20. If A is consentient with B, if C has been abstracted from A, and if D has been abstracted from B (so that $C > A$ and $D > B$), then C is consentient with D.

Thanks to a number of further definitions and three further propositions, Segner applies his system to the verbal expressions of common logic (he also pays attention to singular ideas and strictly particular propositions whose subjects bear the prefix "only"):

"All A are B" means either $A = B$ or $A < B$.

"No A is B" means only $A < -B$.

"Some A is B" means either $A = B$ or $A < B$ or $A > B$ or $A \times B$.

"Some A is not B" means either $A < -B$ or $A \times -B$.

Segner's work proves that not all attempts to construct a logical calculus in an intensional perspective were destined to failure. Segner succeeds where Bernoulli failed (see section 7) because he never assumes that every idea has a (Port-Royalist) comprehension made of necessary attributes that cannot be modified without destroying it. Therefore, given two ideas, he simply considers three possible cases: Provided that A and B are not opposite, either A contains B, and then $A < B$, or A is contained in B, and then $A > B$, or neither idea contains the other, and then $A \times B$. Thus Segner, unlike Bernoulli (apparently unknown to him), is under no obligation to use nouns of ordinary language as signs of ideas so as to recall their unchangeable comprehension, but uses literal symbols. When he considers the expressions of idea-relations in verbal propositions, he is under no obligation to change his intensional perspective and to consider the extensions of ideas whenever a predicate is not contained in the comprehension of a subject. We can affirm a predicate B of a subject A even if they are coordinate ideas, that is, ideas whose relation is, by definition, a relation at one time of consent and of noncontainment.

Some interpreters have suggested that Segner is similar to Leibniz and even a "disciple of Leibniz" (Vailati 1899, 88). Actually, Segner and Leibniz differ at least inasmuch as Segner's relation of coordination—being a relation of noncontainment—does not respond to the Leibnizian predicate-in-notion criterion. But there certainly are striking similarities. In the *Specimen calculi universalis* and in its *Addenda*, Leibniz uses an algebraic notation and an intensional perspective (see section 8). Moreover, what for Leibniz is a per se true proposition of one of the forms "ab is a," "ab is b," or "a is a," has a detailed treatment in Segner. According to Segner, "A is B" is an affirmation that can rest on one of the following relations between A and B, all of which produce truths:

$AB = A$, if $A < B$ (according to Axiom I),

$AB < A$ or $AB < B$ if $A \times B$ (no. 3 in the list of Segner's propositions).

Equally, in his list of valid syllogisms (10–16), Segner includes Leibniz's per se valid conclusion ("if a is b and b is c, then a is c"). As to Leibniz's principle by which the repetition of a term is irrelevant, Segner also maintains that "the idea of the subject composed with itself cannot produce a new idea"

(Segner [1740], 149). Also Leibniz's *praeclarum theorema*, "if A is B, and D is C, then AD will be BC" (see section 8) has an elaborate equivalent in Segner's inferences by composition. Other similarities can be found if one compares Segner's logic with the *Generales Inquisitiones*, notably with respect to the calculus with negative terms (on the problems posed by the use of negative or infinite predicates instead of negative propositions, and on Segner's skilful solution, see Capozzi 1990, clx–clxiv).

Segner did not know these Leibnizian doctrines, but the undeniable similarities we have stressed are not due to a miracle. They have an explanation in the fact that Segner, like Leibniz, was able to unify a variety of existing doctrines in a single system, depending on his practice of mathematics and using only an intensional approach. But that means that Segner's logical background, albeit unsupported by knowledge of Leibniz's relevant texts, was rich enough to offer him a firm ground on which to build his calculus. In this respect, the changes that took place in German logic in the first three decades of the eighteenth century show that Segner is representative of the logic of his time and not an inexplicable exception. But as in the case of Leibniz, this does not detract from his merits: It only emphasizes them. Let us consider one of the most interesting features of his system: the five idea-relations. Segner was not the first to consider such relations, but depended on Rüdiger and Reusch. He was well acquainted with Reusch's logic (the *Specimen* is dedicated to him) and he also knew Rüdiger's idea-relations. In an academic dissertation of 1734, discussed by one of his students under his guidance, he expressly quotes Rüdiger's logical work on such matters (Capozzi 1990, xcix). This does not make him less original, for it is due to him that such idea-relations are proved exhaustive and exclusive and are used as part of an adequate calculus.

In this respect, Segner can be compared to the later mathematician Joseph Diez Gergonne (1771–1859). In his *Essai de dialectique rationelle* (Gergonne 1816–17) Gergonne considers five idea-relations using the notion of containment as basic but giving it an extensional interpretation: "the more general notions are said to contain the less general ones, which inversely are said to be contained in the former; from this the notion of relative extension of two ideas originates" (Gergonne 1816–17, 192). This extensional interpretation of the notion of containment is used by Gergonne (1816–17, 200) to classify the relations between two ideas on a par with the circles of Leonhard Euler (1707–1783) (see later in this section) and to designate each of them by a symbol. Two ideas A and B, where A is the less general idea and B is the more general one, can (1) have nothing in common, so that they stand in the relation H; (2) intersect each other, so that they stand in the relation X; (3) coincide, so that they stand in the relation I; (4) be such that A is contained in B, so that they stand in the relation C; (5) be such that that B is contained in A, so that they stand in the relation Ɔ.

Like Segner, Gergonne also gives a correspondence list between the standard A, E, I, O propositions and the relations of ideas expressed by them:

Gergonne	Propositions	Segner
I	All A are B	$A = B$
C		$A < B$
H	No A is B	$A < -B$
I	Some A is B	$A = B$
C		$A < B$
Ɔ		$A > B$
X		$A \times B$
H	Some A is not B	$A < -B$
X		$A \times -B$
Ɔ		

However, Gergonne and Segner differ in that whereas Gergonne states that a proposition expresses the relation C because A is contained in B, Segner would say that A contains B. Moreover, whereas Gergonne uses the symbol H for the opposition of ideas, Segner uses none, since he treats the relation of opposition as containment of the opposite idea. As a matter of fact he admits that "every affirmative proposition can be made negative without change of sense, and vice versa; to this end one has only to replace the attribute by a word which is its exact negation, as one sees from this example: '*Lagrange* is *dead*, *Lagrange* is not *alive*' " (Gergonne 1816–17, 197 n.). However, he does not change negative propositions into affirmative ones by replacing the predicate by its opposite (this makes him forbid the conversion of particular negative propositions, while in Segner's logic they are convertible). Apart from such differences, the systems of Gergonne and Segner are perfectly comparable, which shows that at least at this level, the perspectives of intension and extension can be translated into one another. If anything, Segner's calculus and symbolism would better suit Ernst Schröder (1841–1902) who acknowledged Gergonne as a forerunner of his own work (Schröder [1890–1895] 1966, II, 106). Segner treated negative concepts and used a symbolism more similar to Schröder's than to that of Gergonne's (see Capozzi 1990, clxx–clxxii).

Segner, who was much appreciated as a mathematician, scientist, and physician, did not meet with the success he deserved for his logical work. His calculus and symbolism is mentioned in Hoffbauer ([1792] 1969) who does not follow him completely for he uses the symbol "#" for opposition. More attention was paid by contemporaries to other interesting attempts to build calculi of ideas, notably the attempts made by Gottfried Ploucquet and Johann Heinrich Lambert.

Ploucquet (1716–1790), professor of philosophy at the university of Tübingen, developed a logical calculus in stages, resorting to different symbolisms. In his *Fundamenta* (Ploucquet 1759), he gives a theory of judgment in which he considers not only the quantity of the subject but also the quantity of the predicate. Ploucquet feels justified in doing so by the traditional doctrine of conversion of propositions. He believes that the *conversio simplex* of a universal negative proposition, for example, the conversion of “No man is stone” in “No stone is man,” depends on the tacit assumption that subject and predicate have the same universal quantity, whereas the *conversio per accidens* of a universal affirmative proposition, for example, the conversion of “All men are rational” in “Some rational (being) is man,” depends on the tacit assumption that the subject is quantified universally and the predicate particularly. In this work we also find two diagrams of universal syllogisms in the first figure, and a symbolism for the calculus that Ploucquet will modify in later essays.

In 1763, Ploucquet published two essays: *Methodus tam demonstrandi directe omnes syllogismorum species* (1763a) and *Methodus calculandi in Logicis* (1763b). Both contain his technical results, but in the second and longer of them Ploucquet explains the meaning of his logical investigations. He begins by defining “calculus”: “In the most generally accepted sense calculus is the method to determine unknown things beginning from known things according to constant rules” (Ploucquet 1763b, 31). Then he discusses the possibility of an *ars calculatoria*, capable of application not only to numbers but also to geometrical quantities, forces, and logical concepts. His conclusion is that there is no *ars combinatoria*, or calculus of forms, containing as its parts the calculus of quantities and the calculus of qualities. In his opinion, every calculus must be adjusted to some particular object of investigation since “by nature and according to logical order every calculus comes after the understanding of the matter to which the calculus is applied. Then, if a universal calculus were to be imagined, one would suppose (a kind of) knowledge of things that could not be supposed for any mortal being. He who invents does not begin from a calculus, but from the consideration of things” (Ploucquet 1763b, 36).

There is no doubt that Ploucquet wants to differentiate his position from that of Leibniz. According to Ploucquet, Leibniz simply proposes a new version of an ancient utopian and unattainable project, as his *characteristica combinatoria* does not differ from Lull’s art. To support this judgment, Ploucquet quotes the rather reductive appraisal of Leibniz’s 1666 essay on the *combinatoria* given by Leibniz himself in the *Acta Eruditorum* of 1700 (Ploucquet 1763b, 39–40). In 1714, in a letter to Remond, he also stresses that Leibniz spoke of a mere wish to find a general *speciosa* in which all the truths of reason would be reduced to a kind of calculus. Ploucquet reserves similar criticisms for the *Inventum novum quadrati Logici Universalis* of Johann Christian Lange (Lange 1714) (Ploucquet 1763b, 43).

In Ploucquet’s view, tentative investigations about the possibility of a universal calculus are to be praised because without them we would be even more ignorant. Nevertheless, he considers dealing with a plurality of calculi for each area of knowledge more realistic. Therefore, his efforts are limited to the construction of a calculus capable of making syllogistic inference intuitive and immediately clear, so as to avoid errors (Ploucquet 1763b, 46–47). In his opinion, because a logical calculus pays attention only to the formal structure of thought, it makes it easier to learn and use logic, so that “even uncultivated people and those who do not clearly perceive the strength of the calculus or of the ratiocination, can be educated, given the premises, to find, without fear of error, those conclusions to which they would not come by themselves" (Ploucquet 1763b, 74). Thus, bringing a recurrent motive in the logic of the first German Enlightenment to unexpected consequences, Ploucquet extends to uncultivated people Leibniz's invitation to calculate as a means to put an end to the disputes of scholars.

As to Ploucquet's actual logical construction, he assumes that when we judge, that is, when we compare a subject and a predicate, we understand (*intelligere*) either their identity or their diversity (Ploucquet 1763b, 48). In the first case we have an affirmative judgment and in the second case a negative judgment. Now, an affirmative judgment as conceived by our mind is not an understanding of two things but of one thing only: "When I intuit a round stone and pronounce these words: this stone is round, what I actually think by this proposition is nothing but one notion, i.e. the notion of round stone" (Ploucquet 1763b, 52). This conception holds also for syllogism: "All affirmative syllogisms reduce to a single notion," and "all negative syllogisms reduce to two notions, one of which is diverse from the other" (Ploucquet 1763b, 73–74). As Ploucquet explains in a polemical essay of 1766 against Lambert, this means that when we think of the shining sun there is no distinction between subject and predicate but, when we want to communicate our thought to others, our language requires more signs, and we say "the sun is shining" (Bök [1766], 245). The same applies to syllogism. It is because of the constitution of our language that we need a rigorous logical calculus adequate to our spoken propositions and syllogisms.

In the first of the two essays of 1763, Ploucquet (1763a, 18) declares that all his previous work devoted to the question of finding the correct conclusion from given premises can be replaced by a single *praeceptum*: The two terms of the conclusion must maintain the same extension they have in the premises. This *praeceptum*, once complemented with the traditional rules that veto four terms and two negative premises in a syllogism, eliminates the traditional classification of syllogisms according to figures and moods, as well as the need to reduce all syllogisms to the first figure. By referring to the extension of both subject and predicate in the conclusion, Ploucquet makes his formerly implicit choice of quantifying the predicate explicit, a choice highly praised in the following century by Hamilton (1860–1869, II, 322; Aner 1909, 22ff.).

The following symbolism used by Ploucquet in 1763 is a variant of that used in 1759 and basically adopted until 1782 (Ploucquet [1782]) (there are small differences also between the two 1763 essays). Uppercase letters stand for universal terms; lowercase letters stand for particular terms; juxtaposition of terms, such as AB or Ab, means affirmation; the sign "$>$" between two letters means negation (see Venn [1894], 499). Ploucquet, who takes an extensional point of view, maintains that in an affirmative proposition the predicate is taken particularly. He says that this kind of particularity holds "in a comprehensive sense [*sensu comprehensivo*]" (Ploucquet 1763b, 52). Thus "All men are rational" means "All men are some rational beings," a proposition that in his symbolism can be written as "Mr," where the letters signify the initials of the subject and the predicate. The particularity "in a comprehensive sense" of the predicate does not exclude that there could be other individuals apart from those under consideration. As for negative propositions, they are usually meant as having a universal predicate, but here too, though this may seem absurd in common language, they can have a particular predicate, as is the case with "All men are not some animals," meaning, for instance, that they are not irrational animals. In this logical framework, Ploucquet develops a doctrine of subalternation and conversion that contrary to accepted rules, allows the conversion of particular negative propositions, for "Some A is not B" correctly converts into "No B is some A." As to the syllogistic calculus, all one has to do is to represent the premises by Ploucquet's symbolism and check that the premises are not both negative and that they do not contain four terms. Under these conditions, one can draw the conclusion by deleting the middle term and relating the two remaining terms, taking care that they preserve the same quantity they had in the premises. In this calculus also a syllogism in the first figure of the form "All M are P, no S are M, then no S are P," which is invalid according to traditional doctrines, becomes acceptable, for it can be symbolized as follows: Mp, $S > M$, then $S > p$ (see Menne 1969).

Another author of logical calculi is Johann Heinrich Lambert (1728–1777). This famous and eclectic scientist contributed to the study of language, metaphysics and logic, in addition to important research in the fields of optics, geometry (conic sections, perspective, theory of parallel lines, writing on the latter subject one of the basic texts in the history of non-Euclidean geometries), astronomy (comets), physics, technical applications of his theoretical works, and cosmology.

Lambert's interest in logic dates back to the fifties when he wrote the so-called *Six essays of an art of the signs in logic* [*Sechs Versuche einer Zeichenkunst in der Vernunftlehre*], published only after his death (Lambert [1782–1787], I). It is rather difficult to explain why Lambert decided not to publish these essays at the time that he wrote them, especially as they contain the general outline of his calculus, as well as comments on the nature of definition and on the representation of relations (see Dürr 1945). According to some interpreters (Barone 1964, 88), this decision depended on the fact that as Lambert confessed, in these writings he was attracted to the idea of discovering what "was concealed in the Leibnizian characteristic and in the *ars combinatoria*." While Segner did not enter into these matters and Ploucquet refused the very idea of a universal calculus, Lambert adopted Leibniz's ideal and, given his ignorance of the latter's relevant texts, wanted to pursue Leibniz's end by his own means. Now, a calculus aiming at being both formal and real presupposes an alphabet of simple elements. Therefore, Lambert had to postpone the publication of his early technical results until his philosophical investigation could establish such an alphabet of simple and first concepts which, not containing composition in themselves, could not contain contradictions.

Meanwhile, Lambert wrote the *Neues Organon* (Lambert [1764]), an important and famous book in which, according to the idea of *mathesis universalis*, he devoted himself to searching for the basic concepts that could help insert already acquired knowledge into a rational system and promote new discoveries. The *Neues Organon* consists of four parts: Dianoiology, Alethiology, Semeiotics, and Phenomenology, to which we will refer beginning with Alethiology and ending with Dianoiology.

Alethiology, the doctrine of truth and error, deals with elementary concepts. Lambert gives a list of the latter, including conscience, existence, unity, duration, succession, will, solidity, extension, movement, and force. In Lambert's view, connecting simple concepts produces truths that are not subject to change, and therefore can be considered as "eternal truths." Eternal truths provide a foundation for all a priori sciences, in particular arithmetic, geometry, and chronometry.

Semeiotics studies the relation between sign and meaning, and therefore introduces both a theory of language and the project of a characteristic.

Phenomenology is the doctrine of appearance. Here Lambert, in discussing certainty and its relations to truth and error, also considers the degrees of possible certainty, and the probability of cognitions of which we have no absolute certainty.

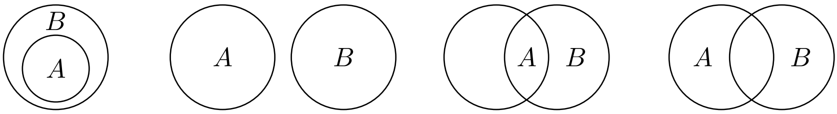
Dianoiology investigates the laws of the understanding. This part of the *Neues Organon* contains diagrams representing concept relations in propositions and syllogisms. Lambert represents a concept—considered in extension, that is, with respect "to all the individuals in which it appears" (Lambert 1764, Dian. §174)—as a line that can either be closed or open. He then represents the relations that two concepts entertain in the four basic propositions of categorical syllogisms:

All A are B	No A is B	Some A is B	Some A is not B
$\ldots B$____$b \ldots$ A____a	A____a B____b	B____b $\ldots A \ldots$	B____b $\ldots A \ldots$

In the diagram representing universal affirmative propositions, what counts, in addition to the length of the lines, is that A is drawn under B. The diagram representing universal negative propositions is clear. As for particular affirmative propositions, the diagram shows that we only know some individuals A that are B, or at least one individual A that is B. Therefore, it remains indeterminate if also all A are B, or even all B are A. In the case of particular negative propositions the diagram shows not only that A is indeterminate but also that A is neither under B nor completely beside it, as in the case of universal negative propositions. On this basis one cannot only immediately make all valid conversions, but also represent all syllogisms, with the advantage of dispensing with the reduction to the first figure (see Wolters 1980, 129–166).

When the *Neues Organon* was published diagrams were no novelty, though Lambert introduced his diagrams before the already mentioned circular diagrams by Leonhard Euler (Euler 1768–1772):

All A are B	No A is B	Some A is B	Some A is not B



In fact, representations of concepts, propositions, and syllogisms by means of circles, lines, and other figures had already been devised by Johann Christoph Sturm (1661), who also introduced circular diagrams representing new syllogisms having negative terms; by Johann Christian Lange (1712, 1714); by Ploucquet (1759); and by Leibniz himself (on the history of diagrams in logic see Gardner 1983; Bernhard 2001).

From what we have seen of Lambert's and Ploucquet's logical work, we can understand why their contemporaries were intrigued by their different approaches to the problem of a logical calculus and wanted to assess their comparative merits. In a public debate in which Lambert and Ploucquet took part directly—reported in Bök ([1766])—Ploucquet's *Methodus calculandi* was compared with Segner's logical work, while Lambert's diagrams in the *Neues Organon* underwent severe criticism. On the occasion of this debate Lambert began a correspondence with Georg Jonathan Holland (1742–1784), a pupil of Ploucquet. In a letter to Holland, Lambert criticized Ploucquet's use of the traditional rule that nothing follows from two negative premises to exclude nonconclusive syllogism. He also criticized Ploucquet for using letters standing for the initials of substantives in syllogisms, thus showing his lack of a true symbolism (Lambert [1782] 1968, 95–96). But in an article of 1765, Lambert, though claiming that his diagrams in the *Neues Organon* were an example of a characteristic, acknowledged that they were only a little thing with respect to his project of a general logical calculus (Bök [1766], 153).

At last, Lambert's logical calculus was published in his *Disquisitio* (Lambert [1765], dated 1765 but actually printed in 1767). Here he states the aim of his calculus and lists the requisites any calculus must satisfy. As to his aim, Lambert says that he wants to find a method for treating qualities similar to the method used in algebra for treating quantities. Just as in algebra we employ the ideas of relation, equality, proportion, and so on, so in the logical calculus we have to employ the ideas of identity, identification, and analogy. As to the requisites a calculus must satisfy, they are the following. (1) For every operation we introduce, there must be the inverse operation, in full analogy with algebra where, when two quantities are added, it is always possible to obtain either by subtracting it from the total. (2) Given the object, the relations and the operations of the logical calculus, an adequate symbolism must be found. The symbols must be a perfect replacement for the things they symbolize to be safely used in their place. This means that we need a characteristic that mirrors things, a real characteristic, in which simple signs stand for simple things and are capable, once composed, to stand for composed things, so that it is also possible to proceed inversely from a composite to its simple elements. (3) Lambert also requires that we have a clear knowledge of the simple elements and the basic relations of the calculus; we must therefore know the combinatorial part of the *ars characteristica combinatoria*. In his opinion, the simple elements are qualities, that is, the special affections of things we can consider as their attributes. Qualities are simple elements because, according to established ontological doctrines, they can be considered as "absolute" attributes, whereas other attributes, notably quantity, must be thought only "relatively" (a similar conception occurs in Leibniz's *De Arte Combinatoria*).

Lambert's calculus in the *Disquisitio*, as it was the case with his unpublished essays of the fifties, is intensional, that is, "does not concern individuals but properties" (letter to Holland 21.4.1765, Lambert [1782] 1968, 37). After trying the extensional perspective in the *Neues Organon*, his return to his early preference for the intension of concepts is undoubtedly due to a conscious choice. For Lambert wants to find what is "simplest" and "first" in concepts, but to obtain what is simpler, it is necessary to consider what is more complex, and in the case of concepts, the more complex concepts are those *containing* the simpler ones. Therefore, it is necessary to consider concepts as properties, as concepts containing other concepts, thus disregarding the class of individuals to which they extend.

When dealing with judgments and syllogisms, Lambert's first aim is to establish the identity of the subject and predicate of judgments. Therefore, given the judgment "All A are B," where A and B are not already obviously identical, Lambert establishes their identity by considering the subject as containing the predicate plus other properties. Hence the following symbolism (Lambert [1765], 461–462):

All A are B	No A is B	Some A is B	Some A is not B
$A = nB$	$A{:}p = B{:}q$	$mA = nB$	$mA{:}p = B{:}q$

In the universal affirmative, $A = nB$, n stands for the qualities which can be found in the subject A but not in the predicate B. In the universal negative, $A{:}p = B{:}q$, the sign ":" stands for a logical division and expresses which qualities, p and q, must be subtracted from the subject and the predicate, because neither belongs both to the subject and the predicate, so as to obtain $A = B$. Similarly, in the particular affirmative, $mA = nB$, and in the particular negative, $mA{:}p = B{:}q$. On this basis, Lambert obtains a general formula expressing any kind of judgment: $A/p = nB/q$ (here Lambert substitutes the sign of fraction for the sign ":"). From this formula one can easily derive a formula expressing any kind of syllogism:

$$\frac{\begin{array}{c} mA/p = nB/q \\ \mu A/\pi = \nu C/\rho \end{array}}{m\nu C/p\rho = \mu n B/\pi q}.$$

To give an example of how this general formula applies to particular syllogisms, a syllogism *Barbara*, whose premises are $B = mA$ and $C = B$, has the conclusion $C = mA$, whereas a syllogism *Celarent*, whose premises are $B/q = A/p$ and $C = \nu B$, has the conclusion $C/q = \nu A/p$ (Lambert [1782–1787], I, 102–103, 107).

Despite the fact that the *Disquisitio*'s treatment of syllogism is very different from that of the *Neues Organon*, it was disappointing for Lambert's most competent readers. Holland sent Lambert a letter in which (beside mentioning his own tentative calculus) he observed that, however good Lambert's calculus was, it did not achieve the declared aim to find symbols mirroring reality. What are A, B, m, n, symbols of? Above all, which are the primitives that they are supposed to be symbols of? A definite answer, Holland concluded, could perhaps be expected from a new work Lambert had announced (Lambert [1782] 1968, 259–266). The work Holland referred to, titled *Architectonic* (Lambert [1771]), was published a few years later. In this treatise, which promised to give a theory of what is simple and first in philosophical and mathematical knowledge, the author collects the results of his philosophical research going back to the mid-forties. But for all its importance as the *summa* of Lambert's thought, the *Architectonic* provides no formal treatment, nor gives a new and complete list of simple elements that could be used as basic elements of a real characteristic, because it contains the same elements already listed in the *Neues Organon*.

The conclusion to be drawn is that Lambert's project shared Leibniz's ambitions and in this respect went far beyond Segner's and Ploucquet's calculi, but perhaps was too ambitious and, though providing interesting details in the application of algebra to logic, can be said to be unachieved. In a sense, Lambert admitted as much in a letter (14.3.1771) to Johann Heinrich Tönnies: "should the universal characteristic belong to the same class as the philosopher's stone or the squaring of the circle, it can at least, just as these, induce other discoveries" (Lambert [1782] 1968, 411).

Ploucquet's refusal of Leibniz's project and Lambert's somber admission to Tönnies may sound too pessimistic if one considers how much they and other eighteenth-century logicians—not to mention Leibniz—had progressed since Bernoulli's failed parallelism. But especially Lambert's assessment of universal characteristic as something similar to the squaring of the circle makes it clear that these authors believed that the construction of a satisfactory logical calculus was hindered by a possibly insurmountable obstacle: the overpowering amount of philosophical analysis to be done in the fields of metaphysics, semiotics, and natural language to reach a suitable alphabet of thoughts. As a matter of fact, unknown to these eighteenth-century logicians, there was an obstacle, not only on the side of philosophical analysis but also on the side of mathematics. Nineteenth-century logicians will find out that one had to reflect also on the nature of mathematics and algebra, especially on their apparently exclusive link with quantity, before an algebra of logic could come to life.

11. Kant

Interest for logical calculi seems to vanish at the end of the eighteenth century. We have mentioned some of the reasons behind this phenomenon, but according to a still widely received opinion this was due to the influence exerted on logic by Immanuel Kant (1724–1804). This opinion is usually justified by saying that Kant introduced confusion in logic through his notion of transcendental logic. As a matter of fact, Kant had a definite concept of logic, related to his transcendental philosophy but not to be confused with it.

To evaluate Kant's concept of logic, one must take into account his 40 years-long activity as a logic teacher, using as a textbook Georg Friedrich Meier's *Auszug aus der Vernunftlehre* (Meier 1752b), a short version of the latter's *Vernunftlehre* (Meier 1752a) (on Meier's logic see Pozzo 2000). We have several texts related to this teaching activity, which constitute the so-called Kantian *logic-corpus*. Apart from the programs of the courses, such texts are (1) Kant's handwritten annotations on Meier's *Auszug* (the so-called logical *Reflexionen*, in Kant 1900, XVI), (2) lecture notes taken by students (Kant 1900–, XXIV; Kant 1998a, 1998b), and (3) *I. Kant's Logik*, a book published in 1800 by Gottlob Benjamin Jäsche by collecting a selection of Kant's annotations on Meier's *Auszug* with Kant's consent (Kant 1900–, IX, 1–150). These texts must be used with care and must always be compared with Kant's published production. Nonetheless, they are essential to assess his views on logic, allowing a deeper insight into the importance of logic for Kant's philosophy, and testifying to his knowledge of the discipline he was due to teach. As it is impossible to give details here of Kant's treatment of logical doctrines, we will only discuss his general concept of logic.[2]

A comparative study of the Kantian *logic-corpus* shows that Kant's concept of logic is the result of a sustained effort of reflection lasting several years. He began as an almost orthodox Wolffian, founding logic on empirical psychology and ontology (*Logik Blomberg*, Kant 1900–, XXIV, 28). In his mature conception, however, he took the opposite view and denied that logic could be founded on either empirical psychology or ontology.

To this effect Kant argues that a logic founded on empirical psychology could describe human logical behavior but not prescribe laws to it. In his opinion, logical rules do not mirror what we actually do when we think, but are the standard to which our thoughts must conform if they are to have a logical form: Logic considers "not how we do think, but how we ought to think"

(*I. Kant's Logik*, Kant 1900–, IX, 14). As to the formerly accepted foundation of logic on ontology, Kant simply suppresses it, to the dismay of many of his contemporaries and later idealist philosophers. In particular, opposing Kant, Hegel proposed a new logic identical to metaphysics which, like old metaphysics, would admit that "thought (with its immanent determinations) and the true nature of things are one and the same content" (Barone 1964, 202). Thus it is rather surprising that William and Martha Kneale claim that it was Kant "with his transcendentalism who began the production of the curious mixture of metaphysics and epistemology which was presented as logic by Hegel and other Idealists of the nineteenth century" (Kneale and Kneale 1962, 355).

The independence of logic from ontology and empirical psychology raises the problem of the origin and justification of logic. Kant gives an indirect answer to the problem of the origin of logic by way of a comparison of logic with grammar. Logic and grammar—he maintains—are similar in as much as we learn to think and to speak without previous knowledge of grammatical and logical rules, and only at a later stage we become conscious of having implicitly used them. Nonetheless logic and grammar differ because, as soon as we become aware of grammatical rules, we easily see that they are empirical, contingent, and subject to variations. On the contrary, once we become conscious of the logical structure of our thought, we cannot fail to appreciate that without that structure we could not have been thinking at all. Therefore logic precedes and regulates any rational thinking and is necessary in the sense that we cannot consider it contingent and variable. Kant concludes that logic "is abstracted [*abstrahirt*] from empirical use, but is not derived [*derivirt*]" from it (*Reflexion* 1612, Kant 1900–, XVI, 36) so that it can be considered a *scientia scientifica*, whereas grammar is only a *scientia empirica* (*Logik Busolt*, Kant 1900–, XXIV, 609). This is important because the logic considered by Kant is not a natural logic that could be investigated by psychology, but is an "artificial logic."

This being the origin of logic, its justification can be reduced to the fact that, according to Kant, logical principles such as the law of contradiction are accepted without proof: "All rules that are logically provable in general are in need of a ground [*Grund*] from which they are derived. Many propositions (e.g. that of contradiction) cannot be proved at all, neither *a priori* nor empirically" (*Logik Dohna-Wundlacken*, Kant 1900–, XXIV, 694). In other words, logical rules, "once known, are clear by themselves" (*Reflexion* 1602, Kant 1900–, XVI, 32).

This means that logic not only is necessary, scientific, and a priori, but also is capable of justifying itself. These features make logic one of the means Kant uses in carrying through his philosophical project of explaining the possibility of experience according to his Copernican revolution. An important part of this project consists in showing that it is possible to find all the general forms of thought—categories—without having to fall back on metaphysics or experience. Now logic (rather, one of its most important parts, i.e., the functions of judgment), which Kant has made no longer dependent on empirical psychology and ontology, qualifies as the perfect clue to the categories. But since categories have to be completely enumerated to be employed in a complete list of principles of the understanding, which define the field of possible experience, logic has to satisfy a further requisite: It has to be complete. Hence Kant's well-known statement that logic "seems to all appearance to be finished and complete" (Kant 1997, B viii). Kant has been criticized for this statement, and in our opinion he lacks conclusive arguments to support it. But one must consider that a proof of the completeness of logic would have been easy if Kant had preserved the foundation of logic on empirical psychology and ontology, both ultimately guaranteed by God. It is also fair to point out that Kant envisages the possibility, for a closed system, of growing "from within," on a par with living organisms that grow with no addition of new parts (Kant 1997, A 832/B 860). Applying this to logic, one could say that some growth in logic is possible, although within the boundaries of a systematic structure.

The scientific, necessary, and self-justifying nature of logic guarantees that it has great autonomy and the maximum spectrum of application. Such prerogatives are counterbalanced by precise limitations: "Nobody can dare to judge of objects and to assert anything about them merely with logic without having drawn on antecedently well-founded information about them from outside logic" (Kant 1997, A 60/B 85). Consequently, logic is the supreme canon of truth with respect to the formal correctness of thought, but must be indifferent to its contents. In this way Kant makes his concept of logic more definite. Logic, having no specific subject matter, is *general*. Having nothing to do with human psychology, it is *pure*. Concerning only the form of thought, it is merely *formal*.

The first consequence of this conception is that logic has to be analytic, although not in the sense that it deals with analytic judgments only. For logic is not concerned with the analytic/synthetic distinction which is left to transcendental logic: "The explanation of the possibility of synthetic judgments is a problem with which general logic has nothing to do, indeed whose name it need not even know" (Kant 1997, A 154/B 193). Logic is analytic in two senses. (1) "General logic analyzes the entire formal business of the understanding and reason into its elements, and presents these as principles of all logical assessment of our cognition" (Kant 1997, A 60/B 84). (2) Logic is analytic inasmuch as it has nothing to do with dialectic, both intended as the rhetorical art of disputation and as the part of logic dealing with probability.

The most evident and better known reason for Kant's separation of logic from dialectic is the connection between dialectic and rhetoric. A rhetorical dialectic is an art for deceiving adversaries in a dispute and for gaining consent not only disregarding truth but also purporting to produce the semblance [*Schein*] or illusion of truth. Kant condemns this kind of "logic" as unworthy of a philosopher (see *I. Kant's Logik*, Kant 1900–, IX, 17). As to the association of dialectic with probability, it goes back to the distinction made by Aristotelian logical treatises between analytic, that is, the part of logic dealing with truth and certainty, and dialectic, that is, the part of logic dealing with what is probable, according to Boethius's translation of the Greek *éndoxos* with the Latin *probabile*. This distinction was adopted by many eighteenth-century logicians, notably by Meier who divides logic into *analytica* or "logic of completely certain erudite cognition," and *dialectica* or *logica probabilium*, defined as "logic of probable erudite cognition" (Meier 1752b, §6, in Kant 1900–, XVI, 72).

Like many philosophers (including Leibniz), till the early seventies Kant hoped that a general logic of the uncertain could be found. Such a logic, although different from the Aristotelian and humanist doctrines of probability and attentive to the late seventeenth-century results in the field of probability calculus, was intended to be capable of also dealing with qualitative matters concerning justice, politics, and so on. Later on, Kant completely changed his view. He considered probability as a measurable degree of certainty—in this he agreed with Wolff—which "can be expressed like a fraction, where the denominator is the number of all possible cases, the numerator is the number of actual cases" (*Logik Pölitz*, Kant 1900–, XXIV, 507). This view restricts probability (*Wahrscheinlichkeit*, *probabilitas*) to matters that can be subjected to a numerical calculus (games of chance and statistically based events such as mortality indexes), and excludes the possibility of an instrumental art for weighing, rather than numbering, heterogeneous reasons pro and contra a given qualitative question. Against this alleged art Kant objects that it concerns the notion of "verisimilitude" (*Scheinbarkeit*, *verisimilitudo*) rather than probability. In his view, if such an art, under the name of dialectic, belonged to logic, the latter would no longer be a canon of truth but would become an instrument for producing an illusion of truth by assigning an alleged probability—in fact a mere verisimilitude—even to questions that are beyond possible experience, such as the existence of the soul. Hence Kant's claim that only probability restricted to matters that can be subjected to a numerical calculus is worthy of this name and, because it is contiguous to truth and certainty, belongs to the analytic part of logic and need not be dealt with in a special part of logic called dialectic (Kant 1900–, A 293/B 349). Kant's position is drastic: Logic and dialectic must part and go separate ways.

The second consequence of Kant's view that logic is a mere formal canon of truth is that the content of logic must be limited to the doctrine of elements: concepts, judgments, and inferences. Therefore, logic must not trespass into the domains of anthropology and psychology, nor give advice for the use of logic in the fields of the natural sciences or of practical life. This means that Kant breaks away from one of the main trends of European logic, which had tried to give new life to the discipline by stressing its usefulness either as a guide for judging, or as a kind of methodology for empirical research, or as a medicine against errors, or as an epistemological exercise. In particular, Kant breaks away from Locke's view of logic, despite the fact that he had formerly praised it, for he maintains that the study of the origin of concepts "does not belong to logic, but rather to metaphysics" (*Logik Dohna-Wundlacken*, Kant 1900–, XXIV, 701).

The fact that Kant separates logic from epistemology does not mean that the texts of the Kantian *logic-corpus* do not contain epistemological parts. On the contrary, these texts make very interesting reading on matters such as opinion, belief, knowledge, hypotheses, probability, and so on. But these matters are no longer intended as belonging to pure logic because, to deal with them, one must take into account the content of knowledge and the human cognitive constitution, including sensibility, or at least the form of sensibility, as well as practical aspects of human action, such as the interest we have for accepting something as true. If Kant had written a logic handbook himself, he probably would have treated such matters at length, in addition to other interesting questions, such as the doctrine of logical essence, in a doctrine of method.

The third consequence of Kant's view of logic is that it is only a canon for checking the correctness of our thoughts but is incapable of invention. Kant's sharp distinction between logic and mathematics contributes to this view. He agrees with Wolff that there is a single logic to be complied with by mathematicians and nonmathematicians alike, but logic is insufficient to explain why mathematics is ampliative. According to Kant, mathematics is the science that constructs a priori its concepts, that is, exhibits a priori the intuitions corresponding to them. Thus, mathematics relies also on the form of sensible intuition, so that it has content and can be inventive with respect to it. This applies not only to arithmetic and geometry but also to algebra, which is inventive because it refers (albeit mediately) to a priori intuitions. Therefore, Kant rejects the view of those who "believe that logic is a heuristic (art of discovery) that is an organ of new knowledge, with which one makes new discoveries, thus e.g. algebra is heuristic; but logic cannot be a heuristic, since it abstracts from any content of knowledge" (*Logik Hechsel*, Kant 1998b, 279 = ms. 9).

These statements are not borne out of ignorance. Kant knew the outlines of Leibniz's *ars characteristica combinatoria*, on whose utopian nature he commented in an essay of 1755 (*Nova dilucidatio*, Kant, 1900–, I, 390) in terms that seem to anticipate analogous statements by Ploucquet and Lambert. Moreover, his *logic-corpus*, as well as his works and correspondence, provide evidence that (1) he was well acquainted with the combinatorial calculus of syllogistic moods; (2) he used Euler's (whom he quotes) circular diagrams to designate concepts, judgments, and syllogisms; (3) he knew the linear diagrams of Lambert, with whom he corresponded; (4) he probably had some knowledge of Segner's and Ploucquet's works; and (5) he actively promoted the diffusion of Lambert's posthumous works containing the latter's algebraic calculus. But all this did not shake his conviction that an algebraic symbolism of idea-relations and the use of letters instead of words are not by themselves a means to invention.

If we consider the development of logic from humanism onward, we see that one of the basic motivations of logical research in the whole period was the demand to make logic inventive. The humanist theories of *inventio*, Bacon's studies on induction, Descartes's theory of problem solving, the *ars inveniendi* and a large part of Leibniz's *ars characteristica combinatoria*, and so on, can be viewed in this perspective. Kant objected that this kind of research, while claiming to be purely formal, was meant to deal with the content of knowledge. The condition (and cost) of his objection to an inventive logic was the separation of logic from mathematics, but in this way he achieved his aim of separating *pure* logic from metaphysics and psychology, as well as from any transcendent foundation. This aspect of Kant's concept of logic reappears in the philosophy of logic of some later logicians. Thus, despite substantial differences, Frege's concept of logic seems indebted to Kant's in several respects, such as the idea that the only logic that really counts is scientific logic, rather than some natural logic; the contention that a scientific or artificial logic provides necessary and universal rules; the condemnation of any intrusion of psychology into logic by the argument that logic is normative on a par with moral laws; the idea that logic is used for justifying knowledge rather than for acquiring new knowledge. But even Venn, who claims that Kant had "a disastrous effect on logical method" (Venn [1894], xxxv) begins his own system of logic by stating: "Psychological questions need not concern us here; and still less those which are Metaphysical" (Venn [1894], xxxix). Perhaps it would have been more difficult for him to make such a statement if Kant had not already made that very same statement.

Notes

1. Reference to secondary literature devoted to Leibniz's notion of complete concept could span over many pages. We will limit ourselves to the seminal papers by Mondadori (1973) and Fitch (1979), to the discussion included in Mates (1986), and—for two recent accounts based on different interpretations—to Zalta (2000) and Lenzen (2003). Mondadori's and Fitch's papers are included, together with other relevant contributions, in Woolhouse (1993).

2. The body of literature on Kant is enormous, and also literature on Kantian logic is very extensive, ranging from the relation between general and transcendental logic to the doctrines of concepts, judgments, and inferences, not to mention topics such as the relations between logic and language and mathematics. We will mention only Shamoon (1981), Capozzi (1987), Pozzo (1989), Brandt (1991), Reich (1992), Wolff (1995), Capozzi (2002), and Capozzi (forthcoming) (the latter containing an extensive bibliography). A precious research tool is provided by an impressive Kantian lexicon, still in progress, many of whose volumes are devoted to Kant's logic-corpus (Hinske 1986–).

References: Primary Sources

Agricola, Rodolphus. [1479] 1967. *De inventione dialectica*. Reprint of the ed. Köln 1539, with Alardus's commentary. Frankfurt: Minerva.

Agricola, Rodolphus. [1479] 1992. *De inventione libri tres. Drei Bücher über die Inventio dialectica*, ed. L. Mundt. Tübingen: Niemeyer.

Aldrich, Henry. [1691] 1849. *Artis Logicae Compendium*. Oxford. Edited by H. L. Mansel as *Artis Logicae Rudimenta*. Oxford: Hammans.

Aquinas, Thomas. 1888–1889. *Opera omnia iussu impensaque Leonis XIII P. M. edita, t. 4–5: Pars prima Summae theologiae*. Rome: Typographia Polyglotta S. C. de Propaganda Fide.

Argens, Jean Baptiste d'. 1737. *La philosophie du bon sens, ou réflexions philosophiques sur l'incertitude des connoissances humaines, à l'usage des cavaliers et du beau-sexe*. London.

Arnauld, Antoine, and Claude Lancelot. [1676] 1966. *Grammaire Générale et raisonnée*, critical ed. of the 1676 edition by Herbert E. Brekle (1st ed. 1660). Stuttgart-Bad Cannstatt: Frommann-Holzboog.

Arnauld, Antoine, and Pierre Nicole. [1683] 1981. *La Logique ou l'art de penser*, critical ed. of the 5th ed., Paris 1683, eds. P. Clair and F. Girbal (1st ed. 1662). Paris: Vrin. English trans. in Arnauld and Nicole 1996.

Arnauld, Antoine, and Pierre Nicole. 1996. *Logic or the Art of Thinking.* Trans. and ed. J. Vance Buroker. Cambridge: Cambridge University Press.

Arnauld, Antoine, and Pierre Nicole. 1704. *Logica sive Ars cogitandi.* Trans. J. Fr. Budde. Halle.

Bacon, Francis. [1620] 2000. *The New Organon.* Trans. L. Jardine and M. Silverthorne. Cambridge: Cambridge University Press.

Becher, Johann Joachim. 1661. *Character pro notitia linguarum universali.* Frankfurt: Ammon and Serlin.

Beck, Cave. 1657. *The Universal Character, by which all the Nations in the World may Understand One Another's Conceptions. . . .* London: Maxey and Weekley.

Bernoulli, Jakob. [1685] 1969. *Parallelismus ratiocinii logici et algebraici.* In *Die Werke von Jakob Bernoulli*, vol. I, ed. Naturforschende Gesellschaft in Basel. Basel: Birkhäuser. English trans. in Boswell (1990), 211–224.

Bernoulli, Jakob. [1713] 1975. *Ars conjectandi.* In *Die Werke von Jakob Bernoulli*, vol. III, ed. Naturforschende Gesellschaft in Basel. Basel: Birkhäuser.

Bök, August Friedrich, ed. [1766] 1970. *Sammlung der Schriften, welche den logischen Calcul Herrn Prof. Ploucquets betreffen, mit neuen Zusätzen.* Reprint, ed. A. Menne. Stuttgart-Bad Cannstatt: Frommann-Holzboog.

Budde, Johann Franz. 1731. *Compendium historiae philosophiae observationibus illustratum*, ed. Johann Georg Walch. Halle.

Burgersdijk, Franco. [1626] 1645. *Institutionum logicarum libri duo.* Leiden: Officina Commelini.

Caesarius, Johannes. [1520] 1545. *Dialectica.* Mainz: Schoeffer.

Clauberg, Johann. [1654] 1658. *Logica vetus et nova.* Amsterdam: Officina Elzeviriana.

Condillac, Étienne Bonnot de. [1780] 1948. *La Logique ou les premiers développements de l'art de penser.* In *Oeuvres philosophiques de Condillac*, ed. G. Le Roy, vol. 2, 368–416. Paris: Presses Universitaires de France.

Condillac, Étienne Bonnot de. [1798] 1948. *La langue des calculs.* In *Oeuvres philosophiques de Condillac*, ed. G. Le Roy, vol. 2, 418–529. Paris: Presses Universitaires de France.

Crousaz, Jean Pierre de. 1724. *Logicae systema.* Genève.

Dalgarno, George. [1661] 1961. *Ars Signorum.* London: Hayes. Reprint Menston: Scholar Press.

Descartes, René. 1964–1976. *Ouvres de Descartes*, ed. C. Adam and P. Tannery. Paris: Vrin.

Descartes, René. [1637] 1994. *Discours de la méthode. A bilingual edition and an interpretation of René Descartes' philosophy of method*. Ed. and trans. G. Heffernan. Notre Dame: Notre Dame University Press.

Duncan, William. [1748] 1968. *Elements of Logick*. London. Reprint Hildesheim: Olms.

Eberstein, Wilhelm Ludwig Gottlob von. [1794–1799] 1970. *Versuch einer Geschichte der Logik und Metaphysik bey den Deutschen von Leibniz bis auf den gegenwärtige Zeit*, 2 vols. Halle. Reprint Bruxelles: Culture et Civilisation.

Euler, Leonhard. [1768–1772] 1911. *Lettres à une Princesse d'Allemagne sur divers sujets de physique et de philosophie*, St. Petersburg. In *Leonhardi Euleri Opera omnia*, ed. Soc. scient. Nat. Helv. Leipzig: Teubner.

Fabri, Honoré. 1646. *Philosophiae tomus primus*. Lyon: Champion.

Fonseca, Pedro da. [1564] 1964. *Instituições dialécticas. Institutionum dialecticarum libri octo*. Introdução, estabelecimento do texto, tradução e notas por J. Ferreira Gomes. 2 vols. Coimbra: Universidade de Coimbra, Instituto de Estudos Filosóficos.

Gergonne, Joseph Diez. 1816–1817. Essai de dialectique rationelle. *Annales des Mathématiques pures et appliquées*, 7, 189–228.

Geulincx, Arnold. 1891–1893. *Opera Philosophica*, ed. J. P. N. Land. The Hague: Nijhoff.

Hobbes, Thomas. [1839] 1961. *Elementorum philosophiae sectio prima. De corpore. Pars prima, Computatio sive logica*. London, 1st ed. 1655. In *Thomae Hobbes Malmesburiensis Opera Philosophica quae latine scripsit omnia*, ed. G. Molesworth, vol. 1. Reprint, Aalen: Scientia Verlag.

Hoffbauer, Johann Christoph. [1792] 1969. *Analytik der Urtheile und Schlüsse. Mit Anmerkungen meistens erläuternden Inhalt*. Halle. Reprint Bruxelles: Culture et Civilisation.

Hutcheson, Francis. 1759. *Logicae compendium*. Glasgow.

Izquierdo, Sebastian. 1659. *Pharus Scientiarum*. Leiden: Bourgeat and Liétard.

Jungius, Joachim. [1638] 1957. *Logica Hamburgensis*, ed. R. W. Meyer. Hamburg: Augustin.

Jungius, Joachim. 1977. *Logica Hamburgensis Additamenta*, ed. W. Risse. Göttingen: Vandenhoeck & Ruprecht.

Kant, Immanuel. 1900–. *I. Kant's gesammelte Schriften*, ed. Königlich Preussische Akademie der Wissenschaften (and successors). Berlin: Georg Reimer. Later publisher Berlin–Leipzig: Walter de Gruyter.

Kant, Immanuel. 1992. *Lectures on Logic*. Trans. M. Young. Cambridge: Cambridge University Press (*Logik Blomberg, Wiener Logik, Logik Dohna-Wundlacken, Logik. Ein Handbuch zu Vorlesungen*, bearing on the margin the number of the pages of Kant 1900–; also contains a partial translation of *Logik Hechsel* in Kant 1998b).

Kant, Immanuel. 1997. *Critique of Pure Reason*. Trans. P. Guyer and W. Wood. Cambridge: Cambridge University Press. (References are to the pages of the 1781 (A) and 1787 (B) edition.)

Kant, Immanuel. 1998a. *Logik-Vorlesung: unveröffentlichte Nachschriften* I, *Logik Bauch*, ed. T. Pinder. Hamburg: Meiner.

Kant, Immanuel. 1998b. *Logik-Vorlesung: unveröffentlichte Nachschriften* II, *Logik Hechsel; Warschauer Logik*, ed. T. Pinder. Hamburg: Meiner.

Keckermann, Bartholomaeus. 1614. *Opera omnia quae extant*. Köln: Aubertus.

Kircher, Athanasius. 1663. *Polygraphia nova et universalis ex combinatoria arte detecta*. Rome: Varesius.

Lambert, Johann Heinrich. [1764] 1965. *Neues Organon*, 2 vols. Leipzig. Reprint in J. H. Lambert, *Philosophische Schriften*, vols. 1–2, ed. H. W. Arndt. Hildesheim: Olms.

Lambert, Johann Heinrich. [1765] 1981. De universaliori calculi idea Disquisitio, una cum adnexo specimine. *Nova Acta Eruditorum*, 441–473. Reprint in microform, Hildesheim–New York: Olms.

Lambert, Johann Heinrich. [1771] 1965. *Anlage zur Architectonic oder Theorie des Einfachen und Ersten in der philosophischen und mathematischen Erkenntnis*, 2 vols. Riga. Reprint in J. H. Lambert, *Philosophische Schriften*, vols. 3–4, ed. H. W. Arndt. Hildesheim: Olms.

Lambert, Johann Heinrich. [1782–1787] 1967–1969. *Logische und philosophische Abhandlungen*, 2 vols. Berlin. Reprint in J. H. Lambert, *Philosophische Schriften*, vols. 6–7. Hildesheim: Olms.

Lambert, Johann Heinrich. [1782] 1968. *Deutscher gelehrter Briefwechsel*. Berlin. Reprint in J. H. Lambert, *Philosophische Schriften*, vol. 9.1. Hildesheim: Olms.

Lange, Johann Christian. 1712. *Nucleus logicae Weisianae . . . variis addimentis novis . . . auctus et illustratus*. Giessen.

Lange, Johann Christian. 1714. *Inventum novum quadrati logici universalis, in trianguli quoque formarum commode redacti*. Giessen.

Le Clerc, Jean. 1692. *Logica sive ars ratiocinandi*. London: A. and J. Churchill.

Leibniz, Gottfried Wilhelm. [1686] 1982. *Generales Inquisitiones de Analysi Notionum et Veritatum*, ed. and German trans. F. Schupp. Hamburg: Meiner.

Leibniz, Gottfried Wilhelm. [1860] 1971. *Briefwechsel zwischen Leibniz und Christian Wolff*, ed. C. I. Gerhardt. Halle: [s.n.]. Reprint Hildesheim–New York: Olms.

Leibniz, Gottfried Wilhelm. [1875–1890] 1978. *Die philosophischen Schriften*, 7 vols., ed. C. I. Gerhardt. Berlin: Weidmannsche Buchhandlung. Reprint Hildesheim–New York: Olms.

Leibniz, Gottfried Wilhelm. [1903] 1966. *Opuscules et fragments inédits de Leibniz*, ed. L. Couturat. Alcan: Paris. Reprint Hildesheim–New York: Olms.

Leibniz, Gottfried Wilhelm. 1923–. *Sämtliche Schriften und Briefen*. Herausgegeben von der Preussischen (Deutschen) Akademie der Wissenschaften zu Berlin. Darmstadt–Leipzig–Berlin: Akademie.

Leibniz, Gottfried Wilhelm. 1973. *Philosophical Writings*, ed. G. H. R. Parkinson, trans. M. Morris and G. H. R. Parkinson. London: J. M. Dent and Sons.

Locke, John. [1690] 1975. *An Essay concerning Human Understanding*, ed. P. H. Nidditch. Oxford: Clarendon Press.

Locke, John. [1706] 1993. *Of the Conduct of the Understanding*. In *Posthumous Works*. London. Reprint Bristol: Thoemmes Press.

Locke, John. [1755] 1996. *Johann Lockens Anleitung zur Erkäntniß der Wahrheit nebst derselben Abhandlung von der Wunderwerken*. Königsberg: Hartnung. Reprint (eds. T. Boswell, R. Pozzo, and C. Schweiger). Stuttgart-Bad Cannstatt: Fromman-Holzboog.

Malebranche, Nicolas. [1674] 1962. *De la recherche de la vérité*. In *Oeuvres complètes*, t. I–III, ed. G. Rodis-Lewis. Paris: Vrin.

Meier, Georg Friedrich. 1752a. *Vernunftlehre*. Halle.

Meier, Georg Friedrich. [1752b] 1914. *Auszug aus der Vernunftlehre*. Halle. Reprint in Kant 1900–, vol. XVI.

Melanchthon, Philipp. 1520. *Compendiaria dialectices Ratio*. Wittebergae: Melchior Lotter.

Melanchthon, Philipp. 1528. *De Dialectica libri IV.* Haganoe: Secerius.

Melanchthon, Philipp. [1547] 1846. *Erotemata dialectices.* In *Corpus Reformatorum*, vol. XIII, ed. C. G. Bretschneider. Halis Saxonum: Schwetschke.

Pascal, Blaise. [1658 or 1659] 1914. *De l'ésprit géométrique.* In *Oeuvres de Blaise Pascal*, vol. IX, 229–294, eds. L. Brunschvicg, P. Boutroux, and F. Gazier. Paris: Hachette.

Petrarca, Francesco. 1933–42. *Epistolae familiares*, ed. V. Rossi and U. Bosco. Firenze: Sansoni.

Ploucquet, Gottfried. 1759. *Fundamenta Philosophiae Speculativae.* Tübingen.

Ploucquet, Gottfried. 1763a. *Methodus tam demonstrandi directe omnes syllogismorum species, quam vitia formae detegendi, ope unius regulae.* Tübingen. In Bök [1766] 1970, 17–28.

Ploucquet, Gottfried. 1763b. *Methodus calculandi in Logicis inventa a G. P. Praemittitur commentatio de arte characteristica.* Frankfurt–Leipzig. In Bök [1766] 1970, 31–80.

Ploucquet, Gottfried. [1782] 2006. *Expositiones Philosophiae Theoreticae.* Stuttgart. Reprint, ed. and German trans. of the logical part by M. Franz. Hildesheim: Olms.

Ramus, Petrus. [1543] 1964. *Dialecticae Institutiones—Aristotelicae Animadversiones.* Repr. with Introduction by W. Risse. Stuttgart: Fromann.

Ramus, Petrus. [1555] 1996. *Dialectique*, ed. N. Bruyère. Paris: Vrin.

Reusch, Johann Peter. [1734] 1990. *Systema logicum antiquiorum atque recentiorum item propria praecepta exhibens*, Jena. In C. Wolff, *Gesammelte Werke*, Sect. III, vol. 26. Hildesheim: Olms.

Rüdiger, Andreas. 1722. *De sensu veri et falsi libri IV.* Halle (2nd ed.; 1st ed. 1709).

Saccheri, Gerolamo. [1697] 1980. *Logica demonstrativa.* Reprint of the ed. Torino 1697 (ed. W. Risse). Hildesheim: Olms.

Saccheri, Gerolamo. [1733] 1986. *Euclides ab Omni Naevo Vindicatus*, Milano, ed. and trans. G. B. Halsted. New York: Chelsea.

Sanderson, Robert. [1615] 1985. *Logicae artis compendium.* Reprint of the ed. Oxford 1618 (ed. J. Ashworth). Bologna: CLUEB.

Schott, Kaspar. 1664. *Technica Curiosa VII—Mirabilia graphica, sive nova aut rariora scribendi artificia.* Nürnberg: Endter.

Schröder, Ernst. [1890–1895] 1966. *Vorlesungen über die Algebra der Logik*, 3 vols. Leipzig. Reprint New York: Chelsea.

Segner, Johann Andreas. [1740] 1990. *Specimen logicae universaliter demonstratae.* Jena. *Appendices: Two dissertations De syllogismo.* Reprint (ed. M. Capozzi). Bologna: CLUEB.

Seton, John. 1545. *Dialectica.* London: Berthel.

Simon of Faversham. 1968. *Commentary on Peter of Spain's "Tractatus,"* ed. by L. M. De Rijk. In *On the genuine text of Peter of Spain's "Summule Logicales". Part II. Simon Faversham (d. 1306) as a commentator of the Tract I–V of the Summule, Vivarium* VI, 2, 69–101, partial trans. J. Longeway, http://uwp.edu/~longeway/peterspa.htm.

Soto, Domingo de. 1543. In *Dialecticam Aristotelis Commentarii.* Salamanca: J. de Junta.

Soto, Domingo de. [1554] 1980. *Summulae* (1st ed. 1529). Salamanca. Reprint (ed. W. Risse), Hildesheim: Olms.

Sturm, Johann Cristoph. 1661. *Universalia Euclidea.* The Hague.

Suárez, Francisco. [1597] 1965. *Disputationes Metaphysicae*. Reprint of vv. XXV–XXVI of *Opera Omnia* (ed. C. Berton, Paris: Vivès, 1861). Hildesheim: Olms.

Suárez, Francisco. [1597] 1996. *On Being of Reason, Metaphysical Disputation LIV*, ed. and trans. J. P. Doyle. Milwaukee: Marquette University Press.

Thomasius, Christian. [1691a] 1968. *Einleitung zu der Vernunftlehre*. Halle. Reprint (ed. W. Schneiders). Hildesheim: Olms.

Thomasius, Christian. [1691b] 1968. *Ausübung der Vernunftlehre*. Halle. Reprint (ed. W. Schneiders). Hildesheim: Olms.

Thomasius, Christian. 1702. *Introductio ad philosophiam aulicam, seu lineae primae libri de prudentia cogitandi et ratiocinandi, ubi ostenditur media inter praejudicia Cartesianorum et ineptias Peripateticorum, veritatem inveniendi via* (1st ed. 1688). Halle.

Toletus, Franciscus. 1985. *Introductio in universam Aristotelis Logicam* (1st ed., 1560) and *Commentaria in universam Aristotelis Logicam* (1st ed., 1572). In *Opera Omnia Philosophica*. Reprint of the ed. Köln 1615–1616 (ed. W. Risse). Hildesheim: Olms.

Tschirnhaus, Ehrenfried Walter von. [1695] 1964. *Medicina mentis, sive artis inveniendi praecepta generalia* (1st ed. 1687). Reprint of the 2nd ed., Leipzig (with an introduction by W. Risse). Hildesheim: Olms.

Valla, Laurentius. 1982. *Repastinatio Dialecticae et Philosophiae*, ed. G. Zippel. Padova: Antenore.

Walch, Johann Georg. [1726] 1968. *Philosophisches Lexicon*. Leipzig. Reprint of the 4th ed. by J. C. Hennings (Leipzig 1775), Hildesheim: Olms.

Wallis, John. 1687. *Institutio logicae*. Oxford: e Theatro Sheldoniano.

Watts, Isaac. 1725. *Logick, or the right use of reason in the enquiry after truth*. London.

Watts, Isaac. 1741. *The improvement of the mind, or a supplement to the Art of Logick*. London.

Wolff, Christian. [1713] 1978. *Vernünfftige Gedanken von den Kräfften des menschlichen Verstandes*. Halle. In C. Wolff, *Gesammelte Werke*, Sect. I, vol. I, ed. H. W. Arndt. Hildesheim: Olms.

Wolff, Christian. [1740] 1983. *Philosophia rationalis sive Logica*. Reprint of the 2nd ed. Frankfurt–Leipzig (1st ed. 1728). In C. Wolff, *Gesammelte Werke*, Sect. II, vols. I/2–3. ed. J. École. Hildesheim: Olms.

Zabarella, Jacopo. [1578] 1966. *Opera Logica*. Reprint of the ed. Köln 1597 (with an introduction by W. Risse). Hildesheim: Olms.

Zabarella, Jacopo. [1572] 1985. *De methodis libri quatuor. Liber de regressu*. Reprint of the ed. Venice 1578 (introduction by C. Vasoli). Bologna: CLUEB.

References: Secondary Sources

Aner, Karl. 1909. *Gottfried Ploucquets Leben und Lehren*. Halle: Niemeyer.

Angelelli, Ignacio. 1990. On Johannes Raue's Logic. In *Leibniz' Auseinandersetzung mit Vorgängern und Zeitgenossen*, eds. I. Marchlewitz and A. Heinekamp, 184–190. Studia Leibnitiana Supplementa XXVII. Stuttgart: Steiner.

Arndt, Hans Werner. 1965. Christian Wolffs Stellung zu "ars characteristica combinatoria." *Studi e ricerche di storia della filosofia*, 71, 3–12.

Arndt, Hans Werner. 1971. *Methodo scientifica pertractatum. Mos geometricus und Kalkülbegriff in der philosophischen Theorienbildung des 17. und 18. Jahrhunderts*. Berlin–New York: de Gruyter.

Ashworth, Earline Jennifer. 1974. *Language and Logic in the Post-Medieval Period*. Dordrecht–Boston: Reidel.

Ashworth, Earline Jennifer. 1978. Theories of the Proposition: Some Early Sixteenth Century Discussions. In *Franciscan Studies* 38, 81–121. Reprinted with the same pagination in Ashworth 1985.

Ashworth, Earline Jennifer. 1980. The Scholastic Background to Locke's Theory of Language. In *Progress in Linguistic Historiography*, ed. Konrad E. F. Koerner, 59–68. Amsterdam: Benjamins.

Ashworth, Earline Jennifer. 1982. The Structure of Mental Language: Some Problems Discussed by Early Sixteenth Century Logicians. *Vivarium* 20, 1, 59–83. Reprinted with the same pagination in Ashworth 1985a.

Ashworth, Earline Jennifer. 1984. Locke on Language. *Canadian Journal of Philosophy*, 14/1, 45–73. Reprinted with the same pagination in Ashworth 1985a.

Ashworth, Earline Jennifer. 1985a. *Studies in Post-Medieval Semantics*. London: Variorum Reprints.

Ashworth, Earline Jennifer. 1985b. Editor's introduction to Sanderson (1985), ix–lv.

Ashworth, Earline Jennifer. 1990. Domingo de Soto (1494–1560) and the Doctrine of Signs. In *De Ortu Grammaticae. Studies in Medieval Grammar and Linguistic Theory in Memory of Jan Pinborg*, eds. G. L. Bursill-Hall, S. Ebbesen, K. Koerner, 35–48. Amsterdam: Benjamins.

Auroux, Sylvain. 1993. *La logique des idées*. Montréal–Paris: Bellarmin–Vrin.

Barone, Francesco. 1964. *Logica formale e logica trascendentale*, vol. I: *Da Leibniz a Kant*. Torino: Edizioni di "Filosofia" (2nd ed., 1st ed. 1957).

Beck, Lewis White. [1969] 1996. *Early German Philosophy. Kant and His Predecessors*. Cambridge, Mass.: Belknap Press at Harvard University Press. Reprint. Bristol: Thoemmes.

Bernhard, Peter. 2001. *Euler-Diagramme. Zur Morphologie einer Repräsentationsform in der Logik*. Paderborn: Mentis.

Blanché, Robert. [1970] 1996. *La logique et son histoire*, rev. by Jacques Dubucs. Paris: Colin/Masson.

Bos, Egbert P., and Henri A. Krop, eds. 1993. *Franco Burgersdijk (1590–1635)*. Amsterdam: Rodopi.

Boswell, Terry. 1990. The Brothers James and John Bernoulli on the Parallelism between Logic and Algebra. *History and Philosophy of Logic* 11, 173–184.

Brandt, Reinhard. 1991. *Die Urteilstafel. Kritik der reinen Vernunft A 67–76/B 92–101*. Hamburg: Meiner.

Braakhuis, Henk A. 1988. Agricola's View on Universals. In *Rodolphus Agricola Phrisius (1444–1485): Proceedings of the International Conference at the University of Groningen*, ed. F. Akkerman and A. J. Vanderjagt. Leiden: Brill.

Bruyère, Nelly. 1984. *Méthode et dialectique dans l'oeuvre de Pierre de La Ramée*. Paris: Vrin.

Buickerood, James G. 1985. The Natural History of the Understanding: Locke and the Rise of Facultative Logic in the Eighteenth Century. *History and Philosophy of Logic* 6: 157–190.

Camporeale, Salvatore. 1986. Lorenzo Valla, "Repastinatio liber primus": retorica e linguaggio. In *Lorenzo Valla e l'umanesimo italiano*, eds. O. Besomi and M. Regoliosi, 217–239. Padova: Antenore.

Capozzi, Mirella. 1982. Sillogismi e proposizioni singolari: due aspetti della critica di Wolff a Leibniz. In *La grammatica del pensiero: Logica, linguaggio e conoscenza nell'età dell'Illuminismo*, eds. D. Buzzetti and M. Ferriani, 103–150. Bologna: Il Mulino.

Capozzi, Mirella. 1987. Kant on Logic, Language and Thought. In *Speculative Grammar, Universal Grammar and Philosophical Analysis of Language*, eds. D. Buzzetti and M. Ferriani, 97–147. Amsterdam–Philadelphia: Benjamins.

Capozzi, Mirella. 1990. Editor's introduction to Segner [1740], xi–clxxii. Bologna: CLUEB.

Capozzi, Mirella. 1994. Algebra e logica in Jakob Bernoulli. *Atti del Congresso "Logica e filosofia della scienza: problemi e prospettive"* (Lucca 7–10 gennaio 1993), 55–74. Pisa: ETS.

Capozzi, Mirella. 2002. *Kant e la logica*, vol. I. Napoli: Bibliopolis.

Capozzi, Mirella. Forthcoming. *Kant e la logica*, vol. II. Napoli: Bibliopolis.

Carboncini, Sonia. 1991. *Transzendentale Wahrheit und Traum. Christian Wolffs Antwort auf die Herausforderung durch den Cartesianischen Zweifel*. Stuttgart-Bad Cannstatt: Frommann-Holzboog.

Cassirer, Ernst. 1922. *Das Erkenntnisproblem in der Philosophie und Wissenschaft der neuern Zeit*, 3rd ed., vol. II. Berlin: B. Cassirer.

Chomsky, Noam. 1966. *Cartesian Linguistics. A Chapter in the History of Rationalist Thought*. New York–London: Harper and Row.

Couturat, Louis. 1901. *La Logique de Leibnitz d'après des documents inédits*. Paris: Alcan.

Daston, Lorraine. 1988. *Classical Probability in the Enlightenment*. Princeton: Princeton University Press.

De Angelis, Enrico. 1964. *Il metodo geometrico nella filosofia del Seicento*. Pisa: Istituto di Filosofia.

Di Liso, Saverio. 2000. *Domingo de Soto. Dalla logica alla scienza*. Bari: Levante.

Dominicy, Marc. 1984. *La naissance de la grammaire moderne. Language, logique et philosophie à Port-Royal*. Liège: Mardaga.

D'Ors, Angel. 1981. *Las "Summulae" de Domingo de Soto*. Pamplona: Universidad de Navarra (tesi doctoral).

Doyle, John P. 1987–1988. Suarez on Beings of Reason and Truth. *Vivarium* 25, 47–75 and 26, 51–72.

Doyle, John P. 1995. Another God, Chimerae, Goat-Staggs, and Man-Lions: a Seventeenth-Century Debate about Impossible Objects. *Review of Metaphysics* 48, 771–808.

Doyle, John P. 2001. *The Conimbricenses: Some Questions on Signs*. Milwaukee: Marquette University Press.

Dürr, Karl. 1945. Die Logistik Johann Heinrich Lamberts. In *Festschrift zum 60. Geburtstag von Prof. Dr. Andreas Speiser*, 47–65. Zürich: Orell Füßli.

Engfer, Hans-Jürgen. 1982. *Philosophie als Analysis. Studien zur Entwicklung philosophischer Analysiskonzeptionen unter dem Einfluss mathematischer Methodenmodelle im 17. und fruhen 18. Jahrhunderts*. Stuttgart-Bad Cannstatt: Fromman-Holzboog.

Feingold, Mordechai, Joseph S. Freedman, and Wolfgang Rother, eds. 2001. *The Influence of Petrus Ramus*. Basel: Schwabe.

Fitch, Gregory. 1979. Analyticity and Necessity in Leibniz. *Journal of the History of Philosophy*, 17/1, 29–42. Reprint in Woolhouse (1993).

Friedman, Russell L., and Lauge O. Nielsen. 2003. *The Medieval Heritage in Early Modern Metaphysics and Modal Theory, 1400–1700*. Dordrecht: Kluwer.

Gardner, Martin. 1983. *Logic Machines and Diagrams*, 2nd ed. Brighton: Harvester Press.

Garin, Eugenio. 1960. La cultura fiorentina nella seconda metà del '300 e i 'barbari britanni'. In *Rassegna della letteratura italiana*, s. VII, 64, 181–195.

Gaukroger, Stephen. 1989. *Cartesian Logic. An Essay on Descartes's Conception of Inference*. Oxford: Clarendon Press.

Giard, Luce. 1985. La production logique de l'Angleterre au XVIe siècle. *Les Ètudes Philosophiques* 3, 303–324.

Gilbert, Neal W. 1960. *Renaissance Concepts of Method*. New York: Columbia University Press.

Gilson, Étienne. 1913. *Index scolastico-cartésien*. Paris: Felix Alcan.

Green-Pedersen, Niels Jorgen. 1984. *The Tradition of the Topics in the Middle Ages: The Commentaries on Aristotle's and Boethius' Topics*. München–Wien: Philosophia Verlag.

Hacking, Ian. 1975a. *Why Does Language Matter to Philosophy?*. Cambridge: Cambridge University Press.

Hacking, Ian. 1975b. *The Emergence of Probability*. Cambridge: Cambridge University Press.

Hamilton, William. 1860–1869. *Lectures on Metaphysics and Logic*, 2nd ed., vol. II, appendix, eds. H. L. Mansel and J. Veitch. Edinburgh–London: Blackwood.

Harnack, Adolf von. [1900] 1970. *Geschichte der Königliche Preussischen Akademie der Wissenschaften zu Berlin*, 3 vols. Berlin. Reprint Hildesheim: Olms.

Hatfield, Gary. 1997. The Workings of the Intellect: Mind and Psychology. In *Logic and the Workings of the Mind: The Logic of Ideas and Faculty Psychology in Early Modern Philosophy*, ed. Patricia A. Easton, 21–45. Atascadero, Calif.: Ridgeview.

Hickman, Larry. 1980. *Modern Theories of Higher Level Predicates. Second Intentions in the Neuzeit*. Munchen: Philosophia.

Hinske, Norbert. 1986–. *Kant-Index*. Stuttgart-Bad Cannstatt: Frommann-Holzboog.

Howell, Wilbur Samuel. 1971. *Eighteenth Century British Logic and Rhetoric*. Princeton, N.J.: Princeton University Press.

Ishiguro, Hidé. 1991. *Leibniz's Philosophy of Logic and Language*. Cambridge: Cambridge University Press (2nd ed., 1st ed. 1972).

Jardine, Lisa. 1982. *Humanism and the Teaching of Logic*. In N. Kretzmann, A. Kenny, and J. Pinborg (1982), 797–807.

Jardine, Lisa. 1988. *Humanistic Logic*. In C. B. Schmitt, Q. Skinner, and E. Kessler (eds.) 1988, 173–198.

Kneale, William, and Martha Kneale. [1962] 1975. *The Development of Logic*. Oxford: Clarendon Press.

Kretzmann, Norman, Anthony Kenny, and Jan Pinborg, eds. 1982. *The Cambridge History of Later Medieval Philosophy*. Cambridge: Cambridge University Press.

Lenzen, Wolfgang. 1986. "Non est" non est "est non"—Zu Leibnizens Theorie der Negation. *Studia Leibnitiana* 18, 1–37.

Lenzen, Wolfgang. 1990. *Das System der Leibnizschen Logik*. Berlin: de Gruyter.

Lenzen, Wolfgang. 2000. Guilielmi Pacidii *Non plus ultra*—oder: Eine Rekonstruktion des Leibnizschen Plus-Minus-Kalküls. *Philosophiegeschichte und logische Analyse* 3, 71–118.

Lenzen, Wolfgang. 2003. Grundfragen des logischen Kalküls—Eine Art Rezension von F. Schupp (Hrg.), G. W. Leibniz, *Die Grundlagen des logischen Kalküls. History and Philosophy of Logic* 24, 141–162.

Lenzen, Wolfgang. 2004. *Calculus Universalis. Studien zur Logik von G. W. Leibniz.* Paderborn: Mentis.

Mack, Peter. 1993. *Renaissance Argument. Valla and Agricola in the Traditions of Rhetoric and Dialectic.* Leiden–New York–Köln: Brill.

Mahoney, Michael S. 1980. The Beginnings of Algebraic Thought in the Seventeenth Century. In *Descartes. Philosophy, Mathematics and Physics*, ed. Stephen Gaukroger, 141–155. Brighton: Harvester Press.

Maierù, Alfonso. 1993. La dialettica. In *Lo spazio letterario del medioevo latino. 1—La produzione del testo* II, eds. G. Cavallo, C. Leonardi, E. Menestò, 273–294. Rome: Salerno.

Mates, Benson. 1986. *The Philosophy of Leibniz: Metaphysics and Language.* Oxford: Oxford University Press.

Menne, Albert. 1969. Zur Logik von Gottfried Ploucquet. *Akten des XIV Internationalen Kongresses für Philosophie*, 45–49. Wien: Herder.

Mondadori, Fabrizio. 1973. Reference, essentialism, and modality in Leibniz's metaphysics. *Studia Leibnitiana* V, 74–101 (reprint in Woolhouse 1993).

Mugnai, Massimo. 1992. *Leibniz' Theory of Relations. Studia Leibnitiana Supplementa* XXVIII. Stuttgart: Steiner.

Muñoz Delgado, Vicente. 1970. La obra lógica de los españoles en París (1500–1525). *Estudios* 26, 209–280.

Murdoch, John. 1974. Philosophy and the Enterprise of Science in the Later Middle Ages. In *The Interaction between Science and Philosophy*, ed. Y. Elkana, 51–74. Atlantic Highlands, N.J.: Humanities Press.

Nuchelmans, Gabriel. 1980. *Late-Scholastic and Humanist Theories of the Proposition*, Amsterdam–Oxford–New York: North-Holland.

Nuchelmans, Gabriel. 1983. *Judgment and Proposition. From Descartes to Kant.* Amsterdam–Oxford–New York: North-Holland.

Nuchelmans, Gabriel. 1991. *Dilemmatic Arguments: Towards a History of Their Logic and Rhetoric.* Amsterdam: North Holland.

Ong, Walter J. 1958. *Ramus, Method and the Decay of Dialogue: From the Art of Discourse to the Art of Reason.* Cambridge, Mass.: Harvard University Press.

Pariente, Jean-Claude. 1985. *L'analyse du langage à Port-Royal. Six études logico-grammaticales.* Paris: Minuit.

Pérez-Ilzarbe, Paloma. 1999. *El significado de las proposiciones. Jerónimo Pardo y las teorías medievales de la proposición.* Pamplona: Eunsa.

Perreiah, Alan R. 1982. Humanist Critiques of Scholastic Dialectic. *Sixteenth Century Journal* 13, 3–22.

Pozzo, Riccardo. 1989. *Kant und das Problem einer Einleitung in die Logik. Ein Beitrag zur Rekonstruktion der historischen Hintergründe von Kants Logik-Kolleg.* Frankfurt–Bern–New York–Paris: Peter Lang.

Pozzo, Riccardo. 2000. *Georg Friedrich Meiers "Vernunftlehre." Eine historisch-systematische Untersuchung.* Stuttgart-Bad Cannstatt: Frommann-Holzboog.

Reich, Klaus. 1992. *The Completeness of Kant's Table of Judgments*, trans. J. Kneller and M. Losonsky of *Die Vollständigkeit der kantischen Urteilstafel*, Berlin (2nd ed.) 1948 Hamburg 1986 (3rd ed.). Stanford, Calif.: Stanford University Press.

Risse, Wilhelm. 1964–70. *Die Logik der Neuzeit*, 2 vols. Stuttgart-Bad Cannstatt: Frommann-Holzboog.

Risse, Wilhelm. 1965. *Bibliographia Logica: Verzeichnis der Druckschriften zur Logik mit Angabe ihrer Fundorte*, vol. I: 1472–1800. Hildesheim: Olms.

Roncaglia, Gino. 1988. Modality in Leibniz' Essays on Logical Calculus of April 1679. *Studia Leibnitiana* 20/1, 43–62.

Roncaglia, Gino. 1996. *Palaestra rationis. Discussioni su natura della copula e modalità nella filosofia 'scolastica' tedesca del XVII secolo.* Firenze: Olschki.

Roncaglia, Gino. 1998. Sull'evoluzione della logica di Melantone. *Medioevo* 24, 235–265.

Roncaglia, Gino. 2003. Modal Logic in Germany at the Beginning of the Seventeenth Century: Christoph Scheibler's *Opus Logicum*. In R. L. Friedman and L. O. Nielsen 2003, 253–308.

Russell, Bertrand. 1900. *A Critical Exposition of the Philosophy of Leibniz.* Cambridge: Cambridge University Press.

Schepers, Heinrich. 1959. *Andreas Rüdigers Methodologie und ihre Voraussetzungen*, (Kant-Studien, Ergänzungsheft 78). Köln: Kölner Universitäts-Verlag.

Schmitt, Charles B., Quentin Skinner, and Eckhard Kessler, eds. 1988. *The Cambridge History of Renaissance Philosophy.* Cambridge: Cambridge University Press.

Schupp, Franz. 2000. Introduction, trans. and commentary to G. W. Leibniz. *Die Grundlagen des logischen Kalküls.* Hamburg: Meiner.

Shamoon, Alan. 1981. *Kant's Logic*, Ph.D. dissertation, Columbia University (1979). Ann Arbor: University Microfilms International.

Swoyer, Chris. 1995. Leibniz on Intension and Extension. *Nous* 29/1, 96–114.

Thom, Paul. 1981. *The Syllogism.* München: Philosophia Verlag.

Tonelli, Giorgio. 1959. Der Streit über die mathematische Methode in der Philosophie in der ersten Hälfte des 18. Jahrhunderts und die Entstehung von Kants Schrift über die "Deutlichkeit." *Archiv für Begriffsgeschichte* 9, 37–66.

Vailati, Giovanni. 1899. La logique mathématique et sa nouvelle phase de développement dans les écrits de M. J. Peano. *Revue de Métaphysique et de Morale* 7, 86–102.

Vasoli, Cesare. 1968. *La dialettica e la retorica dell'umanesimo.* Milano: Feltrinelli.

Vasoli, Cesare. 1974, Intorno al Petrarca e ai logici "moderni." *Miscellanea Medievalia* IX, *Antiqui und Moderni*, Berlin: De Gruyter, 142–154.

Venn, John. [1894] 1971. *Symbolic Logic.* Reprint of the 2nd ed. New York: Chelsea Publ. Co.

Waswo, Richard. 1999. Theories of Language. In *The Cambridge History of Literary Criticism III—The Renaissance*, ed. G. P. Norton, 25–35. Cambridge: Cambridge University Press.

Wolff, Michael. 1995. *Die Vollständigkeit der kantischen Urteilstafel. Mit einem Essay über Freges Begriffsschrift.* Frankfurt: Klostermann.

Wollgast, Siegfried. 1988a. *Ehrenfried Walther von Tschirnhaus und die deutsche Frühaufklärung.* Berlin: Akademie-Verlag.

Wollgast, Siegfried. 1988b (2nd ed.). *Philosophie in Deutschland zwischen Reformation und Aufklärung 1550–1650.* Berlin: Akademie-Verlag.

Wolters, Gereon. 1980. *Basis und Deduktion. Studien zur Entstehung und Bedeutung der Theorie der axiomatischen Methode bei J. H. Lambert (1728–1777)*. Berlin–New York: Walter de Gruyter.

Woolhouse, Roger S., ed. 1993. *Gottfried Wilhelm Leibniz: Critical Assessments*, vols. I–IV. London–New York: Routledge.

Wundt, Max. [1945] 1964. *Die deutsche Schulphilosophie im Zeitalter der Aufklärung*. Tübingen. Reprint Hildesheim: Olms.

Yolton, John. 1986. Schoolmen, Logic and Philosophy. In *The History of the University of Oxford*, vol. V: The Eighteenth Century, eds. L. S. Sutherland and L. G. Mitchell, 565–591. Oxford: Clarendon Press.

Zalta, Edward. 2000. A (Leibnizian) Theory of Concepts. *Philosophiegeschichte und logische Analyse/Logical Analysis and History of Philosophy* 3, 137–183.

4

The Mathematical Origins of Nineteenth-Century Algebra of Logic

Volker Peckhaus

1. Introduction

Most nineteenth-century scholars would have agreed to the opinion that philosophers are responsible for research on logic. On the other hand, the history of late nineteenth-century logic clearly indicates a very dynamic development instigated not by philosophers but by mathematicians. A central outcome of this development was the emergence of what has been called the "new logic," "mathematical logic," "symbolic logic," or, from 1904 on, "logistics."[1] This new logic came from Great Britain, and was created by mathematicians in the second half of the nineteenth century, finally becoming a mathematical subdiscipline in the early twentieth century.

Charles L. Dodgson, better known under his pen name Lewis Carroll (1832–1898), published two well-known books on logic, *The Game of Logic* of 1887 and *Symbolic Logic* of 1896, of which a fourth edition appeared already in 1897. These books were written "to be of *real* service to the young, and to be taken up, in High Schools and in private families, as a valuable addition of their stock of healthful mental recreations" (Carroll 1896, xiv). They were meant "to *popularize* this fascinating subject," as Carroll wrote in the preface of the fourth edition of *Symbolic Logic* (ibid.). But astonishingly enough, in both books there is no definition of the term "logic." Given the broad scope of these books, the title "Symbolic Logic" of the second book should at least have been explained.

The text is based (but elaborated and enlarged) on my paper "Nineteenth Century Logic between Philosophy and Mathematics" (Peckhaus 1999).

Maybe the idea of symbolic logic was so widely spread at the end of the nineteenth century in Great Britain that Carroll regarded a definition as simply unnecessary. Some further observations support this thesis. They concern a remarkable interest by the general public in symbolic logic, after the death of the creator of the algebra of logic, George Boole, in 1864.

Recalling some standard nineteenth-century definitions of logic as, for example, the art and science of reasoning (Whately) or the doctrine giving the normative rules of correct reasoning (Herbart), it should not be forgotten that mathematical or symbolic logic was not set up from nothing. It arose from the old *philosophical* collective discipline logic. It is therefore obvious to assume that there was some relationship between the philosophical and the mathematical side of the development of logic, but standard presentations of the history of logic ignore this putative relationship; they sometimes even deny that there has been any development of philosophical logic at all, and that philosophical logic could therefore justly be ignored.

Take for instance William and Martha Kneale's program in their eminent *The Development of Logic*. They wrote (1962, iii): "But our primary purpose has been to record the first appearances of these ideas which seem to us most important in the logic of our own day," and these are the ideas leading to mathematical logic. Another example is J. M. Bocheński's assessment of "modern classical logic," which he dated between the sixteenth and the nineteenth century. This period was for him noncreative. It can therefore justly be ignored in a problem history of logic (1956, 14). According to Bocheński, classical logic was only a decadent form of this science, a dead period in its development (ibid., 20).

Authors advocating such opinions adhere to the predominant views of present-day logic, that is, actual systems of mathematical or symbolic logic. As a consequence, they are not able to give reasons for the final divorce between philosophical and mathematical logic, because they ignore the seed from which mathematical logic has emerged. Following Bocheński's view, Carl B. Boyer presented for instance the following periodization of the development of logic (Boyer 1968, 633): "The history of logic may be divided, with some slight degree of oversimplification, into three stages: (1) Greek logic, (2) scholastic logic, and (3) mathematical logic." Note Boyer's "slight degree of oversimplification" which enabled him to skip 400 years of logical development and ignore the fact that Kant's transcendental logic, Hegel's metaphysics, and Mill's inductive logic were called "logic," as well.

This restriction of scope had a further consequence: The history of logic is written as if it had been the nineteenth-century mathematicians' main motive for doing logic to create and develop a new scientific discipline as such, namely mathematical logic, dealing above all with problems arising in this discipline and solving these problems with the final aim of attaining a coherent theory. But what, if logic was only a means to an end, a tool for solving nonlogical problems? If this is considered, such nonlogical problems have to be taken note of. One can assume that at least the initial motives of mathematicians working in logic were going beyond creating a new or further developing the traditional theory of logic. Under the presupposition that a mathematician is usually not really interested in devoting his professional work to the development of a philosophical subdiscipline, one can assume that theses motives have to be sought in the mathematician's own subject, namely in foundational, that is, philosophical problems of mathematics.

Today historians have recognized that the emergence of the new logic was no isolated process. Its creation and development ran parallel to and was closely intertwined with the creation and development of modern abstract mathematics which emancipated itself from the traditional definition as a science which deals with quantities and geometrical forms and is therefore responsible for *imaginabilia*, that is, intuitive objects. The *imaginabilia* are distinguished from *intelligibilia*, that is, logical objects which have their origin in reason alone. These historians recognized that the history of the development of modern logic can only be told within the history of the development of mathematics because the new logic is not conceivable without the new mathematics. In recent research on the history of logic, this intimate relation between logic and mathematics, especially its connection to foundational studies in mathematics, has been taken into consideration. One may mention the present author's *Logik, Mathesis universalis und allgemeine Wissenschaft* (Peckhaus 1997) dealing with the philosophical and mathematical contexts of the development of nineteenth-century algebra of logic as at least partially unconscious realizations of the Leibnizian program of a universal mathematics, José Ferreirós's history of set theory in which the deep relations between the history of abstract mathematics and that of modern logic (Ferreirós 1999) are unfolded, and the masterpiece of this new direction, *The Search for Mathematical Roots, 1870–1940* (2000a) by Ivor Grattan-Guinness, who imbedded the whole bunch of different directions in logic into the development of foundational interests within mathematics. William Ewald's "Source Book in the Foundations of Mathematics" (Ewald 1996) considers logical influences at least in passing, whereas the *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, edited by Ivor Grattan-Guinness (1994), devotes an entire part to "Logic, Set Theories and the Foundation of Mathematics" (vol. 1, pt. 5).

In the following, the complex conditions for the emergence of nineteenth-century symbolic logic will be discussed. The main scope will be on the mathematical motives leading to the interest in logic; the philosophical context will be dealt with only in passing. The main object of study will be the algebra of logic in its British and German versions. Special emphasis will be laid on the systems of George Boole (1815–1864) and above all of his German follower Ernst Schröder (1841–1902).

2. Boole's Algebra of Logic

2.1. Philosophical Context

The development of the new logic started in 1847, completely independent of earlier anticipations, for example, those by the German rationalistic universal genius Gottfried Wilhelm Leibniz (1646–1716) and his followers (see Peckhaus 1994a, 1997, ch. 5). In that year the British mathematician Boole published his pamphlet *The Mathematical Analysis of Logic* (1847).[2] Boole mentioned (1847, 1) that it was the struggle for priority concerning the quantification of the predicate between the Edinburgh philosopher William Hamilton (1788–1856) and the London mathematician Augustus De Morgan (1806–1871) that encouraged this study. Hence, he referred to a startling philosophical discussion which indicated a vivid interest in formal logic in Great Britain. This interest was, however, a new interest, just 20 years old. One can even say that neglect of formal logic could be regarded as a characteristic feature of British philosophy up to 1826 when Richard Whately (1787–1863) published his *Elements of Logic*.[3] In his preface Whately added an extensive report on the languishing research and education in formal logic in England. He complained (1826, xv) that only very few students of the University of Oxford became good logicians and that

> by far the greater part pass through the University without knowing any thing of all of it; I do not mean that they have not learned by rote a string of technical terms; but that they understand absolutely nothing whatever of the principles of the Science.

Thomas Lindsay, the translator of Friedrich Ueberweg's important *System der Logik und Geschichte der logischen Lehren* (1857, English translation), was very critical of the scientific qualities of Whately's book, but he nevertheless emphasized its outstanding contribution for the renaissance of formal logic in Great Britain (Lindsay 1871, 557):

> Before the appearance of this work, the study of the science had fallen into universal neglect. It was scarcely taught in the universities, and there was hardly a text-book of any value whatever to be put into the hands of the students.

One year after the publication of Whately's book, George Bentham's *An Outline of a New System of Logic* appeared (1827) which was intended as a commentary to Whately. Bentham's book was critically discussed by William Hamilton in a review article published in the *Edinburgh Review* (Hamilton 1833). With the help of this review, Hamilton founded his reputation as the "first logical name in Britain, it may be in the world."[4] Hamilton propagated a revival of the Aristotelian scholastic formal logic without, however, one-sidedly preferring the syllogism. His logical conception was focused on a revision of the standard forms by quantifying the predicates of judgments.[5] He arrived at eight standard forms (Hamilton 1859–1866, vol. 4, 1866, 287):

1. A "All A is all B" toto-total.
2. A "All A is some B" toto-partial.
3. I "Some A is all B" parti-total.
4. I "Some A is some B" parti-partial.
5. E "Any A is not any B" toto-total.
6. E "Any A is not some B" toto-partial.
7. O "Some A is not any B" parti-total.
8. O "Some A is not some B" parti-partial.

Hamilton's unconsidered transition from the collective "all" to the distributive "any" has already been criticized by William and Martha Kneale (1962, 353). Hamilton used a geometrical symbolism using wedges for illustrating the effects of this modification.[6]

The controversy about priority arose when De Morgan, in a lecture "On the Structure of the Syllogism" (De Morgan 1846) given to the Cambridge Philosophical Society on 9 November 1846, also proposed the quantification of the predicates.[7] Neither had any priority, of course. The application of diagrammatic methods in syllogistic reasoning proposed, for example, by the eighteenth-century mathematicians and philosophers Leonard Euler, Gottfried Ploucquet, and Johann Heinrich Lambert, presupposed a quantification of the predicate.[8] The German psychologistic logician Friedrich Eduard Beneke (1798–1854) suggested to quantify the predicate in his books on logic published in 1839 and 1842, the latter of which he sent to Hamilton (see Peckhaus 1997, 191–193). In the context of this presentation, it is irrelevant to give a final solution of the priority question. It is, however, important that a dispute of this extent arose at all. It indicates that there was a new interest in *research* on formal logic.

This interest represented only one side of the effects released by Whately's book. Another line of research stood in the direct tradition of Humean empiricism and the philosophy of inductive sciences: the inductive logic of John Stuart Mill (1806–1873), Alexander Bain (1818–1903), and others. Boole's logic was in clear opposition to inductive logic. It was Boole's follower William Stanley Jevons (1835–1882; see Jevons 1877–1878) who made this opposition explicit.

As mentioned earlier, Boole referred to the controversy between Hamilton and De Morgan, but this influence should not be overemphasized. In his main work on the *Laws of Thought* (1854), Boole went back to the logic of Aristotle by quoting from the Greek original. This can be interpreted as indicating that the influence of the contemporary philosophical discussion was not as important as his own words might suggest. In writing a book on logic he was doing philosophy, and it was thus a matter of course that he related his results to the philosophical discussion of his time. This does not mean, of course, that his thoughts were mainly influenced by this discussion. In any case, Boole's early algebra of logic kept a close connection to traditional logic, in the formal part of which the theory of syllogism represented its core.[9] Traditional logic not only provided the topics to be dealt with by the "Calculus of Deductive Reasoning,"[10] it also served as a yardstick for evaluating the power and the reliability of the new logic. Even in the unpublished manuscripts of a sequel of the *Laws of Thought* titled "The Philosophy of Logic," he discussed Aristotelian logic at length (see Boole 1997, 133–136), criticizing, however, that it is more a mnemonic art than a science of reasoning.[11]

2.2. The Mathematical Context in Great Britain

Of greater importance than the philosophical discussion on logic in Great Britain were mathematical influences. Most of the new logicians can be related to the so-called Cambridge Network (Cannon 1978, 29–71), that is, a movement that aimed at reforming British science and mathematics which started at Cambridge. One of the roots of this movement was the foundation of the Analytical Society in 1812 (see Enros 1983) by Charles Babbage (1791–1871), George Peacock (1791–1858), and John Herschel (1792–1871). Joan L. Richards called this act a "convenient starting date for the nineteenth-century chapter of British mathematical development" (Richards 1988, 13). One of the first achievements of the Analytical Society was a revision of the Cambridge Tripos by adopting the Leibnizian notation for the calculus and abandoning the customary Newtonian theory of fluxions: "the principles of pure D-ism in opposition to the Dot-age of the University" as Babbage wrote in his memoirs (Babbage 1864, 29). It may be assumed that this successful movement triggered off by a change in notation might have stimulated a new or at least revived interest in operating with symbols. This new research on the calculus had parallels in innovative approaches to algebra which were motivated by the reception of Laplacian analysis.[12] In the first place, the development of symbolic algebra has to be mentioned. It was codified by George Peacock in his *Treatise on Algebra* (1830) and further propagated in his famous report for the British Association for the Advancement of Science (Peacock 1834, especially 198–207). Peacock started by drawing a distinction between arithmetical and symbolic algebra, which was, however, still based on the common restrictive understanding of arithmetic as the doctrine of quantity. A generalization of Peacock's concept can be seen in Duncan F. Gregory's (1813–1844) "calculus of operations."[13] Gregory was most interested in *operations* with symbols. He defined symbolic algebra as "the science which treats of the combination of operations defined not by their nature, that is by what they are or what they do, but by the laws of combinations to which they are subject" (1840, 208). In his much praised paper "On a General Method in Analysis" (1844), Boole made the calculus of operations the basic methodological tool for analysis.

However, in following Gregory, he went further, proposing more applications. He cited Gregory, who wrote that a symbol is defined algebraically “when its laws of combination are given; and that a symbol represents a given operation when the laws of combination of the latter are the same as those of the former” (Gregory 1842, 153–154). It is possible that a symbol for an arbitrary operation can be applied to the same operation (ibid., 154). It is thus necessary to distinguish between arithmetical algebra and symbolic algebra, which has to take into account symbolic but nonarithmetical fields of application. As an example, Gregory mentioned the symbols a and $+a$. They are isomorphic in arithmetic, but in geometry they need to be interpreted differently. a can refer to a point marked by a line, whereas the combination of the signs $+$ and a additionally expresses the direction of the line. Therefore symbolic algebra has to distinguish between the symbols a and $+a$. Gregory deplored the fact that the unequivocity of notation did not prevail as a result of the persistence of mathematical practice. Clear notation was only advantageous, and Gregory thought that our minds would be “more free from prejudice, if we never used in the general science symbols to which definite meanings had been appropriated in the particular science” (ibid., 158).

Boole adopted this criticism almost word for word. In his *Mathematical Analysis of Logic* he claimed that the reception of symbolic algebra and its principles was delayed by the fact that in most interpretations of mathematical symbols the idea of quantity was involved. He felt that these connotations of quantitative relationships were the result of the context of the emergence of mathematical symbolism, and not of a universal principle of mathematics (Boole 1847, 3–4). Boole read the principle of the permanence of equivalent forms as a principle of independence from interpretation in an “algebra of symbols.” To obtain further affirmation, he tried to free the principle from the idea of quantity by applying the algebra of symbols to another field, the field of logic. As far as logic is concerned this implied that only the principles of a “true Calculus” should be presupposed. This calculus is characterized as a “method resting upon the employment of Symbols, whose laws of combination are known and general, and whose results admit of a consistent interpretation” (ibid., 4). He stressed (ibid.):

> It is upon the foundation of this general principle, that I purpose to establish the Calculus of Logic, and that I claim for it a place among the acknowledged forms of Mathematical Analysis, regardless that in its objects and in its instruments it must at present stand alone.

Boole expressed logical propositions in symbols whose laws of combination are based on the mental acts represented by them. Thus he attempted to establish a psychological foundation of logic, mediated, however, by language.[14] The central mental act in Boole’s early logic is the act of election used for building classes. Man is able to separate objects from an arbitrary collection which belong to given classes to distinguish them from others. The symbolic representation of these mental operations follows certain laws of combination that are similar to those of symbolic algebra. Logical theorems can thus be proven like mathematical theorems. Boole's opinion has of course consequences for the place of logic in philosophy: "On the principle of a true classification, we ought no longer to associate Logic and Metaphysics, but Logic and Mathematics" (ibid., 13).

Although Boole's logical considerations became increasingly philosophical with time, aiming at the psychological and epistemological foundations of logic itself, his initial interest was not to reform logic but to reform mathematics. He wanted to establish an abstract view on mathematical operations without regard to the objects of these operations. When claiming "a place among the acknowledged forms of Mathematical Analysis" (1847, 4) for the calculus of logic, he didn't simply want to include logic in traditional mathematics. The superordinate discipline was a *new* mathematics. This is expressed in Boole's writing: "It is not of the essence of mathematics to be conversant with the ideas of number and quantity" (1854, 12).

2.3. Boole's Logical System

Boole's early logical system is based on mental operations, namely, acts of selecting individuals from classes. In his notation 1 symbolizes the Universe, comprehending "every conceivable class of objects whether existing or not" (1847, 15). Capital letters stand for all members of a certain class. The small letters are introduced as follows (ibid., 15):

> The symbol x operating upon any subject comprehending individuals or classes, shall be supposed to select from that subject all the Xs which it contains. In like manner the symbol y, operating upon any subject, shall be supposed to select from it all individuals of the class Y which are comprised in it and so on.

Take A as the class of animals, then x might signify the selection of all sheep from these animals, which then can be regarded as a new class X from which we select further objects, and so on. This might be illustrated by the following example:

animals		sheep		horned sheep		black horned sheep
A		X		Y		Z
$\downarrow$	$\nearrow$	$\downarrow$	$\nearrow$	$\downarrow$	$\nearrow$	
x		y		z		
sheep		horned		black		

This process represents a successive selection which leads to individuals being common to the classes A, X, Y, and Z. xyz stands for animals that are sheep, horned, and black. It can be regarded as the logical product of some common marks or common aspects relevant for the selection. In his major work, *An Investigation of the Laws of Thought* of 1854, Boole gave up this distinction between capital and small letters, thereby getting rid of the complicated consequences of this stipulation.

If the symbol 1 denotes the universe, and if the class X is determined by the symbol x, it is consequent that the class not-X has to be denoted by the symbol $1-x$, which forms the supplement to x, thus $x(1-x)=0$. 0 symbolizes nothing or the empty class. Now one can consider Boole's interpretation of the universal-affirmative judgment. The universal-affirmative judgment "All Xs are Ys" is expressed by the equation $xy=x$ or, by simple arithmetical transformation, $x(1-y)=0$ (p. 22): "As all the Xs which exist are found in the class Y, it is obvious that to select out of the Universe all Ys, and from these to select all Xs, is the same as to select at once from the Universe all Xs." The universal-negative judgment "No Xs are Ys" asserts that there are no terms common in the classes X and Y. All individuals common would be represented by xy, but they form the empty class. The particular-affirmative judgment "Some Xs are Ys" says that there are some terms common to both classes forming the class V. They are expressed by the elective symbol v. The judgment is thus represented by $v=xy$. Boole furthermore considers using $vx=vy$ with vx for "some X" and vy for "some Y," but observes "that this system does not express quite so much as the single equation" (pp. 22–23). The particular-negative judgment "Some Xs are not Ys" can be reached by simply replacing y in the last formula with $1-y$.

Boole's elective symbols are compatible with the traditional theory of judgment. They blocked, however, the step toward modern quantification theory as present in the work of Gottlob Frege, but also in later systems of the algebra of logic like those of C. S. Peirce and Ernst Schröder.[15]

The basic relation in the Boolean calculus is equality. It is governed by three principles which are themselves derived from elective operations (see ibid., 16–18):

1. The *Distributivity of Elections* (16–17):

   > it is indifferent whether from of group of objects considered as a whole, we select the class X, or whether we divide the group into two parts, select the Xs from them separately, and then connect the results in one aggregate conception, in symbols:
   >
   > $$x(u+v)=xu+xv,$$
   >
   > with $u+v$ representing the undivided group of objects, and u and v standing for its component parts.

2. The *Commutativity of Elections*: The order of elections is irrelevant:

   $$xy=yx.$$

3. The *Index Law*: The successive execution of the same elective act does not change the result of the election:

   $$x^n=x, \quad \text{for } n\geq 2.$$

Boole stressed the importance of the Index Law, which is not generally valid in arithmetic (only in the arithmetic of 0 and 1) and therefore peculiar for elective symbols. It allows one to reduce complex formulas to forms more easily capable of being interpreted.

In his *Investigation of the Laws of Thought* (1854) Boole abandoned the Index Law and replaced it by the Law of Duality ("Boole's Law") $xx = x$, or $x^2 = x$.[16] His esteem for this law becomes evident in his claim "that the axiom of the metaphysicians which is termed the principle of contradiction..., is a consequence of the fundamental law of thought whose expression is $x^2 = x$" (Boole 1854, 49). Boole referred to the derivation

$$\begin{aligned} x^2 &= x \\ x - x^2 &= 0 \\ x(1-x) &= 0, \end{aligned}$$

the last formula saying that a class and its complement have no elements in common. Boole was heavily criticized for this "curious error" (Halsted 1878, 86) of considering the Law of Contradiction a consequence of the Law of Duality, not the other way around (the derivation works, of course, also in the other direction).

Boole's revisions came along with a change in his attitude toward logic. His early logic can be seen as an application of a new mathematical method to logic, thereby showing the efficacy of this method within the broad project of a universal mathematics and so serving foundational goals in mathematics. This foundational aspect diminished in later work, successively being replaced by the idea of a reform of logic. Already in the paper "The Calculus of Logic" (Boole 1848), Boole tried to show that his logical calculus is compatible with traditional philosophical logic. Reasoning is guided by the laws of thought. They are the central topic in Boole's *Investigation of the Laws of Thought*, claiming that "there is to a considerable extent an exact agreement in the laws by which the two classes of operations are conducted" (1854, 6), comparing thereby the laws of thought and the laws of algebra. Logic, in Boole's understanding, was "a normative theory of the products of mental processes" (Grattan-Guinness 2000a, 51).

2.4. Symbolic Logic within the Old Paradigm: De Morgan

Although created by mathematicians, the new logic was widely ignored by fellow mathematicians. Boole was respected by British mathematicians, but his ideas concerning an algebraic representation of the laws of thought received very little published reaction.[17] He shared this fate with De Morgan, the second major figure of symbolic logic at that time.[18]

Like Boole, the British mathematician De Morgan was influenced by algebraist George Peacock's work on symbolic algebra, which motivated him to consider the foundations of algebra in connection with logic. He distinguished, for example, algebra as an art associated with what he called "technical algebra" and algebra as science, that is, "logical algebra": "Technical algebra is the art of using symbols under regulations which ... are prescribed as the definition of symbols. Logical algebra is the science which investigates the method of giving meaning to the primary symbols, and of interpreting all subsequent results" (De Morgan 1842, 173–174, reprint p. 338).

He used algebraic symbolism in logic, being mainly interested in a reform and extension of syllogistic logic, but ignoring the operational aspect of logic as calculus. He published his main results in a series of papers in the *Proceedings of the Cambridge Philosophical Society* between 1846 and 1862 (reprinted in De Morgan 1966) and in his book *Formal Logic* (1847).

He has been called "the last great traditional logician" (Hailperin 2004, 346). Among his lasting achievements is the introduction of the technical term of a universe. He spoke, for example, of the "*Universe* of a proposition, or of a name" that may be limited in any matter expressed or understood" (De Morgan 1846/1966, 2) but continued to distinguish two kinds of the universe of a population, "being either the whole universe of thought, or a given portion of it" (De Morgan 1853/1966, 69).

In the first of the papers "On the Syllogism," he introduced an algebraic symbolism for the syllogism, using small letters x, y, z as names contrary to those represented by capitals X, Y, Z (De Morgan 1846/1966, 3). The relations between such names as expressed in standard forms or simple propositions are symbolized as follows (ibid., 4):

$P)Q$	signifies	Every P is Q.
$P.Q$	...	No P is Q.
PQ	...	Some Ps are Qs.
$P{:}Q$	...	Some Ps are not Qs.

The algebraic symbols thus signify both the quantity of the concepts involved and the copula. For the names X and Y and their contraries x and y, the following equations are valid (ibid.):

$$
\begin{array}{ll}
X)Y = X.y = y)x & X.Y = X)y = Y)x \\
X{:}Y = Xy = y{:}x & XY = X{:}y = Y{:}x \\
Y)X = Y.x = x)y & x.y = x)Y = y)X \\
Y{:}X = Yx = x{:}y & xy = x{:}Y = y{:}X
\end{array}
$$

De Morgan used this symbolism to reconstruct the theory of syllogism. It served as representation, not as a calculus.

Only after having written the 1846 paper, De Morgan found "that the whole theory of the syllogism might be deduced from the consideration of propositions in a form in which *definite quantity* of assertion is given both to the subject and the predicate of a proposition," as he reported in an "Addition," dated 27 February 1847 (De Morgan 1966, 17). He claimed to have brought this idea to paper before he learned of Sir William Hamilton's quantification of the predicate, thereby opening the priority quarrel.

De Morgan focused his subsequent logical work on the theory of the copula, following "the hint given by algebra" by separating "the essential from the accidental characteristics of the copula" (1850/1966, 50). The "abstract copula" characterized only by essential features is understood as "a formal mode of joining two terms which carries no meaning, and obeys no law except such as is barely necessary to make the forms of inference follow" (ibid., 51). The abstract copula follows two "copular conditions,"

(1) *transitiveness*

$$X \text{ --- } Y \text{ --- } Z = X \text{ --- } Z$$

(2) *convertibility*

$$X \text{ --- } Y = Y \text{ --- } X$$

Affirmative (—) and negative (−−) copula are contrary to each other. Of X — Y and X −− Y one or the other must be (De Morgan 1850/1966, 51).

De Morgan was the first to take seriously that traditional syllogistics was incapable of dealing with relational properties like "Smith is smaller than Jones." His ideas concerning a logic of (two-place) relations can be regarded as his most important contributions (see Merrill 1990, chs. 5–6; Grattan-Guinness 2000a, 32–34). Already in his second paper on the syllogism, he mentioned the role of the copula for expressing the relation between what is connected. He also considered the composition of relations (1850/1966, 55), that is, in modern terms, the relative product. He studied the subject of relations "as a branch of logic" in his fourth paper on the syllogism (De Morgan 1860/1966, 208). De Morgan used capital letters L, M, N for denoting relations, lowercase letters l, m, n for the respective contraries. Additionally, two periods indicate that a relation holds, only one period that the contrary relations holds. Thus, $X..LY$ or $X.lY$ say that X is "some one of the objects of thought which stand to Y in the relation L, or is one of the Ls of Y" (ibid., 220). X and Y are called "subject" and "predicate," indicating the mode in which they stand in the relation, thus in both $LY.X$ and $X.LY$, Y indicates the predicate. If the predicate is itself the subject of a relation, a composition of relations results. "Thus if $X..L(MY)$, if X be one of the Ls of one of the Ms of Y, we may think of X as an 'L of M' of Y, expressed by $X..(LM)Y$, or simply by $X..LM\ Y$" (ibid., 221). De Morgan used an accent to signify universal quantity as part of the description of the relation. LM' stands for an L of every M, $LM'X$ for the same relation to many (ibid.). The converse relation of L, L', is defined as if $X..LY$, then $Y..L^{-1}X$" (ibid., 222). De Morgan then applied this symbolism to his theory of syllogism, introducing "theorem K" as basic for what he called "opponent syllogism," which is exemplified by the following mathematical syllogism (ibid., 224–225):

> Every deficient of an external is a coinadequate: *external* and *coinadequate* have *partient* and *complement* for their contraries,

> and *deficient* has *exient* for its converse: hence every exient of a complement is a patient; which is one of the opponent syllogisms of that first given.

Theorem K says (ibid., 224) that

> if a compound relation be contained in another relation, by the nature of the relations and not by casualty of the predicate, the same may be said when either component is converted, and the contrary of the other component and of the component change places.

One of the examples is that "if, be Z what it may, every L of M of Z be an N of Z, say $LM))N$, then $L^{-1}n))m$, and $nM^{-1}))l$" (ibid.).

The problematic nature of De Morgan's symbolism becomes obvious in his notation for complex terms. The conjunctive "P and Q" is expressed by PQ, the disjunctive (taken in the inclusive sense) by P, Q. Using this notation he formulated the laws named after him (that can, however, be found already in the work of William of Ockham): "The contrary of PQ is p, q; that of P, Q, is pq" (*1847*, 118). The equivalent in modern notation is $\neg(p \vee q) = \neg p \wedge \neg q$, and $\neg(p \wedge q) = \neg p \vee \neg q$, or in the quantificational version $\neg\exists_x ax = \forall_x \neg ax$ and $\neg\forall_x ax = \exists_x \neg ax$.

2.5. Reception of the New Logic

In 1864, Samuel Neil, the early chronicler of British mid-nineteenth-century logic, expressed his thoughts about the reasons for this negligible reception: "De Morgan is esteemed crotchety, and perhaps formalizes too much. Boole demands high mathematic culture to follow and to profit from" (1864, 161). One should add that the ones who had this culture were usually not interested in logic.

The situation changed after Boole's death in 1864. In the following comments only some ideas concerning the reasons for this new interest are hinted at. In particular the roles of William Stanley Jevons and Alexander Bain are considered. These examples show that a broader reception of symbolic logic commenced only when its relevance for the philosophical discussion of the time came to the fore.

2.5.1. William Stanley Jevons

A broader international reception of Boole's logic began when Jevons (1835–1882) made it the starting point for his influential *Principles of Science* (Jevons 1874). He used his own version of the Boolean calculus introduced in his *Pure Logic* (Jevons 1864). Among his revisions were the introduction of a simple symbolic representation of negation and the definition of logical addition as inclusive "or," thereby creating Boolean algebra (see Hailperin 1981). He also changed the philosophy of symbolism (1864, 5):

> The forms of my system may, in fact, be reached by divesting his [Boole's] of a mathematical dress, which, to say the least, is not essential to it. The system being restored to its proper simplicity, it may be inferred, not that Logic is a part of Mathematics, as is almost implied in Professor Boole's writings, but that the Mathematics are rather derivatives of Logic. All the interesting analogies or samenesses of logical and mathematical reasoning which may be pointed out, are surely reversed by making Logic dependent on Mathematics.

Jevons's interesting considerations on the relationship between mathematics and logic representing an early logicistic attitude will not be discussed here. Similar ideas can be found not only in Gottlob Frege's work, but also in that of Rudolf Hermann Lotze (1817–1881) and Schröder. Most important in the present context is the fact that Jevons abandoned mathematical symbolism in logic, an attitude that was later taken up by John Venn (1834–1923) in his *Symbolic Logic* (Venn 1894). Jevons attempted to free logic from the semblance of being a special mathematical discipline. He used the symbolic notation only as a means of expressing general truths. Logic became a tool for studying science, a new language providing symbols and structures. The change in notation brought the new logic closer to the philosophical discourse of the time. The reconciliation was supported by the fact that Jevons formulated his *Principles of Science* as a rejoinder to John Stuart Mill's (1806–1873) *System of Logic* of 1843, at that time the dominating work on logic and the philosophy of science in Great Britain. Although Mill had called his logic *A System of Logic Ratiocinative and Inductive*, the deductive parts played only a minor role, used only to show that all inferences, all proofs, and the discovery of truths consisted of inductions and their interpretations. Mill claimed to have shown "that all our knowledge, not intuitive, comes to us exclusively from that source" (Mill 1843, bk. II, ch. I, §1). Mill concluded that the question as to what induction is, is the most important question of the science of logic, "the question which includes all others." As a result the logic of induction covers by far the largest part of this work, a subject that we would today regard as belonging to the philosophy of science.

Jevons defined induction as a simple inverse application of deduction. He began a direct argument with Mill in a series of papers titled "Mill's Philosophy Tested" (1877/78). This argument proved that symbolic logic could be of importance not only for mathematics, but also for philosophy.

Another effect of the attention caused by Jevons was that British algebra of logic was able to cross the Channel. In 1877, Louis Liard (1846–1917), at that time professor at the Faculté de lettres at Bordeaux and a friend of Jevons, published two papers on the logical systems of Jevons and Boole (Liard 1877a, 1877b). In 1878 he added a booklet titled *Les logiciens anglais contemporaines* (Liard 1878), which had five editions until 1907 and was translated into German (Liard 1880). Although Hermann Ulrici (1806–1884) had published a first German review of Boole's *Laws of Thought* as early as 1855 (Ulrici 1855, see Peckhaus 1995), the knowledge of British symbolic logic was conveyed primarily by Alois Riehl (1844–1924), then professor at the University of Graz in Austria. He published a widely read paper, "Die englische Logik der Gegenwart" ("English contemporary logic," Riehl 1877), which reported mainly Jevons's logic and utilized it in a current German controversy on the possibility of scientific philosophy.

2.5.2. Alexander Bain

Surprisingly good support for the reception of Boole's algebra of logic came from the philosophical opposition, namely from the Scottish philosopher Bain (1818–1903) who was an adherent of Mill's logical theory. Bain's *Logic*, first published in 1870, had two parts, the first on deduction and the second on induction. He made explicit that "Mr Mill's view of the relation of Deduction and Induction is fully adopted" (1870, I, iii). Obviously he shared the "general conviction that the utility of the purely Formal Logic is but small; and that the rules of Induction should be exemplified even in the most limited course of logical discipline" (ibid., v). The minor role of deduction showed up in Bain's definition "*Deduction* is the application or extension of Induction to *new cases*" (40).

Despite his reservations about deduction, Bain's *Logic* became important for the reception of symbolic logic because of a chapter of 30 pages titled "Recent Additions to the Syllogism." In this chapter the contributions of Hamilton, De Morgan, and Boole were introduced. One can assume that many more people became acquainted with Boole's algebra of logic through Bain's report than through Boole's own writings. One example is Hugh MacColl (1837–1909), the pioneer of the calculus of propositions (statements) and of modal logic.[19] He created his ideas independently of Boole, eventually realizing the existence of the Boolean calculus by means of Bain's report. Even in the early parts of his series of papers "The Calculus of Equivalent Statements," he quoted from Bain's presentation when discussing Boole's logic (MacColl 1877/78). In 1875 Bain's logic was translated into French, in 1878 into Polish. Tadeusz Batóg and Roman Murawski (1996) have shown that it was Bain's presentation which motivated the first Polish algebraist of logic, Stanisław Piątkiewicz (1848–?) to begin his research on symbolic logic.

3. Schröder's Algebra of Logic

3.1. Philosophical Background

The philosophical discussion on logic after Hegel's death in Germany was still determined by a Kantian influence.[20] In the preface to the second edition of his *Kritik der reinen Vernunft* of 1787, Immanuel Kant (1723–1804) had written that logic had followed the safe course of a science since earliest times.

For Kant, this was evident because of the fact that logic had been prohibited from taking any step backward from the time of Aristotle. But he regarded it as curious that logic hadn't taken a step forward either (B VIII). Thus, logic seemed to be closed and complete. Formal logic, in Kant's terminology the analytical part of general logic, did not play a prominent role in Kant's system of transcendental philosophy. In any case, it was a negative touchstone of truth, as he stressed (B 84). Georg Wilhelm Friedrich Hegel (1770–1831) went further in denying any relevance of formal logic for philosophy (Hegel 1812/13, I, Introduction, XV–XVII). Referring to Kant, he maintained that from the fact that logic hadn't changed since Aristotle one should infer that it needs to be completely rebuilt (ibid., XV). Hegel created a variant of logic as the foundational science of his philosophical system, defining it as "the science of the pure idea, i.e., the idea in the *abstract element of reasoning*" (1830, 27). Hegelian logic thus coincides with metaphysics (ibid., 34).

This was the situation when after Hegel's death philosophical discussion on formal logic started again in Germany. This discussion on logic reform stood under the label of "the logical question," a term created by the neo-Aristotelian Adolf Trendelenburg (1802–1872). In 1842 he published a paper titled "Zur Geschichte von Hegel's Logik und dialektischer Methode" with the subtitle "Die logische Frage in Hegel's Systeme." But what is the logical question according to Trendelenburg? He formulated this question explicitly toward the end of his article: "Is Hegel's dialectical method of pure reasoning a scientific procedure?" (1842, 414). In answering this question in the negative, he provided the occasion of rethinking the status of formal logic within a theory of human knowledge without, however, proposing a return to the old (scholastic) formal logic. The term "the logical question" was subsequently used in a less specific way. Georg Leonard Rabus, the early chronicler of the discussion on logic reform, wrote, for example, that the logical question emerged from doubts concerning the justification of formal logic (1880, 1).

Although this discussion was clearly *connected* to formal logic, the so-called reform did not *concern* formal logic. The reason was provided by the neo-Kantian Wilhelm Windelband who wrote in a brilliant survey on nineteenth-century (philosophical) logic (1904, 164):

> It is in the nature of things that in this enterprize [i.e., the reform of logic] the lower degree of fruitfulness and developability power was on the side of formal logic. Reflection on the rules of the correct progress of thinking, the technique of correct thinking, had indeed been brought to perfection by former philosophy, presupposing a naive world view. What Aristotle had created in a stroke of genius, was decorated with the finest filigree work in Antiquity and the Middle Ages: an art of proving and disproving which culminated in a theory of reasoning, and after this constructing the doctrines of judgements and concepts. Once one has accepted the foundations, the safely assembled building cannot be shaken:

> it can only be refined here and there and perhaps adapted to new scientific requirements.

Windelband was very critical of English mathematical logic. Its quantification of the predicate allows the correct presentation of extensions in judgments, but it "drops hopelessly" the vivid sense of all judgments, which tend to claim or deny a material relationship between subject or predicate. It is "a logic of the conference table," which cannot be used in the vivid life of science, a "logical sport" which has its merits only in exercising the final acumen (ibid., 166–167).

The philosophical reform efforts concerned primarily two areas:

1. the problem of a foundation of logic itself. It was dealt with by using psychological and physiological means, thereby leading to new discussion on the question of priority between logic and psychology, and to various forms of psychologism and anti-psychologism (see Rath 1994, Kusch 1995).
2. The problem of the applicability of logic which led to an increased interest in the methodological part of traditional logic. The reform of applied logic attempted to bring philosophy in touch with the stormy development of mathematics and sciences in that time.

Both reform procedures had a destructive effect on the shape of logic and philosophy. The struggle with psychologism led to the departure of psychology (especially in its new, experimental form) from the body of philosophy at the beginning of the twentieth century. Psychology became a new, autonomous scientific discipline. The debate on methodology resulted in the creation of the philosophy of science being finally separated from the body of logic. The philosopher's ignorance of the development of formal logic caused a third departure: Part of formal logic was taken from the domain of the competence of philosophy and incorporated into mathematics where it was instrumentalized for foundational tasks. This was the philosophical background of the emergence of symbolic logic in Germany and especially the logical work of the German mathematician Schröder.

3.2. The Mathematical Context in Germany

3.2.1. Logic and Formal Algebra

The examination of the British situation in mathematics at the time when the new logic emerged has shown that the creators of the new logic were basically interested in a reform of mathematics by establishing an abstract view of mathematics which focused not on mathematical objects like quantities but on symbolic operations with arbitrary objects. The reform of logic was only secondary. These results can be transferred to the situation in Germany without any problem.

Schröder was the most important representative of the German algebra of logic.[21] He was regarded as having completed the Boolean period in logic (see Bocheński 1956, 314). In his first pamphlet on logic, *Der Operationskreis des Logikkalkuls* (1877), he presented a critical revision of Boole's logic of classes, stressing the idea of the duality between logical addition and logical multiplication introduced by Jevons in 1864. In 1890, Schröder started the large project of his monumental *Vorlesungen über die Algebra der Logik*, which remained unfinished, although it increased to three volumes with four parts, of which one appeared only posthumously (1890, 1891, 1895, 1905). Contemporaries regarded the first volume alone as having completed the algebra of logic (see Wernicke 1891, 196). Nevertheless, Schröder's logical theory kept, like the one of Boole, close contact to the traditional shape of logic. The introduction of the *Vorlesungen* is full of references to that time's philosophical discussion on logic. Schröder even referred to the psychologistic discussion on the foundation of logic, and never really freed his logical theory from the traditional division of logic into the theories of concept, judgment, and inference.

Schröder's opinion concerning the question as to what end logic is to be studied (see Peckhaus 1991, 1994b, 2004a) can be drawn from an autobiographical note (written in the third person), published in the year before his death. It contains his own survey of his scientific aims and results. Schröder divided his scientific production into three fields:

1. A number of papers dealing with some of the current problems of his science.
2. Studies concerned with creating an "absolute algebra," that is, a general theory of connections. Schröder stressed that these studies represent his "very own object of research" of which only little was published at that time.
3. Work on the reform and development of logic.

Schröder wrote (1901) that his aim was

> to design logic as a calculating discipline, especially making possible an exact handling of relative concepts, and, from then on, by emancipation from the routine claims of spoken language, and also to remove any breeding ground from "cliché" in the field of philosophy as well. This should prepare the ground for a scientific universal language that, widely differing from linguistic efforts like Volapük [a universal language like Esperanto, very popular in Germany at that time], looks more like a sign language than like a sound language.

Schröder's own division of his fields of research shows that he didn't consider himself a logician: His "very own object of research" was "absolute algebra," which was similar to modern abstract or universal algebra in respect to its basic problems and fundamental assumptions. What was the connection between logic and algebra in Schröder's research? From the passages quoted one could assume that they belong to two separate fields of research, but this is not the case. They were intertwined in the framework of his heuristic idea of a general science. In his autobiographical note he stressed:

> The disposition for schematizing, and the aspiration to condense practice to theory advised Schröder to prepare physics by perfecting mathematics. This required deepening of mechanics and geometry, but above all of arithmetic, and subsequently he became in time aware of the necessity to reform the source of all these disciplines, logic.

Schröder's universal claim becomes obvious. His scientific efforts served for providing the requirements to found physics as the science of material nature by "deepening the foundations," to quote a famous metaphor later used by David Hilbert (1918, 407) to illustrate the objectives of his axiomatic program. Schröder regarded the formal part of logic that can be formed as a "calculating logic," using a symbolic notation, as a *model* of formal algebra that is called "absolute" in its last state of development.

But what is "formal algebra?" The theory of formal algebra "in the narrowest sense of the word" includes "those investigations on the laws of algebraic operations ... that refer to nothing but general numbers in an unlimited number field without making any presuppositions concerning its nature" (1873, 233). Formal algebra therefore prepares "studies on the most varied number systems and calculating operations that might be invented for particular purposes" (ibid.).

It has to be stressed that Schröder wrote his early considerations on formal algebra and logic without any knowledge of the results of his British predecessors. His sources were the textbooks of Martin Ohm, Hermann Günther Graßmann, Hermann Hankel, and Robert Graßmann. These sources show that Schröder was a representative of the tradition of German combinatorial algebra and algebraic analysis (see Peckhaus 1997, ch. 6).

3.2.2. Combinatorial Analysis

Schröder developed the programmatic foundations of absolute algebra in his textbook *Lehrbuch der Arithmetik und Algebra* (1873) and the school program pamphlet *Über die formalen Elemente der absoluten Algebra* (1874). Among the sources mentioned in the textbook, Martin Ohm's (1792–1872) *Versuch eines vollkommen consequenten Systems der Mathematik* (1822) is listed. It stood in the German tradition of the algebraic and combinatorial analysis which started with the work of Carl Friedrich Hindenburg (1741–1808) and his school (see Jahnke 1990, 161–322).

Ohm (see Bekemeier 1987) aimed at completing Euclid's geometrical program for all of mathematics (Ohm 1853, V). He distinguished between number (or "undesignated number") and quantity (or "designated number") regarding the first one as the higher concept. The features of the calculi of arithmetic, algebra, analysis, and so on are not seen as features of quantities but of operations, that is, mental activities (1853, VI–VII). This operational view can also be found in the work of Graßmann, who also stood in the Hindenburg tradition.

3.2.3. General Theory of Forms

Graßmann's *Lineale Ausdehnungslehre* (1844)[22] was of decisive influence on Schröder, especially Graßmann's "general theory of forms" ("allgemeine Formenlehre") opening this pioneering study in vector algebra and vector analysis. The general theory of forms was popularized by Hankel's *Theorie der complexen Zahlensysteme* (1867).

Graßmann defined the general theory of forms as "the series of truths that is related to all branches of mathematics in the same way, and that therefore only presupposes the general concepts of equality and difference, connection and division" (1844, 1). Equality is taken as substitutivity in every context. Graßmann chooses $\frown$ as general connecting sign. The result of the connection of two elements a and b is expressed by the term $(a \frown b)$. Using the common rules for brackets we get for three elements $((a \frown b) \frown c) = a \frown b \frown c$ (§2). Graßmann restricted his considerations to "simple connections," that is, associative and commutative connections (§4). These connecting operations are synthetic. The reverse operations are called resolving or analytic connections. $a \smile b$ stands for the form which results in a if it is synthetically connected with b: $a \smile b \frown b = a$ (§5). Graßmann introduced furthermore forms in which more than one synthetic operation occur. If the second connection is symbolized with $\widehat{\frown}$ and if there holds distributivity between the synthetic operations, then the equation $(a \frown b) \widehat{\frown} c = (a \widehat{\frown} c) \frown (b \widehat{\frown} c)$ is valid. Graßmann called the second connection a connection on a higher level (§9), a terminology that might have influenced Schröder's later "Operationsstufen," that is, "levels of operations."

Whereas Graßmann applied the general theory of forms in the domain of extensive quantities, especially directed lines, that is, vectors, Hankel later used it to erect on its base his system of hypercomplex numbers (Hankel 1867). If $\lambda(a, b)$ is a general connection of objects a, b leading to a new object c, that is, $\lambda(a, b) = c$, there is a connection Θ which, applied to c and b leads again to a, that is, $\Theta(c, b) = a$ or $\Theta\{\lambda(a, b), b\} = a$. Hankel called the operation θ "thetic" and its reverse λ "lytic." The commutativity of these operations is not presupposed (ibid., 18).

3.2.4. "Wissenschaftslehre" and Logic

Graßmann had already announced that his *Lineale Ausdehnungslehre* should be part of a comprehensive reorganization of the system of sciences. His brother, Robert Graßmann (1815–1901), attempted to realize this program in a couple of writings published under the series title *Wissenschaftslehre oder Philosophie*. In its parts on logic and mathematics he anticipated modern lattice theory. He furthermore formulated a logical calculus being in parts similar to that of Boole. His logical theory was obviously independent of the contemporary German philosophical discussion on logic, and he was also not aware of his British precursors.[23] Graßmann wrote about the aims of his logic or theory of reasoning ("Denklehre") that it

> should teach us strictly scientific reasoning which is equally valid for all men of any people, any language, equally proving and rigorous. It has therefore to relieve itself from the barriers of a certain language and to treat the forms of reasoning, becoming, thus, a *theory of forms* or *mathematics*.

Graßmann tried to realize this program in his *Formenlehre oder Mathematik*, published in six brochures consisting of an introduction (1872a), a general part on "Grösenlehre" (1872b) understood as "science of tying quantities," and the special parts "Begriffslehre oder Logik" (theory of concepts or logic), "Bindelehre oder Combinationslehre" (theory of binding or combinatorics), "Zahlenlehre oder Arithmetik" (theory of numbers or arithmetic), and "Ausenlehre oder Ausdehnungslehre" (theory of the exterior or Ausdehnungslehre).

In the general theory of quantities Graßmann introduced the letters a, b, $c, \ldots$ as syntactical signs for arbitrary quantities. The letter e represents special quantities: elements, or in Graßmann's strange terminology "Stifte" (pins), that is, quantities which cannot be derived from other quantities by tying. Besides brackets, which indicate the order of the tying operation, he introduces the equality sign $=$, the inequality sign Z, and a general sign for a tie $\circ$. Among special ties he investigates joining or addition ("Fügung oder Addition") ("+") and weaving or multiplication ("Webung oder Multiplikation") ("·"). These ties can occur either as interior ties, if $e \circ e = e$, or as exterior exterior ties, if $e \circ e \; Z \; e$.

The special parts of the theory of quantities are distinguished with the help of the combinatorically possible results of tying a pin to itself. The first part, "the most simple and, at the same time, the most interior," as Graßmann called it, is the theory of concepts or logic in which interior joining $e + e = e$ and inner weaving $ee = e$ hold. In the theory of binding or combinatorics interior joining $e + e = e$ and exterior weaving $ee \; Z \; e$ hold; in the theory of numbers or arithmetic exterior joining $e + e \; Z \; e$ and interior weaving $ee = e$ hold, or $1 \times 1 = 1$ and $1 \times e = e$. Finally, in the theory of the exterior or Ausdehnungslehre, the "most complicated and most exterior" part of the theory of forms, exterior joining $e + e \; Z \; e$ and exterior weaving $ee \; Z \; e$ hold (1872a, 12–13).

Graßmann thus formulated Boole's Law of Duality using his interior weaving $ee = e$, but he went beyond Boole in allowing interior joining $e + e = e$, so coming close to Jevons's system of 1864.

In the theory of concepts or logic, Graßmann started with interpreting the syntactical elements, which had already been introduced in a general way. Now, everything that can be a definite object of reasoning is called "quantity." In this new interpretation, pins are initially set quantities not being derived from other quantities by tying. Equality is interpreted as substitutivity without value change, inequality as impossibility of such a substitution. Joining is read as "and," standing for adjunction or the logical "or." Weaving is read as "times," that is, conjunction or the logical "and." Graßmann introduced the signs $<$ and $>$ to express sub- and superordination of concepts. The sign $\leqq$ expresses that a concept equals or that it is subordinated another concept. This is exactly the sense of Schröder's later basic connecting relation of subsumption or inclusion. In the theory of concepts, Graßmann expressed this relation in a shorter way with the help of the angle sign $\angle$. The sign T stands for the All or the totality, the sum of all pins. The following laws hold: $a + \mathrm{T} = \mathrm{T}$ and $a\mathrm{T} = a$. 0 is interpreted as "the lowest concept, which is subordinate to all concepts." Its laws are $a + 0 = a$ and $a \cdot 0 = 0$. Finally Graßmann introduced the "not" ("Nicht") or negation as complement with the laws $a + \overline{a} = \mathrm{T}$ and $a \cdot \overline{a} = 0$.

3.3. Schröder's Algebra of Logic

3.3.1. Schröder's Way to Logic

In his work on the formal elements of absolute algebra (1874) Schröder investigated operations in a manifold, called domain of numbers ("Zahlengebiet"). "Number" is, however, used as a general concept. Examples for numbers are "proper names, concepts, judgments, algorithms, numbers [of arithmetic], symbols for quantities and operations, points, systems of points, or any geometrical object, quantities of substances, etc." (Schröder 1874, 3). Logic is, thus, a possible interpretation of the structure dealt with in absolute algebra. Schröder assumed that there are operations with the help of which two objects from a given manifold can be connected to yield a third that also belongs to that manifold (ibid., 4). He chooses from the set of possible operations the noncommutative "symbolic multiplication"

$$c = a \,.\, b = ab$$

with two inverse operations

$$\text{measuring ("Messung")} \quad b \,.\, (a : b) = a,$$
$$\text{and division ("Teilung")} \quad \frac{a}{b} \,.\, b = a.$$

Schröder called a direct operation together with its inverses "level of operations" ("Operationsstufe"). And again Schröder realized that "the logical addition of concepts (or individuals)" follows the laws of multiplication of real numbers.

But there is still another association with logic. In his *Lehrbuch*, Schröder speculated about the relation between an "ambiguous expression" like $\sqrt{a}$

and its possible values. He determined five logical relations, introducing his subsumption relations. Be A an expression that can have different values a, a', $a,''\dots$. Then the following relations hold (Schröder 1873, 27–29):

$$\textit{Superordination} \quad A \mathrel{\ni\!\!=} \begin{cases} a \\ a' \\ a'' \\ \vdots \end{cases} .$$

Examples: metal $\mathrel{\ni\!\!=}$ silver; $\sqrt{9} \mathrel{\ni\!\!=} -3$.

$$\textit{Subordination} \quad \left.\begin{matrix} a \\ a' \\ a'' \\ \vdots \end{matrix}\right\} \mathrel{=\!\!\in} A.$$

Examples: gold $\mathrel{=\!\!\in}$ metal; $3 \mathrel{=\!\!\in} \sqrt{9}$.

$$\textit{Coordination} \quad a \mathrel{\ni\!\!\in} a' \mathrel{\ni\!\!\in} a'' \mathrel{\ni\!\!\in} \cdots .$$

Examples: gold $\mathrel{\ni\!\!\in}$ silver [in respect to the general concept "metal"] or $3 \mathrel{\ni\!\!\in} -3$ [in respect to the general concept $\sqrt{9}$].

$$\textit{Equality} \quad A = B$$

means that the concepts A and B are identical in intension and extension.

$$\textit{Correlation} \quad A(=)B$$

means that the concepts A and B agree in at least one value.

Schröder recognized that if he would now introduce *negation*, he would have created a complete terminology that allows one to express all relations between concepts (in respect to their extension) with short formulas which can harmonically be embedded into the schema of the apparatus of the mathematical sign language (ibid., 29).

Schröder wrote his logical considerations of the introduction of the *Lehrbuch* without having seen any work of logic in which symbolic methods had been applied. It was while completing a later sheet of his book that he came across Robert Graßmann's *Formenlehre oder Mathematik* (1872a). He felt urged to insert a comprehensive footnote running over three pages for hinting at this book (Schröder 1873, note, pp. 145–147). There he reported that Graßmann used the sign + for the "collective comprehension," "really regarding it as an *addition*—one could say a 'logical' addition—that has besides the features of common (numerical) addition the basic feature $a + a = a$." He wrote that he was most interested in the role the author had assigned to multiplication regarded as the product of two concepts which unite the marks being common to both concepts.

In the *Programmschrift* of 1874, Schröder also gave credit to Robert Graßmann, but mentioned that he had recently found out that the laws of the logical operations had already been developed before Graßmann "in a classical work" by Boole (Schröder 1874, 7).

3.3.2. Logic as a Model of Absolute Algebra

In 1877 Schröder published his *Operationskreis des Logikkalkuls*, in which he developed the logic of Boole's *Laws of Thought* stressing the duality of the logical operations of addition and multiplication.[24] An "Operationskreis" (circle of operations) is constituted by more than one direct operation together with their inverses. The "logical calculus" is the set of formulas which can be produced in this circle of operations. Schröder called it a characteristic mark of "*mathematical logic* or the *logical calculus*" that these derivations and inferences can be done in form of calculations, namely, in the first part of logic as calculation with concepts leading to statements about the objects themselves, that is, categorical judgments, or, in Boole's terminology, "primary propositions." In its second part the logical calculus deals with statements about judgments as in conditional sentences, hypothetical or disjunctive judgments, or Boole's secondary propositions. In this booklet Schröder simplified Boole's calculus, stressing, as mentioned, the duality between logical addition and logical multiplication and, thus, the algebraic identity of the structures of these operations.

Schöder developed his logic in a systematic way in the *Vorlesungen über die Algebra der Logik* (1890–1905) designing it as a means for solving logical problems (see Peckhaus 1998, 21–28). Again he separated logic from its structure. The structures are developed and interpreted in several fields, beginning from the most general field of "domains" ("Gebiete") of manifolds of arbitrary distinct elements, then classes (with and without negation), and finally propositions (vol. 2, 1891). The basic operation in the calculi of domains and classes is subsumption, that is, identity or inclusion. Schröder presupposes two principles, *reflexivity* $a \preccurlyeq a$, and *transitivity* "If $a \preccurlyeq b$ and at the same time $b \preccurlyeq c$, then $a \preccurlyeq c$." Then he defines "identical zero" ("nothing") and "identical one" ("all"), "identical multiplication" and "identical addition," and finally negation. In the sections dealing with statements without negation, he proves one direction of the distributivity law for logical addition and logical multiplication, but shows that the other side cannot be proved; he rather shows its independence by formulating a model in which it does not hold, the "*logical calculus with groups*, e.g. functional equations, algorithms or calculi." He thereby found the first example of a nondistributive lattice.[25]

Schröder devoted the second volume of the *Vorlesungen* to the calculus of propositions. The step from the calculus of classes to the calculus of propositions is taken with the help of an alteration of the basic interpretation of the formulas used. Whereas the calculus of classes was bound to a spatial interpretation especially in terms of the part–whole relation, Schröder used in the calculus of propositions a temporal interpretation taking up an idea of Boole from his *Laws of Thought* (1854, 164–165). This may be illustrated regarding subsumption as the basic connecting relation. In the calculus of classes, $a \Subset b$ means that the class a is part of or equal to the class b. In the calculus of propositions, this formula may be interpreted in the following way (Schröder 1891, §28, p. 13):

> In the time during which a is true is completely contained in the time during which b is true, i.e., *whenever ... a is valid b is valid* as well. In short, we will often say: "*If a* is valid, then b is valid," "*a entails b*" ..., "from a follows b."

Schröder then introduces two new logical symbols, the "sign of products" $\prod$, and the "sign of sums" $\sum$. He uses $\prod_x$ to express that propositions referring to a domain x are valid for any domain x in the basic manifold 1, and $\sum_x$ to say that the proposition is not necessarily valid for all, but for a certain domain x, or for several certain domains x of our manifold 1, that is, for at least one x (Schröder 1891, §29, 26–27).

For Schröder the use of $\sum$ and $\prod$ in logic is perfectly analogous to arithmetic. The existential quantifier and the universal quantifier are therefore interpreted as possibly indefinite logical addition or disjunction and logical multiplication or conjunction respectively. This is expressed by the following definition, which also shows the duality of $\sum$ and $\prod$ (Schröder 1891, §30, 35).

$$\sum_{\lambda=1}^{\lambda=n} a_\lambda = a_1 + a_2 + a_3 + \cdots + a_{n-1} + a_n \quad \Big| \quad \prod_{\lambda=1}^{\lambda=n} a_\lambda = a_1 a_2 a_3 \cdots a_{n-1} a_n .$$

With this Schröder had all requirements at hand for modern quantification theory, which he took, however, not from Frege but from the conceptions as developed by Charles S. Peirce (1839–1914) and his school, especially by Oscar Howard Mitchell (1851–1889).[26]

3.3.3. Logic of Relatives

Schröder devoted the third volume of the *Vorlesungen* to the "Algebra and Logic of Relatives," of which only a first part dealing with the *algebra* of relatives could be published (Schröder 1895). The algebra and logic of relatives should serve as an organon for absolute algebra in the sense of pasigraphy, or general script, that could be used to describe most different objects as models of algebraic structures.

Schröder never claimed any priority for this part of his logic, but always conceded that it was an elaboration of Charles S. Peirce's work on relatives (see Schröder 1905, XXIV).

He illustrated the power of this new tool by applying it to several mathematical topics, such as open problems of G. Cantor's set theory (e.g., Schröder 1898), thereby proving (not entirely correctly) Cantor's proposition about the equivalence of sets ("Schröder-Bernstein Theorem"). In translating Richard

Dedekind's theory of chains into the language of the algebra of relatives, he even proclaimed the "*final* goal: to come to a strictly logical *definition* of the *relative* concept '*number of—*' ['*Anzahl von—*'] from which all propositions referring to this concept can be deduced purely deductively" (Schröder 1895, 349–350). So Schröder's system comes close, at least in its objectives, to Frege's logicism, although it is commonly regarded as an antipode.

3.3.4. The Ideas of Peirce

Although Schröder found his way to an algebraic approach to logic independently of Boole, he devoted his early work to a discussion and extension of the Boolean calculus. Main reference point of his mature *Vorlesungen*, however, was the logical work of the American "polymath" (Grattan-Guinness 2004, 545) Charles S. Peirce. Peirce contributed a great wealth of ideas to modern logic. He approached logic to its full range, interested not only in symbolic logic but also in a reform of traditional syllogistics and applications in the philosophy of science.[27]

In one of his first papers on logic, Peirce improved Boole's algebra of logic by introducing the inclusive disjunction as Jevons did before him (see Peirce 1868). He introduced "inclusion" $\prec$ as basic logical operator, in an algebraic spirit both for inclusion between classes and implication between propositions (Peirce 1870, *WCSP* 2, 360). It was later taken up by Ernst Schröder as "subsumption" $\Subset$. Among the five "icons" for nonrelative logic, "Peirce's law" $\{(x \prec y) \prec x\} \prec x$ (see Peirce 1885, *WCSP* 5, 173) is outstanding. It produces an axiom system for classical propositional logic when being added to an axiom system for intuitionistic logic (see Beth 1962, 18, 128).

In the paper "A Boolian Algebra with One Constant" (*WCSP* 4, 218–221), written around 1880, but not published before 1933, Peirce suggested replacing all logical connectors by only one interpreted as "neither P nor Q," thereby anticipating the NOR operator, which was independently rediscovered by H. M. Sheffer in 1913 (Sheffer 1913).

In his paper of 1870, Peirce took the first step for developing a logic of relatives, thereby elaborating the ideas of De Morgan. He distinguished *absolute terms*, such as horse, tree, or man, from terms "whose logical form involves the conception of relation, and which require the addition of another term to complete the denotation" (*WCSP* 2, 365). He discussed *simple relative terms*, that is, two-place relatives, and *conjugate terms*, that is, three- or four-place relatives like "giver of — to —" or "buyer of — for — from —" (ibid.). In his 1880 paper "On the Algebra of Logic," he took up the topic, now speaking of *singular reference* for nonrelative terms and of *dual* and *plural relatives* for two- and more-place relatives. The most elaborated form of his algebra of relatives can be found in his 1885 paper, where he combined it with the theory of quantification, the foundation of which had been formulated entirely independently of Frege by Oscar Howard Mitchell in Peirce's Johns Hopkins logic circle (Mitchell 1883). Whereas Mitchell had developed a system limited to a theory of quantified propositional functions with two prenex quantifiers, Peirce developed quantifiers as operators on propositional functions over specific domains.[28] In his 1885 paper, Peirce gave credit to Mitchell in the following way (*WCSP* 5, 178):

> All attempts to introduce this distinction [of *some* and *all*] into the Boolian algebra were more or less complete failures until Mr. Mitchell showed how it was to be effected. His method really consists in making the whole expression of the propositions consist of two parts, a pure Boolian expression referring to an individual and a Quantifying part saying what individual this is.

Peirce now used an index notation to express relatives. In the first-order part of his logic (first-intentional logic), $x_i y_j$ signifies that x is true of the individual i while y is true of the individual j. The quantifiers Σ and Π are used in analogy to their arithmetical meaning. $\Sigma_i x_i$ means that x is true of some one of the individuals denoted by i, $\Pi_i x_i$ means that x is true of all these individuals. Applied to a ordinary language example: Let l_{ij} denote that i is a lover of j, and b_{ij} that i is a benfactor of j. Then $\Pi_i \Sigma_j l_{ij} b_{ij}$ means that everything is at once a lover and a benefactor of something (*WCSP* 5, 180).

Peirce added considerations on second-intentional logic, that is, second-order logic (ibid., 185–190) and many valued logic (ibid., 166). In later work he used furthermore "existential graphs" for a graphical representation of quantificational logic (see *CP* 4.293–584) which inspired several modern systems for graphical representations of logic (see, e.g., Sowa 1993, 1997).

Peirce's logical considerations were integral part of his triadic category system with firstness (possibility), secondness (existence), and thirdness (law), his semiotics, and his triadic theory of reasoning with deduction, induction, and abduction (see Hilpinen 2004, 622–628, 644–653).

Most of Peirce's path-breaking thoughts remained unpublished during his lifetime. What he was able to publish, however, excited his contemporary logicians. The best example is Schröder, whose *Vorlesungen* were deeply influenced by Peirce, even more, long passages read as critical comments on Peirce's papers, especially on the seminal papers "On the Algebra of Logic" (Peirce 1880, 1885). In an intermediate word separating the halfs of volume two of the *Vorlesungen* Schröder wrote that after the completion of the first half of volume two in June 1891 he had hoped to publish the second half with the logic of relatives in the autumn of the same year, but (Schröder 1905, XXIV):

> It is true, seldom in my life an estimation of mine failed to the same extent as then, when I judged the extension and the seriousness of the gaps in my manuscript. This was due to the fact that the only writing that seemed to be useful, Mr. Peirce's paper on relatives [Peirce 1885], that became indeed the main basis of my volume three, has only a size of 18 pages in print (that could be printed

on half the number of my pages), and that I thought, that I could get away with a largely reproducing report. I became aware of the enormous significance of this paper when I worked at it in detail.

4. Conclusions

Like the British tradition, but independent of it, the German algebra of logic was connected to new trends in algebra. It differed from its British counterpart in its combinatorial approach. In both traditions, algebra of logic was invented within the enterprise to reform basic notions of mathematics which led to the emergence of structural abstract mathematics. The algebraists wanted to design algebra as "pan-mathematics," that is, as a general discipline embracing all mathematical disciplines as special cases. The independent attempts in Great Britain and Germany were combined when Schröder learned about the existence of Boole's logic in late 1873, early 1874. Finally he enriched the Boolean class logic by adopting Peirce's theory of quantification and adding a logic of relatives according to the model of Peirce and De Morgan.

The main interest of the new logicians was to use logic for mathematical and scientific purposes, and it was only in a second step, but nevertheless an indispensable consequence of the attempted applications, that the reform of logic came into the view. What has been said of the representatives of the algebra of logic also holds for the proponents of competing logical systems such as Gottlob Frege or Giuseppe Peano. They wanted to use logic in their quest for mathematical rigor, something questioned by the stormy development in mathematics.

For quite a while, the algebra of logic remained the first choice for logical research. Authors like Alfred North Whitehead (1841–1947), and even David Hilbert and his collaborators in the early foundational program (see Peckhaus 1994c) built on this direction of logic, whereas Frege's mathematical logic was widely ignored. The situation changed only after the publication of Whitehead's and B. Russell's *Principia Mathematica* (1910–1913). But even then important work was done in the algebraic tradition as the contributions of Clarence Irving Lewis (1883–1964), Leopold Löwenheim (1878–1957), Thoralf Skolem (1887–1963), and Alfred Tarski (1901–1983) prove.

Notes

1. Independently of each other, Gregorius Itelson, André Lalande, and Louis Couturat suggested at the 2nd Congress of Philosophy at Geneva in 1904 to use the name "logistic" for, as Itelson said, the modern kind of traditional formal logic. The name should replace designations like "symbolic," "algorithmic," "mathematical logic," and "algebra of logic," which were used synonymously up to then (see Couturat 1904, 1042).

2. For a book-length biography, see MacHale (1985). See also contemporary obituaries and biographies like Harley (1866), Neil (1865), both reprinted. For a

comprehensive presentation of Boole's logic in the context of British mathematics, see Grattan-Guinness (2000a).

3. Whately (1826). Risse (1973) lists 9 editions up to 1848 and 28 further printings to 1908. Van Evra (1984, 2) mentions 64 printings in the United States to 1913.

4. This opinion can be found in a letter of De Morgan's to Spalding of 26 June 1857 (quoted in Heath 1966, xii) which was, however, not sent. Boole lists Hamilton among the "two greatest authorities in logic, modern and ancient" (1847, 81). The other authority is Aristotle. This reverence to Hamilton might not be without irony because of Hamilton's disregard of mathematics.

5. See Hamilton 1859–1866, vol. 4 (1866), 287.

6. See his list of symbols in "Logical Notation" in Hamilton 1859–1866, vol. 4 (1866), 469–486.

7. For the priority struggle, see Heath 1966.

8. For diagrammatic methods in logic, see Gardner (1958), Bernhard (2000).

9. See the section "On Expression and Interpretation" in Boole (1847), 20–25, in which Boole gives his reading of the traditional theory of judgment. The section is followed by an application of his notation to the theory of conversion (ibid., 26–30) and of syllogism (ibid., 31–47).

10. This is the subtitle of Boole's *Mathematical Analysis of Logic* (1847).

11. For the influence of Aristotelian logic on Boole's philosophy of logic, see Nambiar (2000).

12. On the mathematical background of Boole's *Mathematical Analysis of Logic*, see Laita (1977), Panteki (2000).

13. On Gregory with focus on his contributions to the foundations of the calculus see Allaire and Bradley (2002).

14. On Boole's "psychologism," see Bornet (1997) and Vasallo (2000).

15. For the development of quantification theory in the algebra of logic, see Brady (2000).

16. The reason was that already the factorization of $x^3 = x$ leads to uninterpretable expressions. On Boole's *Laws of Thought* see Van Evra (1977); on the differences between Boole's earlier and later logical theory see Grattan-Guinness (2000b).

17. On initial reactions see Grattan-Guinness (2000a), 54–59.

18. For a discussion of De Morgan's logic see Grattan-Guinness (2000a), 25–37; Merrill (1990); Sánchez Valencia (2004), 408–410, 487–515.

19. On MacColl and his logic see Astroh and Read (1998).

20. See for the following chs. 3 and 4 of Peckhaus (1997), and Vilkko (2002).

21. On Schröder's biography, see his autobiographical note, Schröder (1901), which became the base of Eugen Lüroth's widely spread obituary, Lüroth (1903). See also Peckhaus (1997), 234–238; and Peckhaus (2004a).

22. On the various aspects of H. G. Graßmann's work, see Schubring (1996); Lewis (2004).

23. On Robert Graßmann's logic and his anticipations of lattice theory see Mehrtens (1979); Peckhaus (1997), 248–250. On the relation zwischen Schröder and the Graßmann brothers see Peckhaus (1996).

24. On Schröder's algebra of logic see Peckhaus (2004a); Sánchez Valencia (2004), 477–487; Brady (2000).

25. See Schröder (1890), 280. On Peirce's claim to have proved the second form as well (Peirce 1880, 33) see Houser (1991). On Schröder's proof see Peckhaus (1994a), 359–374; Mehrtens (1979), 51–56.

26. See Mitchell (1883), Peirce (1885). On the development of modern quantification theory in the algebra of logic see Brady (2000); Peckhaus (2004b). For Mitchell's biography, see Dipert (1994).

27. For recent work on Peirce's Logic, see Houser, Van Evra, and Roberts (1997); Brady (2000); Grattan-Guinness (2000a), 140–156; Hilpinen (2004).

28. Brady (2000), 6; see Peirce (1883). For Peirce's interpretation of Mitchell see also Haaparanta (1993), 112–116.

References

Allaire, Patricia R., and Bradley, Robert E. 2002. Symbolical algebra as a foundation for calculus: D. F. Gregory's contribution. *Historia Mathematica* 29, 395–426.

Astroh, Michael, and Read, Stephen, eds. 1998. *Proceedings of the Conference Hugh MacColl and the Tradition of Logic. Ernst-Moritz-Arndt-Universität Greifswald, March 29–April 1, 1998*, special issue of *Nordic Journal of Philosophical Logic* 3, no. 1–2.

Babbage, Charles. 1864. *Passages from the Life of a Philosopher*. Longman, Green, Longman, Roberts, and Green: London. Reprinted Gregg: Westmead 1969.

Bain, Alexander. 1870. *Logic*, 2 vols., Pt. 1: *Deduction*, Pt. 2: *Induction*. Longmans, Green: London.

Batóg, Tadeusz, and Murawski, Roman. 1996. Stanisław Piątkiewicz and the beginnings of mathematical logic in Poland. *Historia Mathematica* 23: 68–73.

Bekemeier, Bernd. 1987. *Martin Ohm (1792–1872): Universitäts- und Schulmathematik in der neuhumanistischen Bildungsreform*. Vandenhoeck & Ruprecht: Göttingen. (*Studien zur Wissenschafts-, Sozial- und Bildungsgeschichte der Mathematik*, vol. 4.)

Beneke, Friedrich Eduard. 1839. *Syllogismorum analyticorum origines et ordinem naturalem*. Mittler: Berlin.

Beneke, Friedrich Eduard. 1842. *System der Logik als Kunstlehre des Denkens*, 2 vols. F. Dümmler: Berlin.

Bentham, George. 1827. *An Outline of a New System of Logic. With a Critical Examination of Dr. Whately's "Elements of Logic."* Hunt and Clark: London. Reprinted Thoemmes: Bristol 1990.

Bernhard, Peter. 2000. *Euler-Diagramme. Zur Morphologie einer Repräsentationsform in der Logik*. Mentis: Paderborn.

Beth, Willem Evert. 1962. *Formal Methods*. Reidel: Dordrecht.

Bocheński, Joseph Maria. 1956. *Formale Logik*. Alber: Freiburg/München (*Orbis Academicus*, III, 2), 1978.

Boole, George. 1844. On a general method in analysis. *Philosophical Transactions of the Royal Society of London for the Year MDCCCXLIV*, Pt. 1, 225–282.

Boole, George. 1847. *The Mathematical Analysis of Logic. Being an Essay towards a Calculus of Deductive Reasoning*. Macmillan, Barclay, and Macmillan: Cambridge and George Bell: London. Reprinted Basil Blackwell: Oxford 1951.

Boole, George. 1848. The calculus of logic. *Cambridge and Dublin Mathematical Journal* 3: 183–198.

Boole, George. 1854. *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities*. Walton & Maberly: London. Reprinted Dover: New York [c. 1958].

Boole, George. 1952. *Studies in Logic and Probabilities*, ed. R. Rhees. Watts: London.

Boole, George. 1997. *Selected Manuscripts on Logic and its Philosophy*, ed. Ivor Grattan-Guinness and Gérard Bornet. Birkhäuser Verlag: Basel, Boston and Berlin (*Science Networks. Historical Studies*, vol. 20).

Bornet, Gérard. 1997. Boole's psychologism as a reception problem. In Boole (1997), xlvii–lviii.

Boyer, Carl B. 1968. *A History of Mathematics.* John Wiley & Sons: New York, London, and Sydney.

Brady, Geraldine. 2000. *From Peirce to Skolem. A Neglected Chapter in the History of Logic*. Elsevier: Amsterdam (*Studies in the History and Philosophy of Mathematics*, vol. 4).

Cannon, Susan Faye. 1978. *Science in Culture: The Early Victorian Period.* Dawson Scientific History Publications: New York.

Carroll, Lewis. 1887. *The Game of Logic.* Macmillan: London. Reprinted in Carroll (1958).

Carroll, Lewis. 1896. *Symbolic Logic.* Macmillan: London. Reprinted in Carroll (1958).

Carroll, Lewis. 1958. *Mathematical Recreations of Lewis Carroll. Symbolic Logic and The Game of Logic (both Books Bound as One).* Dover Publications and Berkeley Enterprises: New York.

Couturat, Louis. 1904. *L'Algèbre de la logique.* Gauthier-Villars: Paris (*Scientia. Phys.-mathématique*, vol. 24), 1914. Reprint of the 2nd ed., Olms: Hildesheim 1965.

De Morgan, Augustus. 1842. The foundations of algebra, pt. 2. *Transactions of the Cambridge Philosophical Society* 7: 173–187. Reprinted in Ewald (ed.) (1996), 336–348.

De Morgan, Augustus. 1846. On the syllogism I. On the structure of the syllogism, and on the application of the theory of probabilities to questions of argument and authority. *Transactions of the Cambridge Philosophical Society* 8: 379–408. Reprinted in De Morgan (1966), 1–21.

De Morgan, Augustus. 1847. *Formal Logic: or, The Calculus of Inference, Necessary and Probable.* Taylor and Walton: London. Reprinted De Morgan (1926).

De Morgan, Augustus. 1850. On the syllogism: II. On the symbols of logic, the theory of the syllogism, and in particular of the copula. *Transactions of the Cambridge Philosophical Society* 9: 79–127. Reprinted De Morgan (1966), 22–68.

De Morgan, Augustus. 1860. On the syllogism: IV and on the logic of relations. *Transactions of the Cambridge Philosophical Society* 10: 331–358. Reprinted in De Morgan (1966), 208–242.

De Morgan, Augustus. 1966. *On the Syllogism and Other Logical Writings*, ed. Peter Heath. Routledge & Kegan Paul: London.

Dipert, Randall R. 1994. The live and logical contributions of O. H. Mitchell: Peirce's gifted student. *Transactions of the Charles S. Peirce Society* 30: 515–542.

Enros, Philip C. 1983. The analytical society (1812–1813): Precursor of the renewal of Cambridge mathematics. *Historia Mathematica* 10: 24–47.

Ewald, William, ed. 1996. *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*, 2 Vols. Clarendon Press: Oxford.

Ferreirós, José. 1999. *Labyrinth of Thought. A History of Set Theory and Its Role in Modern Mathematics.* Birkhäuser Verlag: Basel, Boston, and Berlin, (*Science Networks. Historical Studies*, vol. 23).

Gabbay, Dov M. and Woods, John, eds. 2004. *Handbook of the History of Logic*, Vol. 3 *The Rise of Modern Logic: From Leibniz to Frege*. Elsevier North Holland: Amsterdam.

Gardner, Martin. 1958. *Logic Machines and Diagrams*. McGraw-Hill: New York, University of Chicago Press: Chicago, 1982.

Gasser, James, ed. 2000. *A Boole Anthology. Recent and Classical Studies in the Logic of George Boole*. Kluwer: Dordrecht, Boston, and London (*Synthese Library*, vol. 291), 1–27.

Graßmann, Hermann Günther. 1844. *Die lineale Ausdehnungslehre ein neuer Zweig der Mathematik dargestellt und durch Anwendungen auf die übrigen Zweige der Mathematik, wie auch auf die Statik, Mechanik, die Lehre vom Magnetismus und die Krystallonomie erläutert*. Otto Wigand: Leipzig, 1878.

Graßmann, Robert. 1872a. *Die Formenlehre oder Mathematik*. R. Grassmann: Stettin. Reprinted in Graßmann (1966).

Graßmann, Robert. 1872b. *Die Grösenlehre. Erstes Buch der Formenlehre oder Mathematik*. R. Grassmann: Stettin. Reprinted in Graßmann (1966).

Graßmann, Robert. 1966. *Die Formenlehre oder Mathematik*, mit einer Einführung von J. E. Hofmann. Georg Olms: Hildesheim.

Grattan-Guinness, Ivor. 2000a. *The Search for Mathematical Roots, 1870–1940. Logics, Set Theories and the Foundations of Mathematics from Cantor through Russell to Gödel*. Princeton University Press: Princeton and Oxford.

Grattan-Guinness, Ivor. 2000b. On Boole's algebraic logic after *The Mathematical Analysis of Logic*. In Gasser (ed.) (2000), 213–216.

Grattan-Guinness, Ivor. 2004. The mathematical turns in logic. In Gabbay and Woods (eds.) (2004), 545–556.

Grattan-Guinness, Ivor, ed. 1994. *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*, 2 vols. Routledge: London and New York.

Gregory, Duncan Farquharson. 1840. On the real nature of symbolical algebra. *Transactions of the Royal Society of Edinburgh* 14: 208–216.

Gregory, Duncan Farquharson. 1842. On a difficulty in the theory of algebra. *Cambridge Mathematical Journal* 3: 153–159.

Haaparanta, Leila. 1993. Peirce and the logic of logical discovery. In *Charles S. Peirce and the Philosophy of Science. Papers from the Harvard Sesquicentennial Congress*, ed. Edward C. Moore, 105–118. University of Alabama Press: Tuscaloosa and London.

Hailperin, Theodore. 1981. Boole's algebra isn't Boolean algebra. *Mathematics Magazine* 54: 172–184. Reprinted in Gasser (ed.) (2000), 61–77.

Hailperin, Theodore. 2004. Algebraical logic 1685–1900. In Gabbay and Woods (eds.) (2004), 323–388.

Halsted, George Bruce. 1878. Boole's logical method. *Journal of Speculative Philosophy* 12: 81–91.

Hamilton, William. 1833. Logic. In reference to the recent English treatises on that science. *Edinburgh Review* 66 (April 1833): 194–238. Again in Hamilton, *Discussions on Philosophy and Literature, Education and University Reform. Chiefly from the Edinburgh Review; Corrected, Vindicated, Enlarged, in Notes and Appendices*, Longman, Brown, Green and Longmans: London, and Maclachlan and Stewart: Edinburgh 1851, 16–174.

Hamilton, William. 1859–1866. *Lectures on Metaphysics and Logic*, 4 vols., eds. H. L. Mansel and J. Veitch. William Blackwood and Sons: Edinburgh and London.

Hamilton, William. 1866. Extract from prospectus of "Essay towards a new analytic of logical form." In Hamilton (1859–1866), vol. 4 (1866), 251–254.

Hankel, Hermann. 1867. *Theorie der complexen Zahlensysteme insbesondere der gemeinen imaginären Zahlen und der Hamilton'schen Quaternionen nebst ihrer geometrischen Darstellung*. Leopold Voss: Leipzig (Hankel, *Vorlesungen über die complexen Zahlen und ihre Functionen*, pt. 1).

Harley, Robert. 1866. George Boole, F. R. S. *British Quarterly Review* July 1866. Reprinted in Boole (1952), 425–472.

Heath, Peter. 1966. Introduction. In De Morgan (1966), (*Rare Masterpieces of Philosophy and Science*), vii–xxxi.

Hegel, Georg Wilhelm Friedrich. 1812/1813. *Wissenschaft der Logik*, Vol. 1: *Die objective Logik*, Johann Leonhard Schrag: Nürnberg. Critical edition: Hegel, *Wissenschaft der Logik. Erster Band. Die objektive Logik (1812/1813)*, eds. Friedrich Hogemann and Walter Jaeschke. Felix Meiner: Hamburg 1978 (Hegel, *Gesammelte Werke*, vol. 11).

Hegel, Georg Wilhelm Friedrich. 1830. *Encyclopädie der philosophischen Wissenschaften im Grundrisse. Zum Gebrauch seiner Vorlesungen. Dritte Ausgabe*, Oßwald'scher Verlag: Heidelberg. Critical edition Hegel, *Enzyklopädie der philosophischen Wissenschaften im Grundrisse* (1830), eds. Wolfgang Bonsiepen and Hans-Christian Lucas. Felix Meiner: Hamburg 1992 (Hegel, *Gesammelte Werke*, vol. 20).

Hilbert, David. 1918. Axiomatisches Denken. *Mathematische Annalen* 78: 405–415.

Hilpinen, Risto. 2004. Peirce's logic. In Gabbay and Woods (eds.) (2004), 611–658.

Houser, Nathan. 1991. Peirce and the law of distribution. In *Perspectives on the History of Mathematical Logic*, ed. Thomas Drucker, 10–33. Birkäuser: Boston, Basel, and Berlin.

Houser, Nathan, Van Evra, James, and Roberts, Don D., eds. 1997. *Studies in the Logic of Charles Sanders Peirce*. Bloomington: Indiana University Press.

Jahnke, Hans Niels. 1990. *Mathematik und Bildung in der Humboldtschen Reform*. Vandenhoeck & Ruprecht: Göttingen (*Studien zur Wissenschafts-, Sozial- und Bildungsgeschichte der Mathematik*, vol. 8).

Jevons, William Stanley. 1864. *Pure Logic or the Logic of Quality apart from Quantity with Remarks on Boole's System and the Relation of Logic and Mathematics*. E. Stanford: London. Reprinted in Jevons (1890), 3–77.

Jevons, William Stanley. 1874. *The Principles of Science. A Treatise on Logic and Scientific Method*, 2 vols. Macmillan: London [New York 1875].

Jevons, William Stanley. 1877–1878. John Stuart Mill's philosophy tested. *Contemporary Review* 31 (1877/78): 167–82, 256–275; 32 (1878): 88–99. Again in Jevons (1890), 137–172.

Jevons, William Stanley. 1890. *Pure Logic and Other Minor Works*, eds. Robert Adamson and Harriet A. Jevons. Macmillan: London/New York; reprinted Thoemmes Press: Bristol 1991.

Kant, Immanuel. 1787. *Kritik der reinen Vernunft*, 2nd ed. Johann Friedrich Hartknoch: Riga. Again in *Kant's gesammelte Schriften*, vol. 3, ed. Königlich Preußische Akademie der Wissenschaften. Reimer: Berlin 1911.

Kneale, William and Kneale, Martha. 1962. *The Development of Logic*. Clarendon Press: Oxford. Reprinted 1986.

Kusch, Martin. 1995. *Psychologism. A Case Study in the Sociology of Philosophical Knowledge*. Routledge: London and New York (*Psychological Issues in Science*).

Laita, Luis María. 1977. The influence of Boole's search for a universal method in analysis on the creation of his logic. *Annals of Science* 34: 163–176. Reprinted in Gasser (ed.) (2000), 45–59.

Lewis, Albert C. 2004. The unity of logic, pedagogy and foundations in Grassmann's mathematical work. *History and Philosophy of Logic* 25: 15–36.

Liard, Louis. 1877a. Un nouveau système de logique formelle. M. Stanley Jevons. *Revue philosophique de la France et de l'Étranger* 3: 277–293.

Liard, Louis. 1877b. La logique algébrique de Boole. *Revue philosophique de la France et de l'Étranger* 4: 285–317.

Liard, Louis. 1878. *Les logiciens anglais contemporains*. Germer Baillière: Paris, 1907.

Liard, Louis. 1880. *Die neuere englische Logik*, ed. J[ohannes] Imelmann. Denicke's Verlag: Berlin, 1883.

Lindsay, Thomas M. 1871. On recent logical speculation in England. In Ueberweg (1871), 557–590.

Lüroth, Jakob. 1903. Ernst Schröder †. Mitglied der Deutschen Mathematiker-Vereinigung. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 12: 249–265. Again in Schröder (1905), III–XIX.

MacColl, Hugh. 1877/78. The calculus of equivalent statements (second paper). *Proceedings of the London Mathematical Society* 9: 177–186.

MacHale, Desmond. 1985. *George Boole: His Life and Work*. Boole Press: Dublin (*Profiles of Genius Series*, vol. 2).

Mehrtens, Herbert. 1979. *Die Entstehung der Verbandstheorie*. Gerstenberg: Hildesheim (*arbor scientiarum*, vol. A.VI).

Merrill, Daniel D. 1990. *Augustus De Morgan and the Logic of Relations*. Kluwer: Dordrecht, Boston and London (*New Synthese Historical Library*, vol. 38).

Mill, John Stuart. 1843. *A System of Logic, Ratiocinative and Inductive. Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*, 2 vols. J. W. Parker: London.

Mitchell, Oscar Howard. 1883. On a new algebra of logic. In Charles S. Peirce (ed.), *Studies in Logic. By Members of the Johns Hopkins University*. Little, Brown: Boston. Reprinted John Benjamins: Amsterdam and Philadelphia 1983 (*Foundations in Semiotics*, vol. 1), 72–106.

Nambiar, Sriram. 2000. The influence of Aristotelian logic on Boole's philosophy of logic: The reduction of hypotheticals to categoricals. In Gasser (ed.) (2000), 217–239.

Neil, Samuel. 1864. John Stuart Mill. *British Controversialist and Literary Magazine* n. s. (1864): 161–173, 241–256.

Neil, Samuel. 1865. The late George Boole. LL. D., D. C. L., professor of mathematics, Queen's College, Cork; author of "The laws of thought", etc. *British Controversialist and Literary Magazine* n. s., no. 80 (August 1865): 81–94; no. 81 (September 1865): 161–174. Reprinted in Gasser (ed.) (2000), 1–25.

Ohm, Martin. 1822. *Versuch eines vollkommen consequenten Systems der Mathematik*, Pt. 1: *Arithmetik und Algebra enthaltend*, Pt. 2: *Algebra und Analysis des Endlichen enthaltend*. Reimer: Berlin; Jonas: Berlin 1828/1829.

Panteki, Maria. 2000. The mathematical background of George Boole's *Mathematical Analysis of Logic* (1847). In Gasser (ed.) (2000), 167–212.

Peacock, George. 1830. *A Treatise on Algebra.* J. & J. J. Deighton: Cambridge, and G. F. & J. Rivington: London.

Peacock, George. 1834. Report on the recent progress and present state of certain branches of analysis. *Report of the Third Meeting of the British Association for the Advancement of Science Held at Cambridge in 1833.* John Murray: London, 185–352.

Peckhaus, Volker. 1991. Ernst Schröder und die "pasigraphischen Systeme" von Peano und Peirce. *Modern Logic* 1, no. 2/3 (Winter 1990/91): 174–205.

Peckhaus, Volker. 1994a. Leibniz als Identifikationsfigur der britischen Logiker des 19. Jahrhunderts. In *VI. Internationaler Leibniz-Kongreß. Vorträge I. Teil, Hannover, 18.–22.7.1994*, 589–596. Gottfried-Wilhelm-Leibniz-Gesellschaft: Hannover.

Peckhaus, Volker. 1994b. Wozu Algebra der Logik? Ernst Schröders Suche nach einer universalen Theorie der Verknüpfungen. *Modern Logic* 4: 357–381.

Peckhaus, Volker. 1994c. Logic in transition: The logical calculi of Hilbert (1905) and Zermelo (1908). In *Logic and Philosophy of Science in Uppsala. Papers from the 9th International Congress of Logic, Methodology and Philosophy of Science*, eds. Dag Prawitz and Dag Westerståhl. Kluwer: Dordrecht, Boston, and London (*Synthese Library*, vol. 236), 311–323.

Peckhaus, Volker. 1995. *Hermann Ulrici (1806–1884). Der Hallesche Philosoph und die englische Algebra der Logik. Mit einer Auswahl von Texten Ulricis und einer Bibliographie seiner Schriften.* Hallescher Verlag: Halle a. S. (*Schriftenreihe zur Geistes- und Kulturgeschichte. Texte und Dokumente*).

Peckhaus, Volker. 1996. The Influence of Hermann Günther Grassmann and Robert Grassmann on Ernst Schröder's Algebra of Logic. In Schubring (ed.) (1996), 217–227.

Peckhaus, Volker. 1997. *Logik, Mathesis universalis und allgemeine Wissenschaft. Leibniz und die Wiederentdeckung der formalen Logik im 19. Jahrhundert.* Akademie Verlag: Berlin (*Logica Nova*).

Peckhaus, Volker. 1998. Hugh MacColl and the German algebra of logic. *Nordic Journal of Philosophical Logic* 3, nos. 1–2: 17–34.

Peckhaus, Volker. 1999. 19th century logic between philosophy and mathematics. *Bulletin of Symbolic Logic* 5: 433–450.

Peckhaus, Volker. 2004a. Schröder's logic. In Gabbay and Woods (eds.) (2004), 557–609.

Peckhaus, Volker. 2004b. Calculus ratiocinator versus characteristica universalis? The two traditions in logic, revisited. *History and Philosophy of Logic* 25: 3–14.

Peirce, Charles Sanders. 1868. On an improvement in Boole's calculus of logic. *Proceedings of the America Academy of Arts and Sciences* 7: 250–261; *CP* 3.1–19; *WCSP* 2, 12–23.

Peirce, Charles Sanders. 1870. Descriptions of a notation for the logic of relatives. *Memoirs of the American Academy of Arts and Sciences* 9 (1873): 317–378. Opprints printed 1870. Reprinted *CP* 1.27–98; *WCSP* 2, 359–429.

Peirce, Charles Sanders. 1880. On the algebra of logic. *American Journal of Mathematics* 3: 15–57; *WCSP* 4, 163–208.

Peirce, Charles Sanders. 1883. Note B: The logic of relatives. In *Studies in Logic. By Members of the Johns Hopkins University.* Little, Brown: Boston. Reprinted in *Studies in Logic by Members of the Johns Hopkins University (1883)*, ed. Charles S. Peirce. John Benjamins: Amsterdam and Philadelphia 1983, 187–203; *WCSP* 4, 453–466.

Peirce, Charles Sanders. 1885. On the algebra of logic. A contribution to the philosophy of notation. *American Journal of Mathematics* 7: 180–202; *WCSP* 5, 162–203.

Peirce, Charles Sanders. 1931–1935. *Collected Papers of Charles Sanders Peirce*, eds. Charles Hartshorne and Paul Weiss. Harvard University Press: Cambridge, Mass., 1960 [*CP*].

Peirce, Charles Sanders. 1982–2000. *Writings of Charles S. Peirce. A Chronological Edition*, eds. J. W. Kloesel et al., 6 vols. Indiana University Press: Bloomington [*WCSP*].

Rabus, Georg Leonhard. 1880. *Die neuesten Bestrebungen auf dem Gebiete der Logik bei den Deutschen und Die logische Frage*. Deichert: Erlangen.

Rath, Matthias. 1994. *Der Psychologismus in der deutschen Philosophie.* Karl Alber Verlag: Freiburg and München.

Richards, Joan L. 1988. *Mathematical Visions. The Pursuit of Geometry in Victorian England.* Academic Press: Boston.

Risse, Wilhelm. 1973. *Bibliographia Logica. Verzeichnis der Druckschriften zur Logik mit Angabe ihrer Fundorte.* Vol. 2: *1801–1969.* Olms: Hildesheim/New York (*Studien und Materialien zur Geschichte der Philosophie*, vol. 1).

Sánchez Valencia, Victor. 2004. The algebra of logic. In Gabbay and Woods (eds.) (2004), 389–544.

Schröder, Ernst. 1873. *Lehrbuch der Arithmetik und Algebra für Lehrer und Studirende*, Vol. 1 [no further volumes published]: *Die sieben algebraischen Operationen.* B. G. Teubner: Leipzig.

Schröder, Ernst. 1874. *Über die formalen Elemente der absoluten Algebra.* Schweizerbart'sche Buchdruckerei: Stuttgart. At the same time supplement to the program of the Pro- und Real-Gymnasium in Baden-Baden for 1873/74.

Schröder, Ernst. 1877. *Der Operationskreis des Logikkalkuls.* Teubner: Leipzig. Reprinted as "special edition," Wissenschaftliche Buchgesellschaft: Darmstadt 1966.

Schröder, Ernst. 1890. *Vorlesungen über die Algebra der Logik (exakte Logik)*, vol. 1. B. G. Teubner: Leipzig; reprinted Schröder 1966.

Schröder, Ernst. 1891. *Vorlesungen über die Algebra der Logik (exakte Logik)*, vol. 2, pt. 1. B. G. Teubner: Leipzig. Reprinted Schröder (1966).

Schröder, Ernst. 1895. *Vorlesungen über die Algebra der Logik (exakte Logik)*, vol. 3, pt. 1: *Algebra und Logik der Relative.* B. G. Teubner: Leipzig. Reprinted Schröder (1966).

Schröder, Ernst. 1898. Ueber zwei Definitionen der Endlichkeit und G. Cantor'sche Sätze. *Nova Acta Leopoldina. Abhandlungen der Kaiserlich Leop.-Carol. Deutschen Akademie der Naturforscher* 71, no. 6: 301–362.

Schröder, Ernst. 1901, anon. Grossherzoglich Badischer Hofrat Dr. phil. Ernst Schröder[,] ord. Professor der Mathematik an der Technischen Hochschule in Karlsruhe i. Baden. In *Geistiges Deutschland. Deutsche Zeitgenossen auf dem Gebiete der Literatur, Wissenschaften und Musik.* Adolf Eckstein: Berlin-Charlottenburg no year [1901].

Schröder, Ernst. 1905. *Vorlesungen über die Algebra der Logik (exakte Logik)*, vol. 2, pt. 2, ed. Karl Eugen Müller. B. G. Teubner: Leipzig. Reprinted Schröder (1966).

Schröder, Ernst. 1966. *Vorlesungen über die Algebra der Logik (exakte Logik)*, ["second edition"], 3 vols. Chelsea: Bronx, N.Y.

Schubring, Gert, ed. 1996. *Hermann Günther Graßmann (1809–1877): Visionary Mathematician, Scientist and Neohumanist Scholar. Papers from a Sesquicentennial Conference.* Dordrecht: Kluwer. (Boston Studies in the Philosophy of Science, vol. 187).

Sheffer, Henry Maurice. 1913. A set of five independent postulates for Boolean algebras, with application to logical constants. *Transactions of the American Mathematical Society* 14: 481–488.

Sowa, John F. 1993. Relating diagrams to logic. In *Proceedings on Conceptual Graphs for Knowledge Representation*, eds. Guy W. Mineau, Bernard Moulin, and John F. Sowa, 1–35. Springer-Verlag: New York.

Sowa, John F. 1997. Matching logical structure to linguistic structure. In Houser, Van Evra, and Roberts (eds.) (1997), 418–444.

Trendelenburg, Friedrich Adolf. 1842. Zur Geschichte von Hegel's Logik und dialektischer Methode. Die logische Frage in Hegel's Systeme. Eine Auffoderung zu ihrer wissenschaftlichen Erledigung. *Neue Jenaische Allgemeine Literatur-Zeitung* 1, no. 97, 23 April 1842, 405–408; no. 98, 25 April 1842, 409–412; no. 99, 26 April 1842, 413–414. Separately published in Trendelenburg *Die logische Frage in Hegel's System. Zwei Streitschriften.* Brockhaus: Leipzig 1843.

Ueberweg, Friedrich. 1857. *System der Logik und Geschichte der logischen Lehren.* Adolph Marcus: Bonn.

Ueberweg, Friedrich. 1871. *System of Logic and History of Logical Doctrines*, translated from the German, with notes and appendices by Thomas M. Lindsay. Longmans, Green: London. Reprinted Thoemmes Press: Bristol 1993.

Ulrici, Hermann. 1855. Review of Boole (1854). *Zeitschrift für Philosophie und philosophische Kritik* 27: 273–291.

Van Evra, James. 1984. Richard Whately and the rise of modern logic. *History and Philosophy of Logic* 5: 1–18.

Venn, John. 1894. *Symbolic Logic*, 2nd ed., "revised and rewritten." Macmillan: London. Reprinted Chelsea Publishing: Bronx, N.Y. 1971.

Vasallo, Nicla. 2000. Psychologism in logic: some similarities between Boole and Frege. In Gasser (ed.) (2000), 311–325.

Vilkko, Risto. 2002. *A Hundred Years of Logical Investigations. Reform Efforts of Logic in Germany 1781–1879.* Mentis: Paderborn.

Wernicke, Alexander. 1891. Review of Schröder (1891). *Deutsche Litteraturzeitung* 12: cols. 196–197.

Whately, Richard. 1826. *Elements of Logic. Comprising the Substance of the Article in the Encyclopaedia Metropolitana: with Additions, &c.* J. Mawman: London.

Whitehead, Alfred North, Russell, Bertrand. 1910–1913. *Principia Mathematica*, 3 vols. Cambridge University Press: Cambridge.

Windelband, Wilhelm. 1904. Logik. In *Die Philosophie im Beginn des zwanzigsten Jahrhunderts. Festschrift für Kuno Fischer*, vol. 1, ed. Windelband, 163–186. Carl Winter: Heidelberg.

5

Gottlob Frege and the Interplay between Logic and Mathematics

Christian Thiel

Gottlob Frege (1848–1925) has been called the greatest logician since Aristotle, but it is a brute fact that he failed to gain influence on the mathematical community of his time (although he was not ignored, as some have claimed), and that the depth and pioneering character of his work was—paradoxically—acknowledged only after the collapse of his logicist program due to the Zermelo–Russell antinomy in 1902. Because of this lack of influence in his time, a leading historian of logic and mathematics has gone so far as to deny Frege a place in the development of mathematical logic. Other historiographers of science, however, are convinced that the history of visible effects of great ideas on science and scientific communities should be complemented by the recognition even of solitary insights ineffective at their time, because the intellectual status of such insights or discoveries will yield most valuable (and otherwise unobtainable) information about the structure and quality of the community that made them possible by providing, as it were, the native soil for their development. Knowledge of this kind is not historically useless.

The neglect of Frege by the contemporaneous scientific community has two very different reasons. First, there is little doubt that Frege maneuvered himself out of the mainstream of foundational research (or rather, never succeeded in joining this mainstream) by his insistence on using his newly developed "Begriffsschrift," a logical notation the sophistication and analytical power of which the experts of the nineteenth century (as, in fact, most of those of the twentieth and the early twenty-first centuries) failed to recognize. And second, the double disadvantage of working in the no-man's-land between formal logic and mathematics, and of teaching at the then relatively unimportant small university of Jena gave Frege a low status in the academic world.

The superficiality of reception is manifest, for example, in Georg Cantor's (1885) review of *Grundlagen* (1884) where Cantor criticizes in condescending manner an allegedly Fregean definition of (whole) number, whereas the definition actually found in *Grundlagen* is quite different and would have been worthy of a more careful study. Frege's correction in his "Erwiderung" (1885) (which he had to publish as a—presumably paid—advertisement) went practically unnoticed. Similarly, already Ernst Schröder in his (1880) review of Frege's *Begriffsschrift* (1879) had overlooked Frege's revolutionary technique of quantification, claiming (incorrectly) that its effects could have been achieved in a much easier way by Boolean methods.

If Frege has been regarded as the founder of modern mathematical logic, this characterization refers to his creation of classical quantificational logic in his *Begriffsschrift* of 1879 without any predecessor. As to Frege's motivation, one can only surmise that he felt the urgent need for a logically water tight clarification of fundamental concepts of analysis like convergence, continuity, uniform continuity, and so on, the precise definition of which requires nested quantification. The mathematical output of the *Begriffsschrift* approach rested on Frege's replacement of the traditional analysis of elementary propositions into subject and predicate by the general analysis of a proposition into (in our case, propositional) function and argument(s), and its utilization for the expression of the generality of a statement (and of existence statements) by the employment of bound variables and quantifiers. For the antecedent part, classical propositional logic, Frege gave a consistent and complete (although not independent) axiom system in terms of negation and conditional, pointing out that other, equivalent axioms and also other connectives could be used, and he managed to get along with the rule of detachment and (not yet sufficiently precise) substitution rules. In quantificational logic, he restricted himself to universal quantification (which, together with negation, allows the expression of existential statements), and introduced the decisive concepts of the variability domain of a quantifier and the scope of a quantifier and of the quantified variable. The new devices enabled Frege to precisely define, for the first time, one-one relations, a logical successor and predecessor relation, and a logical heredity relation, in such a way that the arithmetical successor and heredity relations are covered as special cases, and mathematical induction can be formulated, and turns out to have a purely logical foundation.

Frege's *Foundations of Arithmetic* (*Die Grundlagen der Arithmetik*, 1884) added, after an elaborate criticism of earlier and contemporary views on the concept of number and on arithmetical statements, a "purely logical" (today dubbed "logicist") notion of whole number by defining the number n as the extension of the concept "equinumerous to the concept F_n," where F_n is a model concept with exactly n objects falling under it, and of a purely logical nature guaranteed by starting with $F_0 = \neg x = x$ and constructing F_{n+1} recursively from F_n. Frege's attainment of this notion is somewhat curious because immediately before that he had described and analyzed an attempt at defining number by abstraction directly from equinumerous concepts, but had repudiated this attempt because of difficulties that he considered insurmountable, so that he decided on the explicit definition just given. *Grundlagen* also introduced important logical distinctions like that of first-order and second-order concepts, with existence and number predicates as examples of the latter.

When Frege published *Function und Begriff* in 1891 and volume 1 of his monumental *Grundgesetze der Arithmetik* in 1893, he had already realized that extensions of concepts, naively regarded as unproblematic in the explicit definition of number, must be introduced by an abstraction principle, too. As extensions of concepts have been a main topic of traditional logic at least since the Logic of Port Royal (1662), Frege's treatment of abstraction in *Grundlagen* and in *Grundgesetze* centered around his discovery of the invariance property of statements about "abstract objects," the logicist definition of number, and the general abstraction principle (exemplified in *Grundgesetze* by Frege's fundamental law V, vide infra) are legitimate and indeed indispensable topics of the history of formal logic.

By contrast, the so-called context principle ("The meaning of a word must be asked for in the context of a proposition, not in isolation," *Grundlagen*, p. X) and the dichotomy of sense and reference developed in *Über Sinn und Bedeutung* (1892), often regarded as his most important contribution to philosophy by drawing guidelines for semantics and for a general theory of meaning, have only a negligible role in the history of logic. However, the latter distinction is put to use by Frege in explaining the informative or cognitive value even of judgments that are derived from and therefore based on purely logical premises (as, e.g., according to the logicist thesis, all nongeometrical mathematical theorems), and is of considerable interest for the philosophy of mathematics.

To derive the fundamental theorems of arithmetic precisely, that is, within a calculus incorporating strict formation rules for "well-formed formulas" and rules for the logical derivation of conclusions from premises, Frege had to revise and to augment his Begriffsschrift. The typically ambiguous "quantification axiom" (*Begriffsschrift*, pp. 51 and 62) is now neatly split into a first-order and a second-order version (*Grundgesetze* I, p. 61), but the most momentous change consists in the introduction of new terms of the general form "$\dot{\varepsilon}\Phi(\varepsilon)$," considered to be names of a new kind of objects called courses-of-values or value-ranges (*Wertverläufe*) the identity condition for which is given by an abstraction principle accepted by Frege as his fundamental law V, the fifth axiom of his new axiom system:

$$\dot{\varepsilon}\Phi(\varepsilon) = \dot{\alpha}\Psi(\alpha) \Leftrightarrow (x)(\Phi(x) = \Psi(x)),$$

where $\Phi(x)$ and $\Psi(x)$ are functions in Frege's general sense and the right side of the equivalence expresses the coincidence of their values for every argument, a state of affairs suggesting the identity of the "courses" (or graphs) of the functions in the case of mathematical functions, and thereby the terminology of "courses-of-values." Frege decided to regard true propositions as names of the "truth value" TRUE and false propositions as names of the "truth value"

FALSE, respectively, and reconstructed the traditional concepts as one-place functions, the function value of which is one of the two truth values for every argument chosen. So the courses-of-values of such functions are nothing else but the traditional extensions of concepts, or mathematically spoken, the sets or classes determined by the associated function as their defining condition. If concepts are taken as the functions in Frege's fundamental law V, we get for this special case (in modern notation),

$$\{\, x \mid F(x) \,\} = \{\, x \mid G(x) \,\} \Leftrightarrow (x)(F(x) \leftrightarrow G(x)).$$

In this way, sets have obviously been integrated into the system of *Grundgesetze*, and since Frege (linking up with the traditional logic of concepts and their extensions) considers abstraction a purely logical operation, set theory becomes (or remains) a proper part of logic. The derivation of arithmetical theorems from the revised and enlarged axiom system of *Grundgesetze* keeps well within the limits of logic, and in this sense the present set-theoretical foundation of mathematics preserves the intentions and the spirit of Fregean logicism.

It was mentioned in the beginning that Frege's *Grundgesetze* system foundered at Zermelo's and Russell's antinomy, as shown in the appendix of volume 2 of *Grundgesetze* as well as in Russell's *The Principles of Mathematics*, both published in 1903. Though Russell proposed to avoid the antinomy by his type theories, Frege suggested a repair of the axiom system by modifying his fundamental law V; it was shown only much later that this attempt, which has been called "Frege's way out," also leads to an impasse by allowing the derivation of other, more complicated antinomies. It is remarkable that the discovery and analysis of Zermelo's and Russell's antinomy was made possible only by the extraordinary precision, explicitness, and cogency of Frege's *Grundgesetze* system, which in spite of its inconsistency remained a paradigm of a well-designed logical system well into the twentieth century. Among the little-known but precious parts of *Grundgesetze*, §§90 ff. deserve to be highlighted because of their clear analysis of the nature and the necessary properties of an elementary proof theory and metalogic ("Die formale Arithmetik und die Begriffsschrift als Spiele": *Grundgesetze* II, p. IX). Attention should also be given to hitherto neglected parts like Frege's derivation of theorem χ in the appendix to *Grundgesetze*, where a diagonal argument is used to exhibit a fundamental inconsistency in the (traditional) notion of the extension of a concept (see Thiel 2003).

Even the origin of the antinomy has not been located unequivocally up to now. According to the received view in current Frege literature, fundamental law V is responsible for the equivalent of Russell's antinomy in *Grundgesetze*. This diagnosis, however, seems a bit rash. It is true that the derivation makes use of fundamental law V, but a careful analysis of it has to inspect not only the logical form of that law but also the structure of the formulae which replace the schematic letters of fundamental law V in every inference that has an instance of it as a premise. Thiel (1975) has tried to show that Frege's formation rules for function names (which include rules for forming function names by the creation of empty places in complex object names) may be too liberal by allowing impredicative function names, and that names of that kind are essentially involved in the derivation of the Zermelo–Russell antinomy in *Grundgesetze*. A decision on this claim and the questions it raises is still open.

A large part of Fregean studies in the past 50 years has been devoted to the investigation of problems that are peculiar to Frege's systems, without visible impact on the development of the mainstream of mathematical logic invoked in the second paragraph of this chapter. Topics of this kind have been skipped here in spite of their intrinsic interest (as, e.g., the "Julius Caesar problem," the permutation theorem, and the identification thesis of *Grundgesetze* §10, Frege's miscarried attempt at a referential completeness proof—which would have implied the consistency of the *Grundgesetze* system—and last but not least "Hume's principle" and "Frege's theorem"). A great thinker's legacy consists not only in far-reaching insights and efficient methods, it also comprises challenging problems, the solutions of which may sometimes occupy whole generations. Frege, by proving his theorem χ without recourse to Wertverläufe, exhibited an inconsistency (or at least an incoherence) in the traditional notion of the extension of a concept. He prompted our awareness of a situation the future analyses of which will hopefully not only deepen our systematic control of the interplay of concepts and their extensions but also improve our understanding of the historical development of the notion of "extension of a concept" and its historiographical assessment.

Primary Texts (Frege)

1879 *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* Halle a.S.: Louis Nebert; repr. in: G. Frege, *Begriffsschrift und andere Aufsätze*, ed. Ignacio Angelelli, Hildesheim: Georg Olms, 1964; Engl. *Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought* (transl. Stefan Bauer-Mengelberg), in: Jean van Heijenoort (ed.), *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931* (Cambridge, Mass.: Harvard University Press, 1967), 1–82; *Conceptual Notation. A Formula Language of Pure Thought, Modelled upon the Formula Language of Arithmetic*, in: G. Frege, *Conceptual Notation and Related Articles* (trans. and ed. Terrell Ward Bynum, Oxford: Clarendon Press, 1972), 101–203.

1884 *Die Grundlagen der Arithmetik. Eine logisch mathematische Untersuchung über den Begriff der Zahl*, Breslau: Wilhelm Koebner; crit. ed. Christian Thiel, Hamburg: Felix Meiner, 1986; Engl. in: G. Frege, *The Foundations of Arithmetic. A logico-mathematical enquiry into the concept of number* (bilingual ed., trans. J. L. Austin), Oxford: Basil Blackwell, 1950, rev. ^{2}1959).

1885 Erwiderung [to Cantor 1885, vide infra], *Deutsche Litteraturzeitung* 6, no. 28 (11 July 1885), Sp. 1030; Engl. "Reply to Cantor's Review of Grundlagen der Arithmetik," in: G. Frege, *Collected Papers on Mathematics, Logic, and Philosophy*, ed. Brian McGuinness (Basil Blackwell: 1984), 122.

1891 *Function und Begriff. Vortrag gehalten in der Sitzung vom 9. Januar 1891 der Jenaischen Gesellschaft für Medicin und Naturwissenschaft.* Jena: Hermann Pohle; Engl. in Beaney (1997), *The Frege Reader* (vide infra), 130–148.

1892 Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100: 25–50; Engl. in *The Frege Reader* (vide infra), 151–171.

1893 *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet*, I. Band, Jena: Hermann Pohle.

1903 *Grundgesetze der Arithmetik. Begriffsschriftlich abgeleitet*, II. Band, Jena: Hermann Pohle.

Secondary Literature

Beaney, Michael. 1997. *The Frege Reader*. Oxford/Malden, MA: Blackwell.

Cantor, Georg. 1885. [Review of] G. Frege, Die Grundlagen der Arithmetik. . . . *Deutsche Litteraturzeitung* 6, no. 20 (16 May 1885): 728–729.

Russell, Bertrand. *The Principles of Mathematics*. Vol. I. Cambridge: Cambridge University Press, 1903; London: George Allen & Unwin, 1937 [“Vol. I” omitted].

Schröder, Ernst. 1880. [Review of] G. Frege, Begriffsschrift. . . . *Zeitschrift für Mathematik und Physik* 25 (1880), Historisch-literarische Abtheilung, 81–94; Engl. Review of Frege’s Conceptual Notation. . . , in: G. Frege, *Conceptual Notation and Related Articles* (vide supra), 218–232.

Thiel, Christian. 1975. Zur Inkonsistenz der Fregeschen Mengenlehre. In: idem (ed.), *Frege und die moderne Grundlagenforschung. Symposium, gehalten in Bad Homburg im Dezember 1973.* Meisenheim am Glan: Anton Hain, 134–159.

Thiel, Christian. 2003. The extension of the concept abolished? Reflexions on a Fregean dilemma. In *Philosophy and Logic. In Search of the Polish Tradition. Essays in Honour of Jan Woleński on the Occasion of his 60th Birthday*, eds. Jaakko Hintikka, Tadeusz Czarnecki, Katarzyna Kijania-Placek, Tomasz Placek, and Artur Rojszczak (†). Dordrecht/Boston/London: Kluwer; *Synthese Library*, vol. 323), 269–273.

Selected Further Readings

Angelelli, Ignacio. 1967. *Studies on Gottlob Frege and Traditional Philosophy*, Dordrecht: D. Reidel.

Demopoulos, William, ed. 1995. *Frege’s Philosophy of Mathematics*. Cambridge, Mass./London: Harvard University Press.

Dummett, Michael. 1973. ^{2}1981. *Frege: Philosophy of Language*. London: Duckworth.

Dummett, Michael. 1981. *The Interpretation of Frege’s Philosophy*. Cambridge, Mass.: Harvard University Press.

Dummett, Michael. 1991. *Frege: Philosophy of Mathematics*. London: Duckworth.

Frege, Gottlob. 1969. *Nachgelassene Schriften*. Hans Hermes, Friedrich Kambartel, and Friedrich Kaulbach (eds.). Hamburg: Felix Meiner, revised and enlarged ^{2}1983; Engl. *Posthumous Writings* (trans. Peter Long/Roger White), Oxford: Basil Blackwell, 1979.

Haaparanta, Leila and Jaakko Hintikka, eds. 1986. *Frege Synthesized. Essays on the Philosophical and Foundational Work of Gottlob Frege*. Dordrecht: D. Reidel.

Kreiser, Lothar. 2001. *Gottlob Frege. Leben–Werk–Zeit.* Hamburg: Felix Meiner.

Schirn, Matthias, ed. 1996. *Frege: Importance and Legacy.* Berlin/New York: Walter de Gruyter.

Sluga, Hans D. 1980. *Gottlob Frege.* London/Boston/Henley: Routledge & Kegan Paul.

Weiner, Joan. 2004. *Frege Explained. From Arithmetic to Analytic Philosophy.* Chicago/La Salle, Ill.: Open Court.

6

The Logic Question During the First Half of the Nineteenth Century

Risto Vilkko

Immanuel Kant wrote, in the preface to the second edition of his *Kritik der reinen Vernunft*, that

> since Aristotle it [logic] has not required to retrace a single step, unless, indeed, we care to count as improvements the removal of certain needless subtleties or the clearer exposition of its recognized teaching, features which concern the elegance rather than the certainty of the science. It is remarkable also that to the present day this logic has not been able to advance a single step, and is thus to all appearance a closed and completed body of doctrine. (*KrV*, B VIII)

Kant's division of logic into its general and transcendental aspects served, during the early nineteenth century, as the basis for the removal of philosophers of logic into, roughly speaking, two opposing camps of the Herbartian formal logicians and the Hegelian idealist metaphysicians. Also it can be assumed that Kant's disbelief in the possibilities of logic to develop any further from its alleged Aristotelian perfection discouraged many philosophers from trying to improve the logic proper and led most of them, instead, to studying the "applications" of logic, that is, the fields of study that are nowadays referred to as epistemology, psychology, methodology, and the philosophy of science. However, not all logicians of the early and mid-nineteenth century took Kant's conception for granted. Herbart saw a promise of further development in Drobisch's *Neue Darstellung der Logik* (Herbart 1836, 1267f.). Beneke wrote a few years later that even though Kant's conception may have felt more or less credible during the 1780s, "since then, the situation has greatly changed"

(Beneke 1842, 1). According to him, "logic has lost its unchangeable character. It has adopted a variety of such aspects of the possibility of which the old logicians, including Kant himself, had no idea" (ibid.). Boole wanted to remark, in 1854, that "syllogism, conversion, &c., are not the ultimate processes of Logic. It will be shown ... that they are founded upon, and are resolvable into, ulterior and more simple processes which constitute the real elements of method in Logic" (Boole 1854, 10). De Morgan had the courage to write in 1860 that in the field of logic "innovations have been listened to in a spirit which seems to admit that Kant's dictum about the perfection of the Aristotelian logic may possibly be false" (De Morgan 1860, 247).

After Hegel's death in 1831, there arose in the academic circles of Germany a lively discussion concerning the makings of logic both as a philosophical discipline and as a formal and fundamental theory of science which might clarify not only the logical but also the metaphysical foundations of science. In fact, this was perhaps the most popular theme in the philosophical exchange of thoughts in Germany during the mid-nineteenth century. The most characteristic slogans in the discussion were "the logic question" and "reform of logic." These slogans did not have very specific meanings. They were used rather loosely to refer to various competing efforts to reform logic. In 1880 Leonhard Rabus (1835–1916) characterized the logic question in his book on nineteenth-century German contributions in the field of logic as circling around the fundamental problems of the possibility and justification of logic (Rabus 1880, 157; see Vilkko 2002).

According to another late nineteenth-century German philosopher, Friedrich Harms (1819–1880), reform of logic could be sought from logic as (1) an organon of sciences, (2) a critique of sciences, or (3) a philosophical science (Harms 1874, 124). As an organon, or as a discipline of the methods of sciences, logic is the science of the forms of thought. As a critique, or as a theory of the necessary preconditions of knowledge and knowing, logic considers such questions as: How are objects given to cognition? What are the basic principles of knowing? What justifies these principles? And how valid are these principles? (Harms 1881, 137). Bacon's "inductive" reform covered logic as an organon, whereas Locke and Kant treated logic as a critique (ibid., 150). As far as Harms was concerned, this was, however, not enough. Harms wanted to stress that these two aspects of logic must be taken simply as two different sides of the one and the same logic. In his view, the most important aspect in this reform concentrated on the form of logic as a philosophical science. He argued as follows: Since logic is a philosophical science due to its content, it must be a philosophical science also due to its form, because the content and the form of a science must coincide. Purely formal logic is, however, originally an empirical science and thus only an instrument for philosophy. When logic is reformed as a philosophical science, it is also reformed as an organon as well as a criterion of correct and consistent thinking (ibid., 121–125, 130).

Hermann Ulrici (1806–1884) defined the logic question as "the question about the place, the context, and the working of logic" (Ulrici 1869/70, 1). He began his most important contribution to this debate, titled "Zur logischen

Frage" (1869/70), by stating that his conception of logic is incompatible not only with that of Hegel but also with every other attempt that denies the purely formal character of logic and tries to identify logic with metaphysics, epistemology, and/or theory of science. According to him, logic "deserves the name of a fundamental science; and it is clearly impossible for such logic to be at the same time also metaphysics and a theory of science" (ibid., 8).

In other words, the logic question sprung from a genuine doubt about the justification of the formal foundations of logic as the normative foundation of all scientific activity. On the one hand, most of the participants of the debate opposed Hegel's attempts to unite logic and metaphysics—on the other, reform was sought to overcome the old scholastic-Aristotelian formal logic. The discussion can thus be characterized as a battle on two fronts. In any case, the need to reform was stimulated by the developments in the field of philosophy. As Volker Peckhaus has put it:

> the reform endeavors that were released through this discussion scarcely considered the formal logic itself, but rather its psychological foundations and its use in theories of science that strove to seize the positive and formal sciences of that time. (Peckhaus 1997, 12)

The very slogan "logic question" was used for the first time by Adolf Trendelenburg (1802–1872). His writings provoked anew an awareness of the problematic philosophical position of formal logic. What is more, it was from his initiative that the reform discussion of logic really started around the turn of the 1840s. In 1842, he asked in his essay "Zur Geschichte von Hegel's Logik und dialektischer Methode" whether Hegel's dialectical method of pure thought should be treated as a scientific one. His own answer to this question was negative (Trendelenburg 1842, 414). However, of more importance was his criticism of both Herbartian formal logic and Hegelian dialectical logic in his two-volume *Logische Untersuchungen* (Trendelenburg 1840, I, 4–99). Before going into the details of Trendelenburg's criticism, let us take a closer look at Herbart's and Hegel's conceptions of the nature and the task of logic.

1. Herbart's Theory of the Structures of Thought

Johann Friedrich Herbart (1776–1841) defined philosophy as cultivation and arranging of conceptual material that is given in sense-experience (Herbart 1813, 38f.). His basic division of the field of philosophy was the that time usual tripartite one: logic, metaphysics, and aesthetics (the most important part of which was ethics). The task of logic was to take care of the first and the foremost duty of philosophy, that is, of conceptual clarity. The task of metaphysics was to justify concepts as objects of thought by analyzing and resolving conceptual contradictions that originate from thought itself. The third constituent, aesthetics, complemented the objects of thought by an analysis of values (Ueberweg 1923, 156f.).

Herbart remained in many ways faithful to Kant's conceptions. In his logic writings he himself sketched only the very basics and trusted, when it came to more advanced issues, the textbooks of, for example, Wilhelm Krug and Jakob Fries. Logic meant for him a regulative science which merely establishes the ways of handling concepts as such and lays down the law of contradiction as their highest standard. In his supplement "Hauptpuncte der Logik" to the second edition of his *Hauptpuncte der Metaphysik* Herbart wrote:

> Indeed logic is concerned with representations but not with the practice of representing. Hence, it is neither concerned with the mode and the manner of how we get to a representation, nor with the conditions of mind that are given thereby, but only with what becomes represented. (Herbart 1808, 217)

In his works, Herbart gives logic such definitions as, for example, "a general science of understanding" (Herbart 1813, 67) and "a theory of the structures of thought" (Herbart 1808, 222). Logic meant for him a fundamental science that occupies itself first of all with separating, classifying, and combining concepts as such; thereafter with making and analyzing judgments; and finally with revealing the modes of inference. The task of logic was to develop the formal consequences from the given premises (Herbart 1808, 218; 1831, 204).

The fundamental point of difference between Hegel's and Herbart's conceptions of logic dealt with the relation between logic and metaphysics. Whereas the former drew an identity between logic and metaphysics, the latter wanted to keep the two strictly separated from each other. Herbart also insisted that for the benefit of pure logic, it is necessary to avoid all psychological considerations. In the second chapter of his *Lehrbuch zur Einleitung in der Philosophie*, Herbart summarizes his logic conception in five theses:

1. Logic provides us with the most general regulations of separating, classifying, and combining concepts.
2. Logic presupposes concepts as known and does not distress itself with their specific contents.
3. Therefore logic is not really an instrument of such an investigation that aims at finding something novel. It rather gives instructions for revealing what we already know.
4. Nevertheless logic also points out the primary conditions of investigation in general and takes care of the important duty of paying attention to the possibility of committing errors.
5. The term "applied logic" refers to a combination of logic and psychology which, however, falls out as defective on issues where psychology does not already lead the way. (Herbart 1813, 41f.)

Even though it is fully justified to characterize Herbart as an advocate of formal logic and to see his philosophy of logic as an offspring of Kant's empirical realism, it can be asked why the logicians of the early nineteenth century did not seem to be very interested in making an effort toward further development of formal logic. One reasonable answer is based on the fact that was already pointed out in the beginning of this chapter. That is, because during the early nineteenth century logic sprung straight from Kant's view, according to which the inherited scholastic-Aristotelian logic was to be seen as closed and complete. The only thing there seemed to be left for logicians to deal with was its applications, such as epistemology, methodology, and the philosophy of science.

2. Drobisch's Formal Philosophy

Around the mid-nineteenth century, the perhaps most eminent opponent to the idealist identification of logic and metaphysics was the mathematician, astronomer, and philosopher Moritz Wilhelm Drobisch (1802–1896). He was one of the most important and insightful thinkers in the Herbartian school.

In the first paragraph of the introduction to his *Neue Darstellung der Logik* (1836), Drobisch introduced philosophy as "the general science" (Drobisch 1836, 1). According to him, it was not the task of philosophy to investigate the auxiliary apparatus of subjective cognition. That was the task of special sciences. For him, like for Herbart, philosophy meant working with purely conceptual material and trying to reach understanding of concepts in themselves (ibid., 2).

In *Neue Darstellung der Logik*, Drobisch focused on what he called "formal philosophy," that is, logic. He introduced logic as the doctrine of the conditions of correct and consistent thinking (ibid., 5). In his view, logic must not be understood as a description of human thinking. It must not be considered as a descriptive natural history of thinking, but rather as a normative discipline of thought in general or as a "Code of Laws of Thought" (ibid., 6). In this connection Drobisch referred to Kant's description of logic not as a descriptive but as a demonstrative a priori science, the function of which is to take care of the necessary laws of thought, that is, as the science of the adequate use of understanding and of reason in general (see *KrV*, B IX, XXII). Drobisch regretted the fact that the law-giving character of logic had been spoiled by the Kantian school. Therefore he saw it necessary to once again impress on philosophers the importance of this normative aspect of logic (Drobisch 1836, 7).

Drobisch was concerned about the purity of logic. Already in the preface to the first edition of *Neue Darstellung der Logik* he made it clear that in what follows, logic will be understood as an independent and autonomous formal foundation of all scientific activity. He wrote that "logic is, in fact, nothing but pure formalism. It is not meant to be, and must not be, anything else" (ibid., VI). Moreover, he did not consider logic as a branch of mathematics (ibid., VIII–X).

At the end of the first edition of *Neue Darstellung der Logik*, there is an exceptional and incisive logico-mathematical appendix (ibid., 127–167). In this appendix Drobisch concentrates on the problematic connection between logic and mathematics, and introduces his algebraic construction of the simplest forms of judgment and derivation of inferences founded thereupon (ibid., 131–136). In effect, he develops an extensional algebraic calculus of classes and elementary judgments. Drobisch's calculation apparatus follows Aristotelian theory of syllogisms. To give just a couple of examples, the classical modi BARBARA and DARAPTI are presented in the following way (ibid., 134f.):

$$\text{BARBARA} \quad \begin{array}{ll} M = p & \\ S = m & (< M = p') \\ \hline S = p' & (\text{where } p' < p) \end{array}$$

$$\text{DARAPTI} \quad \begin{array}{c} M = p \\ M = s \\ \hline s = p \end{array}$$

From today's perspective, Drobisch's calculus appears as one of the most interesting chapters of *Neue Darstellung der Logik*. It has been valued as an improvement in comparison with the intensional systems of his famous predecessors Ploucquet, Lambert, Gergonne, and Jacob Bernoulli (Thiel 1982, 763). Unfortunately, Drobisch removed his calculus from the subsequent editions of *Neue Darstellung der Logik*, which became one of the leading German textbooks of logic during the nineteenth century. He also executed a number of other modifications to the later editions of his book. In the preface to the second edition, he even admitted that the changes that had been carried out were so extensive that one could almost speak of two altogether different books (Drobisch 1851, III). Trendelenburg's hard criticism was undoubtedly an important catalyst for these changes. But was it the decisive one? Drobisch's uncompromising attitude toward Trendelenburg in the preface to the second edition suggest that perhaps he just felt that after all his calculus was not quite ripe to be published.

3. Hegel's Dialectical Logic

The first one of the two most important sources of Georg Wilhelm Friedrich Hegel's (1770–1831) dialectical logic is his monumental *Wissenschaft der Logik* (1812/16). This much debated and thoroughly interpreted work was Hegel's attempt to provide a comprehensive philosophical synthesis of his union of logic, metaphysics, and epistemology. The second source of Hegel's logic is *Encyclopädie der philosophischen Wissenschaften im Grundrisse*, of which Hegel prepared and published three different versions—the first one in 1817 and the two others in 1827 and 1830. Of particular interest here is the first part of the book: "Die *Logik*, die Wissenschaft der Idee an und für sich" (§§19–244). In the following we pay attention only to the third edition, which was published the year before Hegel's death. It can be regarded as Hegel's philosophical testament. Reading it this way requires, however, keeping in mind that it does not provide the reader with a thoroughly elaborated and fully developed system but only with a sketch of the foundations of one. And when considering the relations between *Wissenschaft der Logik* and the *Encyclopädie*, it is worth knowing that Hegel kept on working intensely with the former one until his death (Nicolin and Pöggeler 1959, xxix, xxxviif.).

It is no surprise that Hegel's and Herbart's thought with regard to logic did not meet, because the very foundations and the aims of the two differed from each other as greatly as day and night. The following quotation from the introduction to *Wissenschaft der Logik* gives the reader a hint of the distance between Hegel's understanding of the term "logic" and that of the Herbartians:

> logic is to be understood as the system of pure reason, as the realm of pure thought. *This realm is truth as it is without veil and in its own absolute nature*. It can therefore be said that this content [of pure science] is the *exposition of God as he is in his eternal essence before the creation of nature and a finite mind*. (Hegel 1812, 31)

Hegel built his dialectical logic on the trivet of Kant's transcendental logic, Schelling's identity between the real and the ideal, and Fichte's *Wissenschaftslehre*. The resulting theory was designed to serve the needs of his own monumental philosophical system, which divides into the three main constituents of Science of Logic, Philosophy of Nature, and Philosophy of Spirit. Logic was regarded as the foundational science of the system.

Hegel's works provide the reader with the most comprehensive theory of metaphysical philosophy of logic. In his philosophy there is no way of separating logic and metaphysics from one another. In the *Encyclopädie* Hegel states that "*logic therefore coincides with Metaphysics, the science of things set and held in thoughts*—thoughts accredited able to express the essential reality of things" (Hegel 1830, 58). If Hegel's philosophy in toto is a science about the real world of change and development, understood as the collective self-education of humanity about itself, then logic is the construction of the history of thinking. In the *Encyclopädie* Hegel defines logic as "the science of the pure Idea; pure, that is, because the Idea is in the abstract medium of Thinking" (ibid., 53). His logic does not consider the categories merely as forms of subjective thinking. They are also seen as the forms of objective Being itself. The Absolute or the Reason—which is the ultimate subject matter of Hegel's philosophy—is a union of Thinking and Being, and it is the task of logic to develop this unity. Accordingly Hegel divided his logic into two parts: (1) the objective logic concerned with the Being, and (2) the subjective logic concerned with the Thinking.

In *Wissenschaft der Logik* Hegel attacked fiercely the "dull and spiritless" (Hegel 1812, 34) attempts of formal logicians to elaborate logic as the most general deductive science of thinking. According to him, the deduction of the "so-called rules and laws, chiefly of inference is not much better ... than a childish game of fitting together the pieces of a colored picture puzzle" (ibid.). This "thinking" which formal logic strove to govern constituted for Hegel the mere form of cognition. According to him, formal logic abstracted from all content of cognition and the material constituents of knowledge are, consequently, totally independent of it. Thus, formal logic could provide only "the formal conditions of genuine cognition and cannot in its own self contain any real truth, nor even be the *pathway* to real truth because just that which is essential in truth, its content, lies outside logic" (ibid., 24).

The two most important concepts in Hegel's critique of the traditional scholastic-Aristotelian conception of logic are "formal" and "abstract." His accusation of traditional logic as being merely formal thinking is based on a conception according to which logical form and content should correspond to each other. This requirement has several consequences. It explains why logic embraces for him, in addition to the problem of classification of propositions and inferences, also the study of the categories on which these classifications are based. It implies, for example, that logic also deals with the distinctions between different levels of knowledge correlative to the various aspects of reality given in the categories. It also has implications for Hegel's conceptions of judgment and truth (Kakkuri 1983, 41).

Hegel's conclusion of the history of logic until the early nineteenth century was as follows:

> Before the dead bones of logic can be quickened by spirit, and so become possessed of a substantial, significant content, its method must be that which alone can enable it to be pure science. In the present state of logic one can scarcely recognize even a trace of scientific method. It has roughly the form of an empirical science. (Hegel 1812, 34f.)

This leads us to the central topic of *Wissenschaft der Logik*, that is, the problem of appropriate philosophical method. Hegel's starting point was the assumption that philosophy had not yet found a method of its own, but merely borrowed bits and pieces from the methodologies of various sciences and, in particular, "regarded with envy the systematic structure of mathematics" (ibid., 35). The connection between logic and philosophy is inextricable because "the exposition of what alone can be the true method of philosophical science falls within the treatment of logic itself; for the method is the consciousness of the form of the inner self-movement of the content of logic" (ibid.).

From Kant's remark of elementary logic having neither lost nor gained any ground since the time of Aristotle, Hegel drew a very radical conclusion. He held the same opinion as Kant in stating that logic had not undergone any positive changes in more than 2000 years. But judging by the logic-compendia of his time the few traceable changes appeared to him as consisting "mainly in omissions" (ibid., 33). Therefore he concluded that it is "necessary to make a completely fresh start with this science" (ibid., 6).

Hegel's logic dealt not only with the traditional Aristotelian laws of thought or Kant's logic of the understanding but also with metaphysical issues. He begun the introduction to the first book of *Wissenschaft der Logik* by regretting the fact that "what is commonly understood by logic is considered without any reference whatever to metaphysical significance" (ibid., 28). His at first relative and later absolute identification of logic and metaphysics is the fundamental point of difference between Hegel's philosophy of logic and that of the Herbartian school.

There is no undisputable clear-cut answer to the question "What kind of logic is Hegel's logic?" It is certainly not, for example, a doctrine of the laws of formally impeccable inference. Yvon Gauthier (1984) has suggested an answer that amounts to saying that Hegelian logic is a transcendental logic, which in turn would be the study of the a priori structures of logical thought. According to him, what Hegel calls objective logic is nothing less than metaphysics in the traditional sense, and therefore it is justified to consider Hegelian logic as transcendental-metaphysical. For Hegel, transcendental-speculative logic, which deals with the most general features of thought, reaches even further than it does for Kant (ibid., 303f.). Whatever the truth, in any case Hegel's program was one of the most influential efforts to reform logic during the nineteenth century.

4. Trendelenburg's Logical Investigations

Friedrich Adolf Trendelenburg (1802–1872) was not concerned with logic as mere doctrine of the laws of correct inference. The first two chapters of his greatest work, the two-volume *Logische Untersuchungen* (1840) can well be considered to discuss philosophy of logic in today's sense of the saying, but the rest of the book—the essence of Trendelenburg's logical investigations—is perhaps best characterized as fundamental epistemology with a strong metaphysical flavor. Trendelenburg's intention was to solve what he considered to be the ultimate task of philosophy, namely, the apparent correspondence between "the external reality of Being and the internal reality of Thinking" (Trendelenburg 1840, I, 110). In effect, his *Logische Untersuchungen* was an attempt first to show the defects of both Herbartian and Hegelian logic and then to supplement and reformulate them to achieve a formal and fundamental theory of science and metaphysics.

4.1. Critique of Formal Logic

When talking about formal logicians, Trendelenburg meant those philosophers who attempted to explain the pure forms of thought without paying attention to the contents of thought. This tradition rested, according to Trendelenburg, on a strict distinction between thoughts and their objects, that is, between Thinking and Being. Furthermore, because in Trendelenburg's view these so-called formal logicians took truth as simple correspondence between thoughts and their objects, they also accepted silently the presupposition of harmony between forms of thought and things in themselves. In particular Trendelenburg wanted to criticize those philosophers who subscribed to Herbartian conception of logic (Trendelenburg 1840, I, 4–6).

Formal theories of logic had traditionally begun with a theory of concepts. Accordingly, Trendelenburg begun his criticism with some critical remarks on traditional theories of concepts and specified the target of his criticism: "In particular we consider two ingenious and consistent presentations of formal logic, the famous works of A. D. Ch. Twesten [1825] and the University of Leipzig Professor Moritz Wilhelm Drobisch [1836]" (ibid., 7). He was not content with taking concepts as given and understanding them merely as sub- and superordinate combinations of properties. He criticized this view as much too naive for uncovering the secrets of the foundations of human thought. In his view, the traditional subordination of concepts was based on nothing but simple operations of adding and subtracting properties. According to him, every attempt to find the essence of thought with the help of such basic operations as these—or with any such alternatives as multiplication and division—remains always futile. Every theory that rested on such theory of concepts became thus "more than dubious" (ibid., 8).

Hence, according to Trendelenburg, the whole edifice of formal logic was built on sand. However, for the sake of argument, he assumed that there is nothing wrong with formal theories of concepts and turned to examine the fundamental principle of classical formal logic, that is, the law of identity and contradiction: "*A* is *A* and *A* is not not-*A*." Formal logicians had traditionally believed that in the final analysis everything else in logic derives from this principle. Trendelenburg wanted to find out if this belief really was tenable. Even though he admitted that the principle seemed unassailable at the first sight, he wanted the reader to pay closer attention to the latter part of it: A given concept *A* stands in contradiction with its negation and is logically equivalent with its double negation. According to Trendelenburg, this "blindly accepted" (ibid., 11) interpretation was insufficient for explaining the nuances with regard to contents of concepts. It reduced all of the various conceptual contrasts to the pure formal logical contradiction. Trendelenburg wanted to criticize this inflexibility. In his opinion, every purely formal definition for identity and negation fails to explain them properly (ibid., 11–14).

It may be difficult to understand why Trendelenburg wanted to make such an issue about formal logic not paying attention to the *contents* of judgments if one does not keep in mind that his logical investigations was an attempt to elect the best parts of metaphysics and logic and to reformulate them as a general, formal, and fundamental theory of science. In the introduction to *Logische Untersuchungen*, he wrote that "the range of these investigations must run through the sphere of logic questions and reach for insight on the whole field of science" (ibid., 3). After having scrutinized both formal logic and dialectical method, Trendelenburg announced that the rest of his book will be committed to answering, with regard to the objective foundations of logic, the question about the possibility of knowledge (ibid., 100f.). His project was more ambitious than just explicating the laws of correct and consistent deductions.

Trendelenburg also probed into different types of inference and asked if it was possible to derive all the forms of inference from what the formal logicians regarded as the basic premises of formal logic, that is, the principle of identity and the idea of concepts as combinations of properties. He found nothing to complain about the classical forms of deductive inference. However, the problematic cases were logic of induction and inferences based on analogy. "What a great shame it is," he wrote, "if there is no ability to understand the logic of induction and analogy expressed by science; and in case it is generality and not necessity that follows, then the principle of identity and contradiction is not the [basic] principle of logic" (ibid., 18). Indeed, according to Trendelenburg, inadequate understanding of logic of induction was one of the most alarming shortcomings of early nineteenth-century formal logic.

Trendelenburg dedicated the end of the first chapter to his favorite subject, Aristotelian philosophy. He had noticed that formal logicians often appealed to Aristotle and willingly called themselves Aristotelians. Trendelenburg, however, had found a number of reasons why they should not be regarded Aristotelian. According to him (ibid., 18–21):

1. Aristotle did not propose that the forms of thought should be understood purely in themselves;
2. Understanding concepts simply as given combinations of properties does not correspond to Aristotle's refined theory of concepts;
3. The nineteenth-century formulation for the principle of identity and contradiction, "*A* is *A* and *A* is not not-*A*," differs significantly from Aristotle's original formulation: "The same attribute cannot at the same time belong and not belong to the same subject" (*Met.* 1005b 18–20);
4. Aristotle did not regard affirmation and negation as purely logical forms;
5. Aristotle considered modal judgments of necessity and possibility as rooted in the nature of things;
6. Aristotle did not postulate syllogisms as merely formal relations between judgments.

This was roughly what Trendelenburg left in the hands of the public for deciding whether formal logic could be taken seriously with regard to the logic question.

4.2. Critique of Metaphysical Logic

If it was the most serious defect of Herbartian formal logic to strictly separate Thinking from Being and to concentrate only on the former one, then Hegelian metaphysical logic was guilty of exactly the opposite. According to Hegel, Thinking and Being could not be separated from each other. In his system knowledge of Being derives straight from Thinking in itself. Trendelenburg, however, could not see how this could be possible. Shortly, in his opinion the biggest fault with Hegel's conception of logic was the attempt to completely and unjustly neglect the decisive intermediate role of the Aristotelian concept of motion in obtaining knowledge (Harms 1881, 236).

Trendelenburg's criticism of Hegel's dialectical method ranges over a broad field of philosophical topics. In the following we shall, however, concentrate only on those aspects and arguments that can be regarded as belonging to the field of philosophy of logic. Trendelenburg summarized the basic situation with dialectical method in the following way:

> The dialectical method strives for the greatest possible. It wants to develop and create the pure Idea as if in a divine intellect—solely out of itself. Content and form are supposed to arise simultaneously. Because the pure Idea brings forth only what lies deep in itself, it must create such a world where nothing exists in itself and every thought is a genuine part of the totality. . . . We must, however, resign [from the dialectical method] at once. The means are too frail for executing the plan of such a titanic project. (Trendelenburg 1840, I, 94)

Trendelenburg's first argument concerned the alleged presuppositionless of Hegel's logic. According to Hegel, pure Thought needed no support from perception or sense experience. The pure Idea was the stone foundation of his logic and vice versa. Therefore, according to Hegel, logic was both quite easy and extremely hard:

> From different points of view, Logic is either the hardest or the easiest of sciences. Logic is hard, because it has to deal neither with perceptions nor, like geometry, with abstract representations of the senses, but with the pure abstractions; and because it demands a force and facility of withdrawing into pure thought, of keeping firm hold on it, and of moving in such an element. Logic is easy, because its facts are nothing but our own thought and its familiar forms or terms: and these are the acme of simplicity, the ABC of everything else. (Hegel 1830, §19)

Trendelenburg could not accept this view. In his opinion Hegel's "pure Thought" did not deserve its name because it could not escape from tacitly presupposing the fundamental principle of all knowledge, that is, the Aristotelian idea of motion. According to Trendelenburg, even the most elementary dialectical steps were impossible without support from this concealed principle: "Wherever we turn to, motion remains the presupposed vehicle of the dialectically breeding Thought. . . . This spatial motion is hereupon the first assumption of this presuppositionless logic" (Trendelenburg 1840, I, 24–29; see also Petersen 1913, 156).

Trendelenburg's second argument against the dialectical method concerned the two seemingly logical relations of negation and identity. However, Trendelenburg wanted to point out that with a closer look it can be seen that it is not the *logical* negation that works in Hegel's system:

> What is the nature of this dialectical negation? It can have a twofold character. Either it is understood in a pure logical way, so that it simply denies what the first concept affirms without replacing it with something new, or it can be understood in a real way, so that the affirmative concept is denied by a new affirmative concept, in what way both of the two must be replaced with each other. We call the first instance logical negation, and the second one real opposition. (Trendelenburg 1840, I, 31)

Is it now possible for the logical negation, Trendelenburg asked further, to stipulate such progress that from a given denial a new positive concept arises which exclusively unites in itself both the affirmation and the negation? According to Trendelenburg's definition for logical negation, this was totally out of question. In other words, it would be a mistake to treat the dialectical negation as logical contradiction. Hence, it must be regarded as a real opposition. However, if it is a real opposition, then it is unattainable from the logical point of view and Hegel's dialectic is not the dialectic of pure Thought. Hence the one who takes a closer look on the so-called negations of Hegel's dialectical logic shall in most cases discover ambiguities (ibid., 30–45).

According to the rules of Hegel's dialectical logic, identity creates a new concept of a higher level out of a given concept and its opposite. This dialectical product is the truth of its "ingredients." Hence, dialectical identity appears to be a real unit, even though it is, in the final analysis, only a kind of shallow similarity of abstraction. Trendelenburg could not see how it could be possible for two distinct concepts to mutate into a third, new one. He wrote that "dialectical identity offers more than it has" (ibid., 55). If the dialectical identity was supposed to be some kind of an impetus of the concrete reality, then it surely could not be an identity of abstraction. According to Trendelenburg, there is an obvious contradiction between the origin of the dialectical concept of identity and its alleged effect (ibid., 45–56).

Trendelenburg's third point of criticism concerned Hegel's conception of immediacy. In Aristotle's philosophy, Trendelenburg clarified, every such element of thought is immediate that does not reduce to any other element, for example, the basic elements of representation in general or certain particulars sensed in such a manner that nothing whatsoever comes in between the sensuous representation and its object. In the nineteenth-century philosophy it was, according to Trendelenburg, more customary to use the term "immediate" in the latter sense of the word. Since the whole dialectic was in the final analysis nothing but a chain of mediation, immediacy was out of question in this sense. However, in Hegel's system, the concept of immediacy is prominent everywhere in the dialectical process of mediation. Now it seemed to be, in Trendelenburg's opinion, that in this context immediacy can only mean self-subsistence, that is, Being-for-self. Hegel himself expressed this quite clearly in his *Encyclopädie*: "Being-for-self, as reference to itself, is immediacy, and as reference of the negative to itself, is a self-subsistent, the One. This unit, being without distinction in itself, thus excludes the other from itself" (Hegel 1830, §96). Later in the same volume he exclaims that "the immediate judgment is the judgment of definite Being. The subject is invested with a universality as its predicate, which is immediate, and therefore *a sensible quality*" (Hegel 1830, §172; emphasis added). In Trendelenburg's opinion, this explanation left no room for misunderstanding: In Hegel's logic the term "immediate" refers to something foreign to his system, that is, to something sensuous. And above we just saw how in Hegel's dialectic the function of immediacy was by no means supposed to lead the way from pure thoughts to something sensuous. Thus, Hegel's use of the term "immediate" remains ambiguous (Trendelenburg 1840, I, 56–59).

Trendelenburg closed the first chapter of *Logische Untersuchungen* by estimating whether it is right to regard the Herbartian projects of formal logic as latest extensions to the Aristotelian tradition of logic. Accordingly, at the end of the second chapter, he paid attention to dialectical method having sought for its original from Plato's *Parmenides* dialogue (ibid., 89).

The latter part of the *Parmenides* dialogue (137c–166c), where Socrates and Parmenides discuss the intertwined concepts of the one and many and the problematic relations between parts and wholes, has been interpreted in many different ways since time immemorial. Hegel recognized a resemblance between Parmenides's holistic concept of One and his own Absolute. Admittedly there are certain similarities. However, there are also other ways of understanding the passage. Another possibility is to read *Parmenides* simply as Plato's reply to his critics. According to a number of scholars, the safest principle of interpretation is to excavate the hints that Plato himself gives to the reader for understanding the dialogue. Trendelenburg subscribed to this strategy. One of these hints is also the heart of Trendelenburg's last argument against Hegel: In the beginning of the latter part of the dialogue Parmenides suggests that Socrates could use some training in the art of dialectic so that he might be more successful in searching for solutions to Parmenides's philosophical dilemmas (135c–136a). Thus, the arguments and proofs of the latter part can be regarded as merely heuristic. It can be read as just an evaluation of various juxtaposed philosophical arguments and theses, some of which reappear in other dialogues by Plato—others being mere formal supplements to Parmenides's arguments. Trendelenburg also held the opinion that it is hard to find a credible uniform philosophical doctrine hidden beneath Parmenides's lesson to Socrates (Trendelenburg 1840, I, 89).

The task of Hegel's dialectic was to show how a closed system could seize the whole reality. The outcome, however, failed to convince Trendelenburg. It seemed to him evident that the role of perception is silently assumed everywhere in Hegel's theory and that the concepts of the pure Thought are, in the final analysis, nothing but diluted representations. "Intuition is," he wrote in the closing pages to the second chapter of *Logische Untersuchungen*, "vital for human thought and it starves to death if it must try to live on its own entrails" (ibid., 96). At the end of the whole work, Trendelenburg glanced back to his work and wrote:

> As we have seen, formal logic is essential but not sufficient for accomplishing the logical task. Hegel's dialectic, in its turn, gives a promise of more—as a matter of fact of the greatest that can be imagined—but falls out as impossible. (Trendelenburg 1840, II, 363)

As the discussion on the reform of logic moved on, there appeared certain general points of agreement concerning the basic nature and the task of logic. It became common to accept that the possible reform of logic must go hand in hand with the reform of philosophy. The Kantian appreciation of mathematics against its Hegelian devaluation became rehabilitated even though the question about the relationship between logic and mathematics remained difficult. On the one hand formal logic became almost resistant to philosophical criticism, but on the other hand it lost at least part of its prestige as the foremost constituent of philosophy proper as it gradually was transformed into a subdiscipline of mathematics.

5. Herbartian and Hegelian Reactions to the Criticism

Trendelenburg's *Logische Untersuchungen* had a devastating impact in both the Herbartian and the Hegelian camps. Academic public expected the leading representatives from both sides to formulate and present counterarguments. This they also did. The leading Herbartian philosophers were, however, a little slower in defending themselves than their Hegelian colleagues.

Even though Trendelenburg explicitly aimed his censure at Drobisch and August Twesten (1789–1876), apparently most of those who took part in the evaluation discussion of *Logische Untersuchungen* read the first chapter of the book as censure of Herbartian logic and metaphysics. Drobisch wrote and published several articles in defense of Herbart. Twesten did not reply on *Logische Untersuchungen.*

It took more than 10 years before Drobisch was ready to step forth with counterarguments. The first set of his answers was published in 1851 in the preface to the second edition of his *Neue Darstellung der Logik* (1851). The second one came out the year after, in the form of a journal article (Drobisch 1852). Before these two contributions, only Hermann Kern had dared to defend Herbartian philosophy against Trendelenburg's authority. Kern published an essay for justifying Herbartian metaphysics nine years after *Logische Untersuchungen* (Kern 1849). In addition to Drobisch and Kern, Ludwig Strümpell appears to be the only eminent Herbartian who had the courage to defend Herbartian philosophy in public against Trendelenburg with his essay "Einige Worte über Herbart's Metaphysik in Rücksicht auf die Beurtheilung derselben durch Herrn Professor Trendelenburg" (1855). Even

Trendelenburg himself was astonished to find out how long it took for his opponents to prepare any answers to his criticism (Trendelenburg 1855, 317).

In the second edition of *Neue Darstellung der Logik* (1851), Drobisch still subscribed to formal logic. In general, Drobisch accused Trendelenburg of making formal logic appear as if it was philosophically much less sophisticated than it really was. In particular he emphasized that Trendelenburg's statement about formal logic totally separating thoughts from their objects is incorrect:

> Formal logic *does not* presuppose *pure* thought and does not attempt to analyze or explain the forms of thought *in abstracto*. ... Formal logic does not recognize forms *without* content. It only recognizes such forms that are *independent* of *particular* contents which they might fulfil. Contents, which they cannot completely do without, remain thus *indeterminate* and *accidental*. (Drobisch 1851, IV)

Drobisch also still held that there is no insurmountable dividing wall between Aristotelian logic and formal logic (ibid., III–XIV).

A year later Drobisch admitted, in a journal article "Ueber einige Einwürfe Trendelenburg's gegen Herbart'sche Metaphysik" (1852), that it might have been a good idea to reply on *Logische Untersuchungen* a little sooner. However, he thought that it still was not too late to break that silence. This article was, above all, an act in defense of Herbartian metaphysics. When it comes to Trendelenburg's arguments against Herbartian philosophy of logic, this time Drobisch only referred briefly to the preface to the second edition of his *Neue Darstellung der Logik* (ibid., 11–12)

Evidently Trendelenburg had expected more vivid reactions to the "two-edged critique" of his *Logische Untersuchungen*. This is at least what the critique itself (Trendelenburg 1840, 4–99), his reply to Drobisch (Trendelenburg 1855), and its extension (Trendelenburg 1867) suggest. Perhaps he had even planned the first two chapters of *Logische Untersuchungen* rather as an opening of a polemic than as a coup de grâce. At least Hegelian philosophers reacted a little faster.

Hegelians were naturally very sore with Trendelenburg's criticism. Differences between their published reactions were largely due to different personal temperaments. For instance, the leading Hegelian of the 1840s, Karl Rosenkranz (1805–1879), could not quite understand how "a man, who is so throughly familiar with Aristotle's philosophy, [could] have sunk so deep that he denies νοησιζ τηζ νοησεως from νους" (Rosenkranz 1844, xviif.). There were, however, other Hegelians who did not manage to keep themselves as dispassionate as Rosenkranz. Karl Michelet characterized Trendelenburg's philosophy as "jumble" (Michelet 1861, 126), and Arnold Ruge wrote, in *Deutsche Jahrbücher für Wissenschaft und Kunst*, that

> those who are dull enough to be unable to recognize the progress Hegel has stimulated have no scientific importance—and even less do they possess positive political credibility. Their work is still-born,

> a matter of deepest ignorance and complete lack of strength. Those who cannot digest Hegel, cannot digest either the heroes of the German Spirit: Luther, Leibniz and Kant! (cited in Petersen 1913, 158)

Trendelenburg's victory over both of his opponents seems to have been undisputable. In 1859 Rosenkranz confirmed, in his *Wissenschaft der logischen Idee* (1858/59), that as a consequence of Trendelenburg's censure in his *Logische Untersuchungen* the whole discussion around Hegelian dialectic had come to durable stagnation—the advance of Hegelian philosophy had ceased. Some 30 years after *Logische Untersuchungen*, Hermann Bonitz wrote, in his memorial essay to Trendelenburg, that "in any case it is true that now, after three decades, the substantial influence of Hegelian philosophy has been confined to a very modest group of faithful adherents and that Trendelenburg has had a considerable effect on this change with his criticism" (Bonitz 1872, 23). Forty years after *Logische Untersuchungen*, Friedrich Harms valued Trendelenburg's contribution to the nineteenth-century philosophy of logic as easily the most significant. Either neglecting or not knowing what for instance Gottlob Frege and Ernst Schröder had recently accomplished, he wrote that

> Trendelenburg's *Logische Untersuchungen* is the latest significant attempt to reform logic. ... we are living in a time of fragmentary efforts to reform logic. These attempts do not have any accurate continuity. They attempt to remodel logic from greatly varying starting-points and with greatly varying results. The future will make the best out of what Lotze, Ulrici, Ueberweg, Chr. Sigwart, and others have accomplished. (Harms 1881, 238)

References

Aristotle. 1928. *Metaphysica.* The Works of Aristotle. Translated into English, vol. 8, ed. W. D. Ross. Oxford: Clarendon Press.

Beneke, Friedrich E. 1842. *System der Logik als Kunstlehre des Denkens.* Berlin: Dümmler.

Bonitz, Hermann. 1872. Zur Erinnerung an Friedrich Adolf Trendelenburg. *Abhandlungen der königlichen Akademie der Wissenschaften zu Berlin aus dem Jahre 1872.* Berlin: Buchdruckerei der königlichen Akademie der Wissenschaften, 1–39.

Boole, George. 1854. *An Investigation of the Laws of Thought, on which are Founded the Mathematical Theories of Logic and Probabilities.* London: Walton & Maberly; Cambridge: Macmillan.

De Morgan, Augustus. [1860] 1966. Logic. In De Morgan, *On the Syllogism and Other Logical Writings*, ed. Peter Heath, 247–270. Reprint, London: Routledge & Kegan Paul.

Drobisch, Moritz W. 1836. *Neue Darstellung der Logik nach ihren einfachsten Verhältnissen. Nebst einem logisch-mathematischen Anhange.* Leipzig: Voß.

Drobisch, Moritz W. 1851. *Neue Darstellung der Logik nach ihren einfachsten Verhältnissen, mit Rücksicht auf Mathematik und Naturwissenschaft*, Zweite, völlig umgearbeitete Auflage. Leipzig: Voß.

Drobisch, Moritz W. 1852. Ueber einige Einwürfe Trendelenburg's gegen Herbart'sche Metaphysik. *Zeitschrift für Philosophie und philosophische Kritik* (*Neue Folge*) 21: 11–41.

Gauthier, Yvon. 1984. Hegel's Logic from a Logical Point of View. In *Logic, Methodology and Philosophy of Science VI. Proceedings of the Sixth International Congress of Logic, Methodology and Philosophy of Science, Hanover, 1979. Studies in Logic and Foundations of Mathematics*, no. 104, eds. L. Jonathan Cohen, Jerzy Łoś, Helmut Pfeiffer, and Klaus-Peter Podewski, 303–310. Amsterdam, New York, Oxford: North-Holland.

Harms, Friedrich. 1874. *Die Reform der Logik* (*aus den Abhandlungen der Königlichen Akademie der Wissenschaften zu Berlin*). Berlin: Vogt.

Harms, Friedrich. 1881. *Philosophie in ihrer Geschichte, Zweiter Theil, Geschichte der Logik*. Berlin: Hofmann.

Hegel, Georg W. [1812] 1923. *Wissenschaft der Logik. Erster Band: Die objective Logik*. Ed. Georg Lasson. Reprint, Leipzig: Meiner.

Hegel, Georg W. [1830] 1911. *Encyclopädie der philosophischen Wissenschaften im Grundrisse. Zum Gebrauch seiner Vorlesungen. Dritte Ausgabe*. Ed. Georg Lasson. Reprint, Leipzig: Meiner.

Herbart, Johann F. [1808] 1887. Hauptpuncte der Logik. In the 2nd edition of *Hauptpuncte der Metaphysik*, ed. Karl Kehrbach, 217–226. Reprint, Langensalza: Beyer & Söhne.

Herbart, Johann F. [1813] 1891. *Lehrbuch zur Einleitung in die Philosophie*, ed. Karl Kehrbach. Reprint, Langensalza: Beyer & Söhne.

Herbart, Johann F. [1831] 1897. *Kurze Encyklopädie der Philosophie aus praktischen Gesichtspuncten entworfen*, ed. Karl Kehrbach. Reprint, Langensalza: Beyer & Söhne.

Herbart, Johann F. 1836. Rezension von Neue Darstellung der Logik nach ihren einfachsten Verhältnissen. Nebst einem logisch-mathematischen Anhange. Von M. W. Drobisch. *Göttingische gelehrte Anzeigen* 10: 1267–1274.

Kakkuri, Marja-Liisa. 1983. Abstract and concrete: Hegel's logic as logic of intensions. *Ajatus* 39: 40–106.

Kant, Immanuel. [1787] 1904. *Kritik der reinen Vernunft.* Zweite Auflage, ed. by Königlich Preußische Akademie der Wissenschaften. Reprint, Berlin: Reimer.

Kern, Hermann. 1849. *Ein Beitrag zur Rechtfertigung der herbartschen Metaphysik.* Coburg: Gymnasium in Coburg.

Michelet, Karl L. 1861. Die dialektische Methode und der Empirismus. In Sachen Trendelenburgs gegen Hegel. *Der Gedanke* 1: 111–126, 187–201.

Nicolin, Friedhelm and Otto Pöggeler. 1959. Zur Einführung. In Hegel, *Enzyklopädie der philosophischen Wissenschaften im Grundrisse*, IX–LII. Hamburg: Meiner.

Peckhaus, Volker. 1997. *Logik, Mathesis universalis und allgemeine Wissenschaft. Leibniz und die Wiederentdeckung der formalen Logik im 19. Jahrhundert.* Berlin and New York: Akademie Verlag.

Petersen, Peter. 1913. *Die Philosophie Friedrich Adolf Trendelenburgs. Ein Beitrag zur Geschichte des Aristoteles im 19. Jahrhundert.* Hamburg: Boysen.

Plato. 1953. Parmenides. In *The Works of Plato*, Translated into English with Analyses and Introductions by B. Jowett, vol. 2, 669–718. Oxford: Clarendon Press.

Rabus, Georg L. 1880. *Die neuesten Bestrebungen auf dem Gebiete der Logik bei den Deutschen und Die logische Frage.* Erlangen: Deichert.

Rosenkranz, Karl. 1844. *Georg Wilhelm Friedrich Hegel's Leben. Supplement zu Hegel's Werken.* Berlin: Duncker und Humblot.

Rosenkranz, Karl. 1858/59. *Wissenschaft der logischen Idee.* Königsberg: Bornträger.

Strümpell, Ludwig. 1855. Einige Worte über Herbart's Metaphysik in Rücksicht auf die Beurtheilung derselben durch Herrn Professor Trendelenburg. *Zeitschrift für Philosophie und philosophische Kritik (Neue Folge)* 27: 1–34, 161–192.

Thiel, Christian. 1982. From Leibniz to Frege: Mathematical Logic between 1679 and 1879. In *Logic, Methodology and Philosophy of Science VI. Proceedings of the Sixth International Congress of Logic, Methodology and Philosophy of Science, Hanover, 1979. Studies in Logic and Foundations of Mathematics*, no. 104, eds. L. Jonathan Cohen, Jerzy Łoś, Helmut Pfeiffer, and Klaus-Peter Podewski, 755–770. Amsterdam, New York, Oxford: North-Holland.

Trendelenburg, Adolf. 1840. *Logische Untersuchungen.* Berlin: Bethge.

Trendelenburg, Adolf. 1842. Zur Geschichte von Hegel's Logik und dialektischer Methode. Die logische Frage in Hegel's Systeme. Eine Auffoderung [*sic*] zu ihrer wissenschaftlichen Erledigung. *Neue Jenaische Allgemeine Literatur-Zeitung* 1(97): 405–408; 1(98): 409–412; 1(99): 413–414.

Trendelenburg, Adolf. [1853] 1855. Ueber Herbart's Metaphysik und eine neue Auffassung Derselben. In *Historische Beiträge zur Philosophie, Zweiter Band, vermischte Abhandlungen*, 313–351. Berlin: Bethge.

Trendelenburg, Adolf. 1867. Ueber Herbart's Metaphysik und neue Auffassung Derselben, Zweiter Artikel. In *Historische Beiträge zur Philosophie, Dritter Band, vermischte Abhandlungen*, 63–96. Berlin: Bethge.

Twesten, August. 1825. *Die Logik, insbesondere der Analytik.* Schleswig: Königlichen Taubstummen-Institut.

Ueberweg, Friedrich. 1923. *Grundriss der Geschichte der Philosophie, Vierter Theil, Die deutsche Philosophie des neunzehnten Jahrhunderts und der Gegenwart.* Berlin: Mittler und Sohn.

Ulrici, Hermann. 1869/70. Zur logischen Frage. (Mit Beziehung auf die Schriften von A. Trendelenburg, L. George, Kuno Fischer und F. Ueberweg.) I. Formale oder materiale Logik? Verhältniß der Logik zur Metaphysik. *Zeitschrift für Philosophie und philosophische Kritik* (*Neue Folge*) 55: 1–63; II. Die logischen Gesetze, ibid.: 185–237; III. Die Kategorieen, 56: 1–46; IV. Begriff, Urtheil, Schluß, ibid.: 193–250.

Vilkko, Risto. 2002. *A Hundred Years of Logical Investigations. Reform Efforts of Logic in Germany, 1781–1879.* Paderborn: Mentis.

7

The Relations between Logic and Philosophy, 1874–1931

Leila Haaparanta

One who seeks to discuss the relations between logic and philosophy in the nineteenth century and the early twentieth century has to pay special attention to his or her use of the term "logic." In the context of nineteenth-century and early twentieth-century philosophy, that term occasionally refers to similar activities to those we now call logic. In those days, logic could mean what we nowadays tend to call logic proper, that is, working with formal systems that resemble those of mathematics. However, it could also mean activities that we would now wish to label as "epistemology," "philosophy of science," "philosophy of language," or "philosophy of logic." Therefore, it may sound strange to promise to discuss the relations between nineteenth-century and early twentieth-century logic and philosophy. It is more to the point to claim that this chapter gives a survey of the field of philosophy where (1) the philosophical foundations of modern logic were discussed and (2) where such themes of logic were discussed that were on the borderline between logic and other branches of the philosophical enterprise, such as metaphysics and epistemology. What will be excluded in this chapter are the formal developments on the borderline between logic and mathematics, hence, contributions made by such logicians as Augustus De Morgan (1806–1871), George Boole (1815–1864), Ernst Schröder (1841–1902), and Giuseppe Peano (1858–1932), for example (see chapters 4 and 9 in this volume). Gottlob Frege (1848–1925) and Charles Peirce (1839–1914) are included, since their work in logic is closely related to and also strongly motivated by their philosophical views and interests. In addition, this chapter pays attention to a few philosophers to whom logic amounted to traditional Aristotelian logic and to those who commented on the nature of logic from a philosophical perspective without making any significant contribution to the development of formal logic.

The choice of the years 1874 and 1931 has its reasons. In 1874 Franz Brentano (1838–1917) held his inaugural lecture "Über die Gründe der Entmutigung auf philosophischem Gebiete" in Vienna. The lecture showed the way to philosophers who made a sharp distinction between subjective psychological acts studied by the empirical sciences and the objective contents of those acts represented by means of logic (Brentano 1929, 96). Basically the same distinction had already been made by Bernard Bolzano (1781–1848) in his *Wissenschaftslehre* (1837). In 1931, which is the last year that is taken into account in this chapter, a volume of *Erkenntnis* was published which contained Rudolf Carnap's (1891–1970) criticism of Martin Heidegger (1889–1976) titled "Überwindung der Metaphysik durch logische Analyse der Sprache" and Arendt Heyting's (1898–1980) "Die intuitionistische Grundlegung der Mathematik," which was inspired by Edmund Husserl's (1859–1938) and Heidegger's thoughts. Those articles were important in view of the division of philosophical schools in the twentieth century. This chapter is far from being the whole story of the relations between logic and philosophy 1874–1931. Instead, it consists of a number of themes and opens up a few perspectives on the period. There is slight emphasis on German philosophy. The chapter focuses on Frege, Husserl, and Peirce. Frege and Peirce are chosen because of their central role in the development of modern logic. Husserl is chosen because he wrote a great deal on the philosophical problems related to the logical enterprise. If we use the labels of our time, we would say that Husserl was one of the most important philosophers of logic of his own time.

1. The Historical Setting, 1874–1931

Even if Kant thought that no significant changes are possible in logic, his own transcendental logic raised several new themes that we could now call philosophy of logic. Transcendental logic was philosophy of certain logical categories, especially of their metaphysical limits and epistemological import. After Kant, the role of those categories was discussed in various ways. There were philosophers such as Johann Gottlieb Fichte (1764–1814), Friedrich Wilhelm Schelling (1775–1854), and G. W. F. Hegel (1770–1831) who anchored logical categories to the world, who argued that logical categories are categories of being, of what there is (see chapter 6). In the second half of the century, the situation changed. Philosophers started to debate on the relation between logic and psychology. That debate increased interest in the epistemological questions related to logic, but it also brought about a new formulation of metaphysical or ontological problems. The basic question was no longer what the most general structure of the world is. Instead, philosophers pondered on whether there was a specific abstract realm that had thoughts as its denizens and that logic could represent.

In his *Lehrbuch der Logik* (1920), Theodor Ziehen listed and characterized various groups of nineteenth-century logic (Ziehen 1920, 155–216). The main opposition in the last decades of the century was that between psychologists and antipsychologists. In the first half of the century, there were a number of psychologists such as Friedrich Beneke (1798–1854), Otto Friedrich Gruppe (1804–1876), William Whewell (1794–1866), and August Comte (1798–1857). In Germany several schools arose in the late nineteenth century that attacked psychologism, such as neo-Kantians like Hermann Cohen (1842–1918) and Paul Natorp (1854–1924) and logicists like Frege and Husserl. Husserl's early views are usually considered psychologistic. Among logicists there were philosophers whom Ziehen called value-theoretical logicists, such as Wilhelm Windelband (1848–1915) and Heinrich Rickert (1863–1936), and moderate logicists like Hermann Lotze (1817–1881) and Gustav Teichmüller (1832–1888). Moderate or weak psychologists included several thinkers, for example, Christoph Sigwart (1830–1904), Wilhelm Wundt (1832–1920), Benno Erdmann (1851–1922), and Theodor Lipps (1851–1914).[1] The contents of these doctrines will be clarified later in this chapter.

The opposition between logic and metaphysics became important in a new way in the beginning of the twentieth century, when Carnap raised criticism against Heidegger's views. Researchers who have tried to trace the origin of the distinction between the analytic and the phenomenological, more generally, the Continental tradition, have paid attention to the debate between Carnap and Heidegger concerning the relation between logic and metaphysics. Various interpretations can be proposed concerning the core of Carnap's criticism. It is not clear how Heidegger would have defended his view or attacked Carnap's position. Michael Friedman (1996, 2000), among others, has studied the theme by taking the historical context into account. He has argued that the roots of Carnap's thought were in the neo-Kantianism of the Marburg school, while Heidegger's philosophy ensued from the Southwest school. According to Friedman, that difference largely explains the fact that Carnap emphasized the role of logic, while Heidegger stressed the centrality of questions concerning human beings and their values.

There are various ways of making the distinction between the analytic and the phenomenological tradition. Several criteria have been suggested, such as their attitudes toward the history of philosophy, toward their own history, toward science, and toward the idea of scientific philosophy; their views on what are the central problems of philosophy, the objects of philosophical research, and the methods of philosophy; and their attitudes toward the ideal of clarity in philosophy. It has been suggested that views on the relation between logic and metaphysics are an important criterion if we wish to divide philosophers into the two camps. The different criteria turn out to be problematic in closer scrutiny.[2] One who seeks to locate the differences between the two traditions cannot ignore the fact that there were at least two ideas that early phenomenology, especially Husserl's thought, and most of early analytic philosophy shared. First, there was the idea of pure philosophy, which presupposed a belief in the sharp distinction between knowledge a priori and knowledge a posteriori. That belief had one of its origins in late nineteenth-century antipsychologism. Second, there was the belief in the method of analysis as the method of philosophy. That method, though in different versions, was used by Frege, the "godfather" of analytic philosophy, and Husserl, the pioneer of phenomenology. These two common features were intertwined in various ways.[3]

Carnap's and Heidegger's debate had its background in lively discussion concerning the philosophy of logic and mathematics that was going on particularly in German philosophy at the end of the nineteenth century and at the beginning of the twentieth century. Hermann Lotze (1817–1881), who was professor at the University of Göttingen, influenced a number of those who took part in the discussion. On Lotze's view, objectivity is not the same as that actuality (*Wirklichkeit*) which belongs to concrete beings. Lotze also regarded abstract objects like thoughts and values as objective in the sense that they are valid. Frege was one of Lotze's students, and so was Bruno Bauch, Frege's colleague in Jena, a neo-Kantian philosopher and the founder of the society for German idealism. Frege also belonged to that society. Heinrich Rickert, who was professor in Freiburg, was also influenced by Lotze's philosophy. Carnap was a student of Bauch's, while Heidegger was a student of Rickert's. Rickert and Windelband were central figures of the neo-Kantianism of Southwest Germany.[4] Frege's philosophical environment was not the Southwest school but rather the Marburg school. Frege was most likely to receive his concept of truth value from Windelband, who was one of the so-called value-theoretical logicists; they were philosophers who thought that besides the moral values the realm of values includes the truth values studied by logic.[5]

Like Frege, Husserl criticized psychologists and held the view that logic and mathematics study abstract objects, such as numbers and thoughts, that is, the structure of thoughts and the inferential relations between thoughts. At the beginning of the twentieth century, Husserl was professor in Göttingen, until he moved to Freiburg in 1916, to follow Rickert in the professorship. Husserl's follower in Freiburg was Heidegger. Frege was a logicist in two meanings, Husserl in one meaning of the word. A competing doctrine, namely formalism, was represented by David Hilbert (1862–1943), who was Husserl's colleague and friend in Göttingen. In the philosophy of mathematics, logicism meant two things; on the one hand, it was the view that arithmetic or even the whole of mathematics can be reduced to logic; on the other hand, it was the view that numbers are abstract objects that are independent of the human mind. On this latter view, mathematical knowledge has to do with these very objects. Frege's logicist program had to do with arithmetic, and it included defining the concept of number by means the concepts of "extension of a concept" and "equinumerous." Frege took extensions of concepts to be logical objects.[6] Husserl's studies in the foundations of logic and mathematics were closely connected to the rise of phenomenology. Logicism in the latter meaning was a natural starting point of Husserl's phenomenology; namely, if the objects of logic and mathematics have their origins in consciousness, even if not in empirical consciousness studied by psychology, as Husserl argued, one has to find out how to make the distinction between empirical and non-empirical consciousness. If psychology is interested in empirical consciousness, and if philosophy, including the philosophy that studies the relations between the subject and logic and mathematics, is interested in pure consciousness, how are these consciousnesses distinguished from each other?[7] Are we dealing with ontologically two different consciousnesses, or are we talking about two different points of view to the same consciousness? If the latter holds, what do we mean by saying that in the last analysis it is the same consciousness we are talking about, and if it is one and the same, how can we justify the claim that it is one and the same consciousness? Several difficult ontological and epistemological questions arise. Therefore, it is not surprising that Husserl moved from studies in the philosophical foundations of logic to studies of consciousness.

Husserl asked how logic as science is possible. He wished to justify the field of knowledge called logic, but it often seems that he also wished to justify a certain logic, namely, classical logic, by studying its origin in consciousness. There has been a debate on whether Husserl wished to take a position on the correctness of logical systems, and if he did, whether he was a conservative or a revisionist in logic. One has raised the question whether Husserl would have suggested giving up the law of excluded middle or any other law of logic, if we cannot find philosophical justification for those laws. Phenomenologists often emphasize the incommensurability of philosophy and the sciences. We could think that Husserl's philosophical studies and logicians' debate on the acceptability of various logics are incommensurable. Dieter Lohmar has, for his part, sought to show that Husserl was a moderate revisionist. In his view, Husserl thought that it is possible that we cannot find justification for all laws of classical logic.[8] Radical revisionists in logic were Oskar Becker (1889–1964) and Heyting, who sought to change logic on the basis of Husserl's and Heidegger's thought. They tried to develop intuitionistic logic by using Husserl's concepts of meaning intention and meaning fulfillment or disappointment.[9] As noted, one of Heyting's papers, inspired by Husserl and Heidegger, came out in the same volume of *Erkenntnis* where Carnap had his criticism of Heidegger, the criticism where Carnap sought to show by means of logic that Heidegger's sentences are meaningless.

In this chapter, I proceed as follows. First, I consider logic as a category theory, hence, as a doctrine that is interested in the categories of thought and being. I discuss those views in which logic is understood as the ideal language that mirrors reality in the right way. Husserl's formal ontology is related to these doctrines. This consideration brings us to Heidegger's view of metaphysics. I then take up the ideas of one world and of the plurality of worlds and consider the theories of modalities. The doctrine of three realms was much discussed in the late nineteenth century. I will consider the acknowledgment of the third realm and the reasons for such an acknowledgment. This question is tied to the debate on psychologism and antipsychologism and the problem of the objectivity of the realm of thoughts that logic speaks about. The important point made by antipsychologists was epistemological rather than ontological. That point brings us to the question of the possibility of logical knowledge and the various ways of answering the question given by Kant, Frege, and Husserl. I then continue with epistemological considerations and discuss Frege's idea that in logical inference no intuitive gaps are allowed. If a logical theorem is justified, there is no reference to intuition in the inferential chain. In addition, philosophers raised the question concerning the justification of traditional logical laws and a specific logical language. Husserl was one of those who raised such questions. That theme also brings us to intuitionistic logic and to the relations between logic and experience. Finally, Frege's and Peirce's methodologies of logic are discussed, and Frege's semantic views are presented.

2. The Relations between Logic, Metaphysics, and Ontology

In nineteenth-century logic and philosophy, logic was often understood contentually or materially (*inhaltlich*). The idea that logic has content received various meanings. (1) Logic was regarded as contentual in the sense that it was assumed to speak about the objects of the world. Kant's transcendental logic was contentual in this sense in a peculiar way; it showed us the form of the phenomenal world. Hegelian logic was contentual, because it sought to mirror the historical development of reality. (2) Logic was taken to be contentual in the sense of being transcendental, that is, being a picture of the a priori conditions of all thought. (3) Logic was thought to have content in the sense that it was assumed to speak about the objects of the abstract realm, that is, to convey thoughts, which were considered objective. (4) Logic was thought to have content in the sense that it was assumed to mirror the structure of the psychological realm.

Philosophers who regarded logical categories as categories of being or as categories of objects of knowledge and experience represented the first or the second position. Leibniz and Kant belonged to that tradition of logic, even if their views otherwise differed radically. Frege also thought that an ideal language can be discovered that is the correct mirror of the universe. However, Frege is also famous for his writings about objective thoughts and of his view of logic as a representative of the realm of abstract objects. Hence, besides being a mirror of all that there is, for Frege logic was a mirror of a specific realm; the problem for interpreters has been whether Frege considered that realm in the framework of epistemology only, or whether he regarded its objects as having an ontological status. This doctrine, whether in its epistemological version or both its epistemological and its ontological version, was in opposition with the fourth doctrine listed, which was called psychologism.

2.1. The Leibnizian Starting Point: Logic as the Mirror of Reality

Leibniz was the most prominent of the pre-Fregean thinkers who maintained that the terms of our natural language do not correspond to the things of the world in a proper way and that we should therefore construct a new language which mirrors correctly the whole universe. One important characteristic of modern logic was that unlike traditional logic, it proposed a new language—*mathesis universalis*, *lingua characteristica*, begriffsschrift, or whatever it was called by various authors. Modern logicians, primarily Frege, wished to establish a new language that mirrors the world and replaced the grammatical subject-predicate analysis of sentences by the argument-function analysis. Therefore, the term "linguistic turn," as applied to Frege, may lead us astray, if we do not remember that Frege also turned away from language. That is, unlike traditional logicians, he paid little attention to grammatical concepts like those of subject and predicate in his logical studies.

In addition to the dream of ideal language, there was the idea of calculus strongly emphasized by Boole and his followers. It meant the effort to formulate the rules of logical inference explicitly by presenting logical and non-logical vocabulary, formation rules, and transformation rules. Boole stated as follows:

> We might justly assign it as the definitive character of a true Calculus, that it is a method resting upon the employment of symbols, whose laws of combination are known and general, and whose results admit of a consistent interpretation. (Boole 1965, 4)

The nineteenth century saw a breakthrough of the two ideas, even if emphases varied among logicians. Frege stressed that he did not want to put forward, in Leibniz's terms, only a *calculus ratiocinator*, by which he primarily meant the rules of logical inference. He argued that his conceptual notation was to be a *lingua characterica*, which was the term that he used for Leibniz's *lingua characteristica*. That is, his notation was to be a proper language which speaks about all that there is.[10] Frege raised criticism against Boole, because in his view Boole merely focused on developing a Leibnizian calculus in his logical works. However, this was not exactly what Boole himself thought of his project, because he included the idea of logic as a mental or philosophical language in his philosophical remarks on logic (Boole 1958, 11, and Boole 1965, 5).

It has been argued in the literature that since different logicians emphasized different sides of the Leibnizian project, they finally came to advocate conflicting views of the basic nature of logic. It has been claimed that Boole, Peirce, and Schröder, for example, were inclined to stress the importance of developing a calculus, whereas Frege and the early Russell were among those who laid more emphasis on the idea of logic as a universal language. The systematic consequences of the two views have been studied by a number of authors, especially Jean van Heijenoort (1967), Warren D. Goldfarb (1979), and Jaakko Hintikka (1979, 1981a, 1981b). According to these studies, those who stressed the idea of logic as language thought that language speaks about one single world. This was the position to which Frege was committed. He thought that there is one single domain of discourse for all quantifiers, as he assumed that any object can be the value of an individual variable and any function must be defined for all objects. This is what was stated by his principle of completeness (*Grundsatz der Vollständigkeit*) (GGA II, §§56–65). On the other hand, those who supported the view that logic is a calculus were ready to give various interpretations or models for their formal systems. This appears to have been Boole's standpoint. Boole wrote:

> Every system of interpretation which does not affect the truth of the relations supposed, is equally admissible, and it is thus that the same process may, under one scheme of interpretation, represent the solution of a question on the properties of numbers, under another, that of a geometrical problem, and under a third, that of a problem of dynamics or optics. (Boole 1965, 3)

However, it is not clear how this passage ought to be interpreted. It is noteworthy that Boole's statement is not far from what Frege thought. Frege also wished to construct such a language as can be applied to various fields like arithmetic and geometry (BS, 1964, "Vorwort," XII). However, fields of application are not what is meant by the distinction between the one-world and the many-world view. Moreover, Frege wanted to develop both a language and a calculus; if he wanted to develop them as they are understood by contemporary scholars, he could not consistently support both of the implications stressed by those scholars, that is, he could not preach for the one-world view and for the plurality of worlds at the same time.

The twentieth-century perspective has also given more content to the two views. It has been claimed that those who support the idea of logic as language tend to think that they cannot step beyond the limits of language and that this prevents them from developing a proper semantic theory for their language. On the other hand, it has been argued that those who endorse the view of logic as calculus are inclined to think that it is possible to look at a formal system, as it were, from the outside and develop a semantic theory for it. For example, even if Frege introduced his doctrine of senses (*Sinne*) and references (*Bedeutungen*), which is a semantic doctrine, he did not believe that he could propose a proper semantic theory for a formal or a natural language. He repeatedly pointed out that he can only give suggestions and clues concerning his basic semantic concepts and the semantic properties of his conceptual notation.[11]

Frege made the distinction between language and calculus on the basis of his interpretation of Leibniz's project, but he was not conscious of all the implications of the two views of logic which have been detected in the literature. Hence, there are at least three different (though closely connected) stories to be told, as far as the ideas of a universal language and calculus are concerned. There is the story of the content which Leibniz gave to his idea, the story Frege and Boole told about their projects, and the story told from the twentieth-century perspective that tries to capture the far-reaching systematic implications of the two extreme positions.

In Frege's hands, the dream of a universal language was tied to the task of philosophy. Otherwise Frege did not write much about the task of philosophy. In the beginning of his *Begriffsschrift* (1879) he writes that if one task of philosophy is to free the human mind from the power of word by revealing the mistakes that are often almost unavoidably caused by the use of language, then his conceptual notation, which has been constructed for this purpose, will be a useful tool for a philosopher (Frege, BS, 1964, XII–XIII). Frege often complains that natural language leads us astray. However, he nowhere states that it would be the only task of philosophy to clarify language.

There is one story to be told concerning the relations between Kant and Frege which illuminates Frege's position among the opponents and the supporters of metaphysics. In the preface of his *Begriffsschrift*, Frege states that he tries to realize Leibniz's idea of *lingua characterica*. The term was most likely to come from the Leibniz edition by J. E. Erdmann from the years 1839 and 1840, as the word *characterica* is used there instead of the word *characteristica* used by Leibniz (see Haaparanta 1985, 102–117). Adolf Trendelenburg also used the same word in his writing "Über Leibnizens Entwurf einer allgemeinen Charakteristik" (1867). According to Trendelenburg, philosophers ought to construct a Leibnizian universal language, *Begriffsschrift*, by taking Kant's theory of knowledge into account. In his view, Kant's contribution was that he distinguished the conceptual and the empirical component of thought and stressed the importance of studying the conceptual component. Trendelenburg also tells us about Ludwig Benedict Trede, who in his article "Vorschläge zu einer nothwendigen Sprachlehre" in 1811 tried to create a universal language by following Leibniz and Kant. Frege also called his language conceptual notation, which he, it is true, took to be a less successful name for it. He also used the expression "the formula language of pure thought" in the subtitle of his book *Begriffsschrift* and the expression "the intuitive representation of the forms of thought" in his article "Über die wissenschaftliche Berechtigung einer Begriffsschrift" (1882) (Frege, BS, 1964, 113–114). The above-mentioned connections have been noticed and also stressed by a few scholars several years ago (see Sluga 1980; Haaparanta 1985). Even if there were no similarities whatsoever between Trede's notation and Frege's language, we can say that by his reference to Trendelenburg Frege told us something about the philosophical background of his conceptual notation.

On the basis of what has been said, we may argue that Frege's conceptual notation was itself a philosophical position taking. It was not in favor of psychologistic transcendentalism, according to which the necessary conceptual conditions which make knowledge and experience possible are typical of the human mind. Nor was it in favor of transcendental idealism, if we think that a transcendental idealist is one who acknowledges a transcendental subject. We can say, however, that Frege was a transcendentalist in a very weak sense; he tried to write down the forms of thought, which Kant would have called the necessary conditions of knowledge and experience. It is, of course, obvious that Frege's conceptual notation was not a codification of those forms that we find in Kant's table of categories.

The Vienna Circle gave a special treatment to the new logic that Frege had developed. The manifest of the Vienna Circle was directed against metaphysics, and the same spirit can be found in many other writings of the members of the circle. In the manifest, the new logic was described as a neutral system of formulas, a symbolism which is free from the slag of the historical languages, as a tool by means of which it is possible to show that the statements made by metaphysicians and theologians are pseudo-statements, that they express feeling of life, which would be properly expressed by art. The Vienna Circle regarded the close relation with traditional languages as the main problem of metaphysics. They also blamed metaphysics for assuming that thought can know itself without empirical material; that kind of knowing was sought by transcendental philosophy. The Vienna Circle declared that it is not possible to develop metaphysics from "pure thought" (Der Wiener Kreis 1973, 308). They believed that logical analysis overcomes not only scholastic metaphysics but also Kantian and modern apriorism. That position taken by the Vienna Circle meant the rejection of synthetic judgments a priori and hence the rejection of transcendental knowledge.

Hence, if we draw a line from Kant to Frege and further to the Vienna Circle, there is a crucial change in how the relations between being and the pure forms are understood. It is as early as in his *Allgemeine Erkenntnistheorie* (1918) that Schlick raised the question of whether there are any pure forms of thought and answered that thought with its judgments and concepts does not press any form on reality (Schlick 1918, 304–305). For Schlick, that means the repudiation of Kant's philosophy (ibid., 306). In his article "Die Wende der Philosophie" (1930) he argued that the greatest change is due to a new insight concerning the nature of the logical, which was made by Frege, Russell, and particularly Wittgenstein. According to that new understanding, the pure form is merely the form of an expression, but that form cannot be presented (Schlick 1938, 33–34). It is true that Frege did not present the system of signs called conceptual notation, if presenting it had meant giving a semantic theory for the system in a metalanguage. If Frege thought that forms of thought are proper objects of knowledge, that knowledge was for him a kind of immediate recognition. Recognition of the correct forms, the result of which is the creation of conceptual notation, can be called immediate intellectual seeing or intuition. In his late writings in 1924 and 1925, Frege stressed that we see correctly, if natural language does not disturb our intellectual seeing. Moreover, when Frege discussed certain important features of his language, such as the distinctions between the different meanings of "is," which are existence, predication, identity, and class inclusion, he gave lengthy arguments for the distinctions. One of the most central reasons he put forth was that his new language takes care of the difference between individuals and concepts, which is missing both in Aristotelian logic and in Boole's logic, and that the difference is mirrored by the distinction between identity and predication as well as by the distinction between predication and class inclusion. For example, to preserve the distinction between objects and concepts, Frege considered it necessary to realize that the "is" of identity differs from the "is" of predication and, moreover, that this distinction reflects how things really are ("Über Begriff und Gegenstand," 1892, KS, 168). Moreover, the motivation for denying that existence is a first-order predicate came from Kant's thought. Frege also gave a positive contribution by trying to tell what existence is, namely, that it is a second-order concept. We can say that Frege had not only a view of the word "being" but also a view of the forms of being, which are forms of thought, and those forms were meant to be codified as his ideal language.[12]

There was a well-known controversy between logic and metaphysics in the early days of the analytic tradition and the phenomenological movement, to which I already referred. The Vienna Circle declared in 1929 that the new logic, the ideal language developed by Frege, Russell, and Whitehead, frees philosophy from considering the true nature of reality. It was believed that by means of the new formula language, it was possible to show that metaphysical statements are meaningless. It was not thought that the very ideal language would have a metaphysical content. For a logical empiricist, Heidegger's philosophy was an example of the meaninglessness of metaphysics. In 1931 Carnap published his article "Überwindung der Metaphysik durch logische Analyse der Sprache," in which he studied Heidegger's sentences and stated that the sentences of a metaphysician cannot be combined with the ways in which logic and science proceed. In his *Was ist Metaphysik?* (1929) as well as in the afterwords of its later editions, Heidegger discussed the criticism that had been raised against the way he used the word "nichts." According to Heidegger, nothing is the origin of negation, not the other way round. His message was that logic has its origin in the being of *Dasein* (Heidegger 1992, 37) and philosophy can never be measured by means of the standards of the idea of science (ibid., 41). Hence, for Heidegger the origin of the logical concept of being was the being of *Dasein.* There thus seemed to be a sharp contrast between Heidegger, who spoke about the meaning of being and a linguistic philosopher who spoke about the different meanings of the word "is." It was Frege who distinguished the different meanings of "is" in his conceptual notation, and therefore it may seem that Frege was clearly among those who wished to limit the talk about being to the word "is." This is not the case, as Frege was not an opponent of metaphysics. It is more to the point to say that Frege's thought lay somewhere between the philosophy of the Vienna Circle and Heidegger's fundamental ontology.

The view of philosophy held by the Vienna Circle was characterized by the fact that philosophy was taken to be an art of using a tool and the good tool was Frege's, Russell's, and Whitehead's formula language. However, the language lost the metaphysical content that it had for Frege. The pure forms were interpreted as the forms of a system of signs; the system of signs was no more "an intuitive representation of the forms of thought," as Frege wrote. In its manifest the Vienna Circle declared that there are no depths in science but there is surface everywhere (Der Wiener Kreis 1973, 306). In that sense, the circle also wanted philosophy to be like science.

Both Frege and Heidegger were interested in the philosophical basis of logic, Frege mainly in the epistemological basis and Heidegger in the origin of logic in the being of *Dasein*. Both thought that there is something under the "surface." The Vienna Circle thought that philosophy is activity; that was especially emphasized by Schlick in "The Future of Philosophy" (1931). Schlick referred to Wittgenstein, for whom philosophy was not a theory but a certain kind of activity, that is, of clarifying meanings and writing formulas which do the job of clarification (Schlick 1938, 132). It is true, the incentive for that kind of philosophizing was given by Frege, but it would be far from the truth to argue that Frege held that view.

Edmund Husserl touched on the relations between logic and being in several connections, for example, when he distinguished between formal ontology and material ontologies in his *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie I* (1913). It was already in the first volume of the *Logische Untersuchungen* (1900) that he divided logic into two parts according to the two tasks that he believed logic to have. One of the tasks was to give the formal categories of meaning, whereas the other was to put forward the formal categories of objects. Husserl listed the basic concepts of pure logic or analytical categories both in the *Logische Untersuchungen I* (A 244–245/B 243–244) and in the *Ideen I* (Husserl 1950b, 26–32). His categories of meaning include such concepts as belong to the essence of the proposition or apophansis, such as subject and predicate, conjunctive, disjunctive, and hypothetical connections, that is, what we would call logical connectives, and the concepts of concept, proposition, and truth. In addition to the categories of meaning, he gave a list of pure formal objective categories, such as object, property, relation, state of affairs, identity, whole and part, number, and genus and species. Husserl called these categories of objects substrate-categories. In the *Ideen I*, Husserl states that "formal ontology contains the forms of all ontologies . . . and prescribes for material ontologies a formal structure common to them all" (Husserl 1950b, 27; Kersten's translation, 21), and then goes on with treating formal ontology and pure logic as synonymous terms. He also claims that pure truths of meaning can be converted into pure truths of objects (ibid., 28).

In the *Logische Untersuchungen*, Husserl pays attention to the distinction between formal and empirical (or material) concepts, as well as to the distinction between formal or analytic propositions and laws and material propositions and laws. He states that concepts like something, one, object, quality, relation, association, plurality, number, order, ordinal number, whole, part, magnitude, and so on, have a basically different character from concepts like house, tree, color, tone, space, sensation, feeling, and so on, which for their part express genuine contents. It is not clear how we should make the distinction between form and content. What is clear, however, is that Husserl and his contemporaries took the very distinction between form and matter, or form and content, to be essential to logical studies, whether logicians were interested in concepts or in inferences in which the concepts were used.

2.2. The Absence of the Metaphysics of Modalities

Our contemporary modal logic is usually considered as an extension of the two-valued predicate calculus that was developed in the late nineteenth century. However, the roots of our modal theory reach far back to Aristotelian logic, which regarded modal logic as a legitimate branch of logical studies. Interest in modal notions is a new phenomenon among logicians only when it is considered in the framework of the developments of those late nineteenth-century logicians who are honored as the pioneers of modern logic.

In the beginning of the twentieth century, logicians were not willing to discuss modal concepts. They were mainly inspired by the extensionalist program which was preached by Frege, among others, and codified in the *Principia Mathematica*. Modal notions seemed to escape all treatments that are interested only in references. Later in the twentieth century, logicians proposed axiomatic systems for modal logic, which, however, first avoided all systematic semantic considerations. Since the late 1950s, they introduced and developed interpretations for the axioms of modal systems. These interpretations are useful for clarifying which systems of axioms most naturally correspond to our intuitions concerning modal notions and their relations.

Leibniz is an important figure behind our contemporary modal logic. It is also known that he is an important figure behind Frege's logical work. Nonetheless, given that Frege set out to realize what Leibniz had dreamt of, it is surprising that he was reluctant to develop modal logic in the early twentieth century. Even if he started from Leibniz's program in arguing that we must construct a proper language that represents the world, he was not true to Leibniz's view that there could be alternative worlds to which our ideal language would be related. Frege's conceptual notation was meant to represent only one world. As already noted, this doctrine of Frege's is most clearly visible in his requirement that all of the predicates of the language must be defined for all objects. Frege's formula language was thus meant to speak about all that there is, and its quantifiers were meant to range over all individuals. Modal logic, as we understand it nowadays, was thus blocked out in the very beginning.

Frege gave another reason for his unwillingness to discuss the concepts of necessity and possibility within the limits of his logic. The reason was that those concepts do not concern logic at all but that they have to do with the nature of the grounds of our judgments (BS, §4). For Frege, logic is interested in the objective realm of thoughts.[13] Frege regarded the act of judging as a psychological phenomenon, which belongs to the realm of our private minds.[14] Hence, even if Frege severely criticized all efforts to reduce logical laws to psychological laws, he restricted modal notions to the realm of psychology, thus agreeing with psychologists. He did not believe that thoughts are necessary or possible as such, but he insisted that they are necessary or possible for our private minds. Like psychologists, he connected modal concepts with the concept of certainty and took them to modify our acts of thinking, which are units of the subjective realm. Signs for modal concepts did not play any role in his ideal language (for Frege's views in more detail, see chapter 12 in this volume).

3. The Relations between Logic, Epistemology, and Psychology

3.1. Logical Psychologism

In the late nineteenth century, the question of what logic mirrors, if it mirrors something, was mostly discussed in a way that was determined by the debate on the relation between logic and psychology. Contemporary naturalism has its roots in late nineteenth-century psychologism. The word "naturalism" was also used in the late nineteenth century. Like contemporary naturalism, late nineteenth-century naturalism and its version called psychologism had various contents(see Haaparanta 1995, 1999b; Kusch 1995).

In her book *Philosophy of Logics* (1978) Susan Haack distinguishes between strong and weak logical psychologism. According to the strong view, logic describes our thought and may also tell us how we ought to think (Haack 1978, 238). In his book *Husserl and Frege* (1982), J. N. Mohanty describes strong logical psychologism as a doctrine according to which logic is a branch of psychology, the laws of logic describe actual human thought, and psychological study is therefore both sufficient and necessary for studying the foundations of logic (Mohanty 1982, 20). In Haack's terminology, weak logical psychologism is the view that logic determines how we ought to think (Haack 1978, 38). Mohanty, for his part, characterizes the weak version as a thesis that it is necessary but not sufficient to study human thinking processes if we want to clarify the theoretical foundations of logic (Mohanty 1982, 20). Many logicians who are regarded as antipsychologists (Frege, for example) might accept what Haack calls weak logical psychologism. However, they would not say that determining the norms of thought would be the only or the basic task of logic (GGA I, "Einleitung," XV).

Logical psychologism had two different roots in nineteenth-century philosophy. First, there was an interpretation of transcendentalism which regarded the transcendental conditions of experience as the conditions determined by the mental structure of the human race. Second, there was the tradition of empiricism, which attempted to base all knowledge on experience. German, French, and British logical psychologism in the nineteenth century was so complicated a doctrine that many ways of classifying it are possible. It could be a doctrine concerning the basic concepts of logic, the basic laws of traditional logic, or the nature of logical inference. If logical psychologism was understood as a theory that primarily concerns the conceptual tools of logic, it either claimed that such concepts as unity, plurality, negation, and possibility are structural features of the human mind, or it argued that those concepts are abstracted from sense perception. The latter position is linked with the empiricist tradition of the modern times. The former position followed if one interpreted transcendentalism by saying that the transcendental conditions of experience are determined by the structure of our factual minds.

If logical psychologism was a doctrine concerning such laws as the law of excluded middle or the law of noncontradiction, it either maintained that those laws are structural features of the human mind or claimed that those laws have their origin in sense perception. There were a number of philosophers who stressed that the laws of logic have an empirical basis in sense perception, but who did not call themselves psychologists. J. S. Mill, for example, did not want to take that label (Mill 1906, Book II). If a psychologist claimed that the basic laws of logic represent the constant and innate structures of the human mind, he regarded the laws of logic as factual in the sense that he took them to be research objects of the science called psychology. Hence, both the empiricist and the transcendentalist version of psychologism were epistemological theories that tried to reveal the natural origin of logic and to justify certain logical concepts, logical laws, and logically valid inferences by means of the revealed origin.

The foregoing classification contained two basic forms of psychologism. Husserl also hinted at a similar division, when he distinguished between empirical and transcendental psychology as two different bases of psychologism (LU I, A 123/B 123). One of the versions abstracts such laws as the law of noncontradiction from the objects of experience, whereas the other version pushes the structure of the mental realm into the objects of experience.

In his *Philosophie als strenge Wissenschaft* (1910–1911) and in his lecture notes "Logik als Theorie der Erkenntnis" (1910–1911) Husserl characterized naturalism in various ways.[15] He stated that naturalism is a phenomenon consequent on the discovery of nature, which is to say, nature considered as a unity of spatiotemporal being subject to exact laws of nature (PsW, 79). He also remarked that psychology is concerned with "empirical consciousness," with consciousness from the empirical point of view, whereas phenomenology is concerned with "pure consciousness," which is consciousness from the phenomenological point of view (PsW, 91). For Husserl, the phenomenological point of view was the philosophical point of view. Moreover, he continued that any psychologistic theory "naturalizes" pure consciousness (PsW, 92). Naturalizing pure consciousness amounts to identifying it with empirical consciousness. If the realm of pure consciousness had been the realm of norms for Husserl, his criticism would have been that naturalism deduces norms from facts. However, the core of the distinction between pure and empirical consciousness was not at that point.

In his lectures on logic as the theory of knowledge (1910–1911), Husserl distinguished between laws of logic, laws of natural sciences, and norms created by human beings. The distinction corresponds to that made by Frege in his preface to the first volume of the *Grundgesetze*. Husserl admitted that there are major reasons which speak in favor of logical psychologism. However, he explained by means of an analogy that logical psychologism is not true. In his "Göttinger Vorlesungen über Urteilstheorie" in the summer term of 1905 Husserl talked about the analogy between geometry and logic. There he points out that it is common to draw a false analogy; the psychologistic view is that the art of logical reasoning is related to psychology as geodesy is related to geometry or as technical physics is related to theoretical physics (19b). In his lectures in 1910 and 1911, Husserl explained what he thought is the right analogy (20b). Just as geodesy is related to ideal geometry, normative logic is related to logic as a theoretical discipline. Moreover, just as behind geodesy there is a natural science or several of them, likewise behind normative logic there is psychology. In Husserl's view, the norms of logic are inferred from the facts of pure or theoretical logic, not from the facts given by psychology; the facts given by pure logic have to do with the structures of propositions and with the inferential relations between propositions.

Husserl also stated in his lecture notes that naturalistic philosophy is characterized by the fact that it acknowledges only one field of possible knowledge, which is nature (17a). Moreover, he stated that naturalism recognizes only one method of giving foundations for knowledge; it argues that all knowledge is based on experience (17b). In Husserl's view, the essential difference between naturalism and antinaturalism was that naturalism does not acknowledge the ideal realm. Husserl characterized the ideal realm as eternal, self-identical, timeless, spaceless, unmovable, and unchangeable; he did not state that it is something that is expressed by normative propositions. He also remarked that there is no mysticism in such a view (28a, 28b). As we will see in the next section, in his later writings he expressed his view in constructivist terms and stressed the difference between two attitudes more than the difference between the two realms.

3.2. Antipsychologism and the Doctrine of the Third Realm

In the passages quoted, Husserl acknowledged what is called "the third realm" by Frege. The doctrine of the three realms can be found in Lotze. According to Lotze, the being of abstract objects is not like the being of concrete objects. Instead, abstract objects are valid, *geltend*. Lotze took it to be important to distinguish between what is valid and what is (*was gilt* and *was ist*) (Lotze 1874, 16 and 507).

Frege presented a doctrine of three realms, by means of which he expressed his view on the being and the being known of logical categories and of thoughts that are constituted by those categories. In the first volume of his *Grundgesetze der Arithmetik* (1893) and in his article "Der Gedanke" (1918) he made a distinction between the subjective realm of ideas (*Vorstellungen*), the objective realm of actual (*wirklich*) objects, and the realm of objects that do not act on our senses but are objective, that is, the realm of such abstract objects as numbers, truth values, and thoughts (GGA I, XVIII–XXIV; KS, 353). His conceptual notation, which he called the formula language of pure thought, was meant to mirror parts of the third realm, as it was meant to present the structure of thoughts and the inferential relations between thoughts.

It is usually assumed that Frege's acknowledgment of the third realm was a Platonic doctrine. Some interpreters have challenged the received view, but others, most notably Tyler Burge (1992), have given strong arguments for the view that Frege held a Platonic ontology; Burge also emphasizes that Frege did not seek to defend his position, except for showing problems in competing views, and that he did not make any effort whatsoever to develop a sophisticated version of his ontology.

In spite of Burge's carefully documented argumentation, other interpretations remain serious candidates. When Frege discussed his third realm in his "Der Gedanke," he remarked that he must use metaphorical language. In other words, such expressions as "the content of consciousness" and "grasping the thought" must not be understood literally (KS 359, n. 6). As Frege expressed his worry about the fact that natural language leads us astray as early as in the preface of his *Begriffsschrift*, the interpretation that Frege did not take numbers or thoughts to have being in the proper sense of the word "being" is at least worth considering. Frege did think that the objects of the third realm are objective, hence, independent of subjective minds. That is not yet an ontological position. On Frege's view, thoughts and their constitutive logical categories are denizens of the third realm, but their being is not like the being of the denizens of the objective and actual realm. Thomas Seebohm has argued that Frege presented a transcendental argument to the effect that the existence of mathematical objects and logical categories is a necessary condition of the meaningfulness of mathematical and logical practice (Seebohm 1989, 348). If that argument holds, Frege's acknowledgment of the third realm would have ensued from his epistemological views.

Husserl argued that we must acknowledge an ideal realm of abstract objects to avoid psychologism. He pointed out that there is an unbridgeable difference between the sciences of the real and the sciences of the ideal, as the former are empirical, while the latter are a priori. Husserl realized that if we acknowledge the ideal realm, we must face an epistemological problem concerning our access to this realm. Most of Husserl's logical studies after his *Logische Untersuchungen* are an effort to answer this question by means of phenomenology. In his last logical works, titled *Formale und transzendentale Logik* (1929) and *Erfahrung und Urteil* (1939), which was published posthumously, he sought to show that we have an access to the denizens of the ideal realm, because we have set the structure of transcendental consciousness to those denizens, hence, we have maker's knowledge of that realm.

Husserl did not think that we could be mistaken about what the correct logical categories are. His problem was how to give a justification for what he regarded as our true beliefs concerning those categories. In his sixth logical investigation (LU II, 1901, 1921) Husserl studies the components of meaning which determine the form of a proposition and calls them categorial meaning-forms (*Bedeutungsformen*). In his view, those forms are expressed in natural language in several ways, for example, by definite and indefinite articles, numerals, and by such words as "some," "many," "few," "is," "and," "if—then," and "every" (LU II, A 601/B_2 129; LU II, A 611/B_2 139). Husserl asked what the origin of logical forms is, when nothing in the realm of real objects seems to correspond to them (LU II, A 611/B_2 139). He took it to be a problem how the logical words originally get their meaningfulness, hence, what kind of activity of a subject is required so that the logical words become meaningful. In his last works on logic, Husserl sought to show that that activity is precisely the activity of transcendental consciousness.

Kant interpreted logical categories as the pure concepts of understanding, which correspond to certain types of judgments and which give form to the objects of experience. In his *Begriffsschrift*, Frege, for his part, introduced eight signs as the basic signs of his formula language of pure thought; those signs expressed the basic logical categories and made it possible for Frege to present most types of judgments listed in Kant's table. As was noted, in Frege's doctrine of the three realms the logical categories were regarded as constitutive for the denizens of the third realm called thoughts.

3.3. On the Possibility of Logical Knowledge

3.3.1. What Is Logical Knowledge?

Kant is famous for his effort to answer the so-called transcendental questions, such as "How is pure mathematics possible?", "How is pure natural science possible?", and "How is metaphysics as science possible?" This type of questions have two readings. One either wants to know whether x is possible and wishes to have a justification for its possibility, or one assumes that x is possible and tries to find out the conditions of its possibility.[16]

If one raises the question concerning the possibility of logical knowledge, one may think of two questions, first, whether logical knowledge is possible at all, and second, if it is, under what conditions it is possible. This section is a short study of a few late nineteenth-century and early twentieth-century logicians' and philosophers' views of that possibility. Frege's and Husserl's views will again be in focus. By logical knowledge, one may mean knowledge which is reached by means of logical inference, hence, knowledge based only on logical truths. For example,

$$\text{knowing that } p \,\&\, q \rightarrow p$$

would be an example of logical knowledge. By logical knowledge, one may also mean knowledge concerning the basic concepts of logic, the logical forms of propositions, the basic laws of logic, or the rules of logical inference. For example, statements like

> Existence is a logical concept.
>
> The logical form of the sentence "Man is an animal" is "$\forall x(F(x) \rightarrow G(x))$".
>
> The law of noncontradiction holds.

would express logical knowledge in the intended meanings.

In his *Grundlagen der Arithmetik* (1884, §3) Frege stated that the distinctions between a priori and a posteriori, synthetic and analytic, concern not the content of the judgment but the justification for making the judgment. He excluded the naturalistic interpretation of his claim and stated that by his distinctions he intends to refer to the ultimate ground on which rests the justification for holding a proposition to be true. He continued that the problem becomes that of finding the proof of the proposition. By his characterizations of analytic and synthetic truths and truths a priori and a posteriori, he expressed the view that the justification of analytic truths a priori comes from general logical laws and definitions. In his view, logical laws neither need nor admit of justification. However, the question remains what Frege would have named as the source of knowledge if he had thought that we can know the structure of the ideal logical language in the proper sense of knowing. Did he think that we *know* that existence is a logical concept? If he thought that way, what would he have labeled as the source of knowledge, hence, what would have been a justification for such a claim?

As already stated in section 3.2, Husserl studied the components of meaning which determine the form of a proposition and called them categorial meaning-forms (*Bedeutungsformen*). He took it to be a problem how the logical words originally get their meaningfulness, hence, what kind of activity of a subject is required so that the logical words become meaningful. In Husserl's thought, the questions of origin were linked with the questions of justification.

3.3.2. Can We Have Logical Knowledge?

Emil Lask on Kant's View Kant thought that categories, hence, logical concepts, have their origin in the logical forms of propositions. However, he took the list of the logical forms of propositions for granted. If Kant's transcendental deduction was a justification of certain logical concepts, the idea of that justification was to show the role of those concepts in cognition and experience; it was to show how the pure concepts of understanding contribute to making objects of knowledge possible and how they are linked with the forms of intuition. Expressed in contemporary terminology, Kant sought to give us the epistemological foundation of logic by showing how the pure forms of thought are applied to sensuous experience. Moreover, by that project Kant also tried to show us how logical knowledge in general is possible. He argued that logical concepts are hidden in objects of experience, they have their origin in those objects, and we can have knowledge of them precisely via their link to what is given in intuition.

Emil Lask, a student of Rickert, whose thought influenced Heidegger's early philosophy, praised Kant's Copernican revolution in his work on logic and the doctrine of categories. According to his writing published in 1911, Kant had shown that certain questions concerning objects belong to logic, hence not to metaphysics (LP, 31).[17] However, Lask argued that Kant's critique of knowledge could not touch on the questions concerning logic or the logical forms of objects in a proper manner (LP, 260–262). In his view, that followed because Kant was committed to a two-world doctrine in which a distinction was made between the world of sensory objects and the transcendent world. Lask argued that as Kant neither regarded logic as sensory nor took it to be metaphysical, he made it homeless (*heimatlos*; LP, 263).

We may disagree on Lask's two-world interpretation of Kant's thought. However, it is interesting to find out how Lask solved the problem concerning the homelessness of logic which he thought to have found in Kant's philosophy. His starting point was to give up the two-world doctrine and replace it by an epistemological doctrine concerning the concept of objectivity. That doctrine came from Lotze. As was noted, according to Lotze the being of abstract objects is not like the being of concrete objects. Lotze took them to be valid, and he considered it to be important to distinguish between what is valid and what is (Lotze 1874, 16 and 507). Lask supported that kind of division between two worlds (LP, 6), but he did not consider it an ontological distinction. Instead, for him that was a distinction between two different attitudes or points of view, which we can take toward our sensory experience (LP, 48–49, and 88–91). Lask thought that the logical attitude considers psychological, physical, and cultural beings in a way that differs from the attitude of everyday experience and scientific activity; it is interested in what is valid for those beings.

On Frege's View In his "Über die wissenschaftliche Berechtigung einer Begriffsschrift" (1882) Frege writes: "a perspicuous representation of the forms of thought (eine anschauliche Darstellung der Denkformen) has ... significance extending beyond mathematics. May philosophers, then, give some attention to the matter!" (BS, 1964, 114). The forms of thought Frege talked about were not meant to be the forms which the human mind happens to have. However, Frege thought that we (or he) can have an access to those forms and they can be written down as a language, as a conceptual notation (begriffsschrift), as he thought to have done in his book titled *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. These forms were not tied to human psychology, but they were pure, hence, not naturalistically characterizable forms. If Frege thought that logical forms can be known in the proper sense of knowing, he must have meant by "knowing" some kind of immediate recognition of the presence of the forms of thought. That recognition, the result of which is Frege's conceptual notation, could be called intuition, in the sense of immediate intellectual seeing.

How is this intellectual seeing possible at all? In view of Frege's conceptual notation, which is meant to be a genuine language that speaks about the world and carries fixed meanings, intellectual seeing presupposes grasping the correct structure of thoughts. Frege did not think that meanings could be given syntactically, hence, for him, giving meanings to logical constants did not amount to giving inferential rules, say, rules of introduction and elimination. For him, meanings of logical function names were found by means of grasping thoughts and by analyzing them. Still, Frege assumed that meanings are present in the syntax of the conceptual notation and there is no way of giving a semantic theory for that notation. He thought that our knowledge concerning the structure of the ideal logical language, hence, the basic logical concepts and the logical forms of propositions, and concerning the basic laws of logic carries its own justification, which has to do with "immediate seeing," which is not disturbed by sensory data. However, even if Frege did not seek to present any theory of logical knowledge, that did not mean that his ideal language would not have been motivated by epistemological considerations.[18]

On Husserl's View Husserl's doctrine of categorial perception in the *Logische Untersuchungen* was meant to be a solution to the problem concerning the origin of logical knowledge. Husserl introduced a new concept of perception that was not sensuous perception. In his view, categorial meanings are originally related to objects of sense perception but in a peculiar manner; logical forms are in the objects of sensuous acts but hidden in them as it were. In categorial perception, which was the term Husserl used, the subject sees the sensuous object differently; he or she perceives the object via logical forms, hence, the object is for him or her in these forms (LU II, A 615/B_2 143). Sensuous objects are objects of sensuous acts, whereas ideal objects are objects which arise in that kind of "seeing differently" (LU II, A 617/B_2 146). In Husserl's view, such acts as the act of conjunction, disjunction, and generalization need sensuous acts which are their foundation (LU II, A 618/B_2 146). In the later edition of the sixth logical investigation in 1921, Husserl remarked that these acts that are not founding acts are in relation to what appears in the sensuous founding acts (LU II, B_2 146).[19] Lask and Husserl thus shared the idea that the philosophical nature of logic must be studied by studying the attitudes or the points of view which the subject of knowledge has to the objects of knowledge.

Hence, in Husserl's view the origin of logical concepts is in sense perception; logical forms can become ideal objects studied by the science called logic, because there is a subject who sees the objects of perception in an explicating manner (LU II, A 623–625/B_2 151–153). Logical concepts, like the concepts of whole and part, are possibilities in objects which become articulated in categorial acts (LU II, A 627/B_2 155). Logical forms are not in themselves but the subject makes them exist. Husserl thus found the origin of logical concepts in the activity of the subject. For Husserl, the forms of thought have been set into sensuous objects and as objective they can be known by us. To know them is to construct new categorial objects, and that constructing is categorial intuition. On this reading of Husserl's text, logical knowledge is possible because it is knowledge concerning our own constructions. Frege thought that we cannot take distance from logical categories, we can only write them down when we see them correctly; we cannot present an epistemological theory for them. Nonetheless, Frege gave us several epistemological arguments which aimed at supporting his choice of a certain kind of logical language. Unlike Frege, Husserl thought that logical categories and even logical laws need and can be given justification, which means giving an epistemological theory for logic.[20]

A somewhat surprising conclusion can be drawn if we pay attention to the connection between the views of the possibility of logical knowledge just discussed and of the nature of philosophy. The philosophers of the Vienna Circle thought that Frege's, Russell's, and Whitehead's logic was a neutral system of formulas and a useful tool for clarifying thoughts, hence, philosophy was for them a certain kind of activity, namely, the activity of clarifying thoughts by means of the new tool. They did not suggest that philosophers ought to present theories of anything, not even of logical knowledge. Later, it has been typical of the analytic tradition to put forward philosophical, formal, and semiformal theories of various kinds, including theories of logic, or logics, and of natural language. From that perspective, Husserl's way of thinking of the possibility of logical knowledge and his search for a theory of that kind of knowledge is a more natural background for the analytic tradition than Frege's approach.

4. Discovery, Justification, and Intuition

4.1. The Rejection of Intuition

The problem of justification became a central theme in the philosophy of logic during the first decades of the twentieth century. The role of intuition as a justifier was discussed by logicians and philosophers. From what has been said, it seems that Husserl opposed reference to intuition in cases where Frege was ready to rely on intuitive knowledge. If we argue that way, we suggest that Husserl's demand for justification goes further than that of Frege's. It is true Husserl thought that even if propositions that are taken to be basic in a formal system are not in need of justification in terms of logical inference, they need another kind of justification, namely, a philosophical justification. Unlike Frege, Husserl thought that logical categories and logical laws need and can be given philosophical justification in the sense of giving a philosophical or an epistemological theory for logic. In section 4.2.1, I trace Husserl's view back to its Kantian origins.

In a postumous work published in 1991, J. Alberto Coffa considers the semantic tradition from Bolzano to Carnap, hence from the early nineteenth century until the early twentieth century. According to Coffa, the semantic tradition reacts against Kant's philosophy. He claims that that tradition tried to get rid of all references to intuition, which it took to be Kant's great problem. Coffa points out that the semantic tradition can be defined by means of its problem, its enemy, its goal, and its strategy. According to Coffa, its problem was a priori, its enemy was pure intuition, on which Kant relied when he studied the possibility of mathematics, its aim was to develop a concept of a priori in which pure intuition played no role, and its strategy was to base that theory on the development of semantics (Coffa 1991, 22). Coffa also argues that in geometry it is particularly necessary to refer to constructions which are seen immediately but that even calculus, which was the strongest branch of eighteenth-century mathematics, had the same practice (ibid., 23). Coffa remarks that by the end of the nineteenth century Bolzano, Helmholz, Frege, Dedekind, and many others helped settle that Kant was not right, that concepts without intuition were not empty (ibid., 140). The pioneers of logic at the end of the nineteenth century stressed that in the field of logic one is not allowed to refer to intuition; each inferential step must be written down. Particularly, Frege's conceptual notation was meant to be a tool by means of which each step in the process of inference can be written down exactly without any resort to intuition.

However, even if Frege and Peirce, among others and maybe most prominently, were creating a new logic and Frege tried to carry out a program which aimed at reducing arithmetic to logic, they did not, and they did not even want to, get rid of intuition altogether. Of course the very concept of intuition was problematic for them. If we look at the pages of Frege's *Begriffsschrift*, we notice that he appeals to what we would nowadays call our pattern recognition abilities both in his analysis of sentences and in his ways of presenting inferences. Peirce laid even more emphasis on the role of intuition. For example, in 1898 he praised Kant for understanding the role of constructions or diagrams in mathematical inference. He wrote that mathematical inference proceeds by means of observation and experiment and that the necessary nature of this inference is merely caused by the fact that a mathematician observes and tests a diagram which is his own creation ("The Logic of Mathematics in Relation to Education," 1898, CP, 3.560). In 1896, Peirce noted that logic has to do with observing facts concerning mental constructions (NE 4, 267). He very often stressed the value of figures in inference and states in 1902 that all knowledge has its origin in observation (NE 4, 47–48). Of course these pioneers of modern logic did not assume that a logician is able to see all the consequences of given premises, and they did not give a logician a permission to refer to seeing the conclusion of given premises immediately. Nevertheless, they thought that when taking the shortest steps in an inferential process, a logician does something that can be naturally called perceiving or seeing.

4.2. Husserl's Problem of Justification and Frege's and Peirce's Discoveries

Even if the idea of axiomatic science in logic is not discussed in this section, the methods of logical discovery and justification deserve attention. Frege claimed that all great scientific improvements of modern times have their origin in the improvement of method. He wrote:

> I would console myself on this point with the realization that a development of method, too, furthers science. Bacon, after all, thought it better to invent a means by which everything could easily be discovered than to discover particular truths, and all great steps of scientific progress in recent times have had their origin in an improvement of method. (Frege, BS, 1964, XI; Frege 1972, 105)

The method Frege proposed for science was his begriffsschrift, the new logic, but there was even a deeper truth in his statement. A better method was also needed if one wished to improve logic.

In his paper "Explanation of Curiosity the First" (1908) Peirce described Euclid's procedure in proving theorems. Euclid first presented his theorem in general terms and then translated it into singular terms. Peirce paid attention to the fact that the generality of the statement was not lost by that move. The next step was construction, which was followed by demonstration. Finally, the *ergo*-sentence repeated the original general proposition. Peirce laid much emphasis on the distinction between corollarial and theorematic reasoning in geometry. He took an argument to be corollarial if no auxiliary construction was needed. For Peirce, construction was "the principal theoric step" of the demonstration (CP, 4.616). Peirce also stressed that it is the observation of diagrams that is essential to all reasoning and that even if no auxiliary constructions are made, there is always the step from a general to a singular statement in deductive reasoning; that means introducing a kind of diagram to reasoning.

Peirce's methodological interests are well known. For example, in 1882 he stated in his "Introductory Lecture on the Study of Logic": "This is the age of methods; and the university which is to be the exponent of the living condition of the human mind, must be the university of methods" (W 4, 379). Moreover, in his "Introductory Lecture on Logic" (1883) he made an interesting remark on methodology. He wrote:

> But modern logicians generally, particularly in Germany, do not regard Logic as an art but as a science. They do not conceive the logician as occupied in the study of methods of research, but only as describing what they call the *normative laws of thought*, or the essential maxims of all thinking. Now I have not a high respect for the Germans as logicians. I think them very unclear and obtuse. But I must admit that there is much to be said in favor of distinguishing

> Logic from Methodology. . . . Let us say then that Logic is not the *art of method* but the science which analyzes method. (W 4, 509–510)

As Peirce thus regarded logic as science, it is no surprise that he was also interested in the methodological commitments and choices of the one who works in the science of logic.

There is an interesting history of method from Kant to Frege and Husserl. What was especially important is that all the way the task is twofold. On the one hand, Kant considered transcendental forms, that is, logical concepts, to be our method or tools for reaching the phenomenal world, as he considered them to be our tools for constructing that world. On the other hand, he regarded it as necessary to have a proper method, which is transcendental analysis, for knowing those very tools. Frege's task was also twofold. Frege set out to find a new method for science, which would be his begriffsschrift, but he also needed a new method of discovering that very method.

In Husserl's philosophy the method of finding the method came to be a method of knowing the ideal world. That happened because Husserl considered the logical tools to have being as structures of that world. According to Husserl, logic tries to claim something about the structure of the realm of ideal objects, which is strange for us in the sense that it is independent of our subjective mental realms. Husserl's question brings us back to the question of method, as Husserl assumed that we have knowledge of the ideal realm only if there is a reliable route from our subjective minds to the objective realm, that is, only if we have proper tools for reaching that realm. Therefore, for him the foundational task was to know and describe these very tools.

4.2.1. Husserl and the Justification of Logic

Husserl's question "How is logic as science possible?" amounted to the question "What were the methods of discovery and justification that justified modern logic as science?" Husserl also proposed this problem for those who are interested in the foundations of logic. He compared the activities of a practicing artist with those of a scientist. He argued that both of them are in an equally bad shape if we think of how conscious they are of the principles of their creation or their evaluation. Husserl even claimed that mathematics has no special position on this issue. A mathematician is often unable to inform us of his steps of discovery or to give us a proper theoretical evaluation, that is, a justification, of his results (LU I, A 9–10/B 9–10). Husserl proposed that all discovery and testing rest on regularities of form and that regularities of form also make the theory of science, that is, logic, possible (LU I, A 22/B 22).

Husserl's thought lends itself easily to the framework of the philosophical tradition introduced by Kant. His main works in the field of logic bear Kantian labels in their very titles. His trilogy of logic consisted of the book titled *Logische Untersuchungen* I–II (1900–1901), the first volume of which he calls *Prolegomena zur reinen Logik, Formale und transzendentale Logik* (1929), and *Erfahrung und Urteil* (1939), which was posthumously completed and published by Ludwig Landgrebe. Even if Husserl attached his philosophy to Kantian themes, he was convinced that he had to raise heavy criticism against Kant's ideas. He blamed Kant for having failed to achieve a "pure" theory of knowledge, which would be free from all naturalistic elements, such as psychological and anthropological assumptions. No more was he satisfied with neo-Kantians' developments, which he called transcendental psychology (LU I, A 92–97/B 92–97).[21] He admitted, though, that Kant's philosophy also had features that go beyond psychologism (LU I, A 94/B 94, note).

In his early writings, Husserl seemed to speak in favor of psychologism, for example, in his book *Philosophie der Arithmetik* (1891), which Frege, the devoted antipsychologist, heavily attacked in 1894 ("Rezension von: E. Husserl, Philosophie der Arithmetik, Erster Band, Leipzig, 1891," KS, 179–192). Some scholars, for example Mohanty (1982), have disputed that Husserl was a psychologist in the sense that Frege gave to the term. Mohanty stresses that Frege's criticism led Husserl to revise some parts of his theory of number and it may have made him pay more attention to distinguishing between act, content, and object. However, Mohanty points out that it could not lead Husserl to reject such a version of psychologism which Frege attacked simply because Husserl never subscribed to that version (Mohanty 1982, 22–26). However that may be, it was at the very end of the nineteenth century that Husserl clearly joined the antipsychologistic camp, which his *Logische Untersuchungen* testified. It may be noted that in that work he also pointed out that he does not want to reject everything that he has done in his *Philosophie der Arithmetik* (LU II, B_1 283, note).

In the *Logische Untersuchungen*, the main starting points for Husserl were Bolzano, Lotze, and Brentano, to whom Husserl paid homage in those two logical works (LU I, A 219–227/B 219–227, and LU II, A 344–350/B_1 364–370). Bolzano (1837) had introduced *Sätze an sich* and *Vorstellungen an sich*, which he regarded as neither existing in space and time nor depending on our mental acts (Bolzano 1929, §19). Hence, Bolzano distinguished the proposition itself from our thinking of it and acknowledged a specific realm of ideal objects, for which he did not admit proper existence, though. As was noted, Lotze, for his part, considered being and validity to be two senses of actuality (*Wirklichkeit*) and distinguished between the being of concrete things and the validity of abstract objects. For him, validity was a way of being independent of subjective mental acts (Lotze 1874, 507). Even if Brentano was not a defender of abstract entities, he distinguished between mental acts and their objects, which have intentional inexistence in those acts but need not have any real existence (Brentano 1924, 124–125). Husserl was influenced by Brentano already from the middle of the 1880s, when he was Brentano's student in Vienna.[21]

As was noted in section 3.2, Husserl approved of those ideas and made a distinction between the real and the ideal. He stated:

> There is an essential, quite unbridgeable difference between sciences of the ideal and sciences of the real. The former are a priori, the latter

> empirical. The former set forth ideal general laws grounded with intuitive certainty in certain general concepts; the latter establish real general laws, relating to a sphere of fact, with probabilities into which we have insight. (LU I, A 178/B 178; LI I, 185)

Husserl observed that once the distinction between the ideal and the real realm is acknowledged, we quite naturally come to realize one crucial problem. This problem constituted an important part of Husserl's criticism against Kant. In 1929 Husserl maintained that because Kant did not make the distinction between the ideal and the real, he failed to ask one important question. Because Kant did not assume any world of ideal objects of thought, he could not ask how we can have an access to these objects (FTL, 233–235). In the *Formale und transzendentale Logik*, Husserl was explicit in stressing the importance of Kant's theories concerning the Humean problem, which include his doctrine of transcendental synthesis and of transcendental abilities in general. Husserl praised Kant's questions concerning our knowledge and its presuppositions. However, he blamed Kant for not asking transcendental questions about formal logic (FTL, 228–230). Kant took Aristotelian logic to be a complete system, which needs no major corrections. All we can do for what he called general logic was to make it more elegant; the proper task of that logic, which is to expose and prove the formal rules of all thought, had already been accomplished, in Kant's view (KRV, B viii–ix). Kant asked how pure mathematics is possible, how pure natural science is possible, and how metaphysics as natural disposition and as science is possible (KRV, B 20–22), but he did *not* ask how logic as science is possible. Husserl believed that if Kant had distinguished between the ideal and the real realm, it would have occurred to him to ask such an epistemological question.

Husserl concluded that both Hume and Kant realized the transcendental problem of the constitution of what he called the real realm. He thought that they failed to see the corresponding problem concerning the constitution of the ideal objects, such as the judgments and the categories which belong to the sphere of reason and which logic is interested in. In other words, Kant did not make his analytic a priori a problem (FTL, 229–230). Husserl's question in his logical works can thus be formulated in three ways: (1) How can we have knowledge of the realm of ideal objects? (2) How can we rely on what logic claims? (3) How can we justify the analytic truths a priori? These formulations have close connections. The ideal realm consists of abstract objects like numbers and thoughts, and it is precisely logic that tries to say something about the structure of thoughts and about the inferential links between thoughts. Therefore, because Kant did not ask how we can know anything about the ideal realm, he did not ask how logic as science is possible, either. Moreover, since logical laws are analytic a priori, Husserl asked how we can rely on the analytic a priori claims which logic offers to us.[22]

Husserl thus blamed Kant for not being able to ask how we can have knowledge of the ideal realm. We could certainly defend Kant by the argument that because he did not postulate any such problematic realm as the realm of abstract objects, he did not need to face such epistemological problems as Husserl. We may also say that even if Kant did not ask Husserl's question, his *Kritik der reinen Vernunft* served as an answer to that question. However, the point in Husserl's argument might be construed as the statement that we cannot know anything that is not made objective, hence that the proper definition of the concept of knowledge implies that the object of knowledge is distinguished from the knowing subject. On this construal of his argument, Husserl required that logical concepts and laws are something that can be known in the proper sense of knowledge. If this is what he meant, the point of his criticism was that Kant did not set the conceptual tools of logic outside consciousness to study those tools.[23]

Husserl thus asked the question which Kant did not ask and tried to do what Kant did not do, namely, lay the epistemological foundations of logic. But what was actually the philosophical incentive of the question concerning how logic as science is possible? From Galilei and Descartes to Kant, philosophers had sought for a firm foundation for modern natural science, for mathematics and even for metaphysics. If we believe that the history of logic can be reconstructed as a Kuhnian science, hence, that the question of foundations arises in logic when the received framework is threatened, we quite naturally see the nineteenth century as a revolutionary period in logic. Aristotelian logic was losing ground in those days, and new formal developments arose. What this period needed, then, was an epistemological justification for either the old logic or for those new suggestions. Hence, on this construal, Husserl's question was necessitated by the new developments of logic in the nineteenth century. Husserl remarked: "how could such a logic [scientific logic] become possible while the themes belonging to it originally remained confused?" (FTL, 158; Husserl 1969, 178).

The foundational crisis was not the most perspicuous reason for the question concerning the possibility of logic as science. The question arose as a natural consequence of the various confrontations within logic and philosophy of logic in the nineteenth century. As we saw in Husserl's case, it arose from a philosophical position that postulated a specific realm of abstract objects like thoughts which logic speaks about. If that kind of realm is assumed and acknowledged, it is quite natural to ask how we can have knowledge of it, that is, how we can rely on logic which is supposed to speak about it.

But why does anyone want to assume such an objective realm? I already suggested one answer that had to do with the proper concept of knowledge. Other guesses can also be made. Husserl's argumentation suggests that historically the objectivity of the field of interest of logic was probably necessitated by a proper criticism against a psychological or anthropological interpretation of Kant's transcendentalism, which was represented by such logicians as Jakob Fries and Benno Erdmann, for example. Fries thought that logical concepts must be understood as the ways in which the human species organizes experience, and the logical laws must be construed as anthropological laws.[24]

On Husserl's judgment, the philosophy of logic of his own day was strongly anthropologistic; he even argued that it was rare to find a thinker who would be free from the influence of that doctrine (LU I, A 116/B 116). In addition to the empiricist tradition, psychologism in logic had a natural connection with Kant's transcendentalism, for the transcendental structure of human thought was easily construed as a psychological structure, which is typical of the human race. If one wanted to save transcendental logic from that kind of reading, one had better regard the transcendental structure as the structure of some objective realm.[25]

4.2.2. The Role of Judgments in Frege's and Peirce's Logical Discoveries

Frege and Peirce discovered quantification theory independently of each other. They both introduced a new formula language in which arguments or indices were distinguished from functions or relative terms. In his paper "Über den Zweck der Begriffsschrift" (1883) Frege remarked:

> In fact, it is one of the most important differences between my way of thinking and the Boolean way—and indeed I can add the Aristotelian way—that I do not proceed from concepts but from judgements. (BS 1964, 101)

That Frege opposed Aristotle and Boole has been noticed by all interpreters, but it was about 30 years ago that Frege's way of thinking was taken under more extensive historical consideration. Interpreters such as Hans Sluga (1980) linked Frege's view with Kant's idea that judgments have priority over their constitutive concepts. Kant was also one of Peirce's philosophical heroes. Murray Murphey (1961) already noted that Peirce's logical discovery brought him closer to Kant, as Peirce distinguished between indices and relative terms, hence, as it were, wrote down Kant's distinction between intuitions and concepts.

In his paper "Booles rechnende Logik und die Begriffsschrift" (1880/81), Frege clarified the difference between his conceptual notation and Boolean logic. He stated that the real difference is that in logic he avoids a division into two parts, of which the first is dedicated to the relation of concepts, that is, to primary propositions, and the second to the relation of judgments, that is, to secondary propositions, by construing judgments as prior to concept formation (ibid., 14 and 52). He continued that unlike Boole, he reduces his primary propositions to the secondary ones, which comes up in that he construes the subordination of two concepts as a hypothetical judgment (ibid., 17–18). This result came out when Frege broke up the judgment which contained subordination, which is a relation between two concepts. Before Frege was able to do this, he had to realize the distinction between individuals and concepts. This is what he also emphasized. In the article he remarked that his view does justice to that distinction. In Frege's view, the problem with Boole's notation lay in that Boole's letters never meant individuals but always extensions of concepts. The distinction between individuals and concepts, or more generally functions, hence, between proper names and function names, was a crucial part in Frege's discovery. It seems on the basis of Frege's remarks in the *Grundlagen* that even if Frege criticized Kant's concept of intuition, he viewed the distinction between intuitions and concepts as a precursor of his own distinction (GLA, §27, n.). The same methodological change from the Boolean method to the analysis of judgments was essential to Peirce's discovery. I already mentioned that Kant was an important figure behind Peirce's philosophy. As Murphey remarked, it was the manner in which Kant discovered his categories that interested Peirce most of all (Murphey 1961, 33).

In the 1870s, Peirce discovered his logic of relatives, which was inspired by De Morgan's ideas and Boole's algebra of logic. Peirce's articles titled "The Logic of Relatives" (1883) and "On the Algebra of Logic: A Contribution to the Philosophy of Notation" (1885) contained the first presentation of his quantification theory, which he himself called his general algebra of logic and which, as he wrote, he developed on the basis of O. H. Mitchell's, his student's, ideas (CP, 3.363 and 3.393). The first important change from Boole's logical algebra was that Peirce added indices to relations. Indices referred directly to individuals. Second, he introduced the quantifiers "some" and "every."

When he introduced his two improvements of logic, Peirce referred to Mitchell's article "On a New Algebra of Logic" (1883). He expressed his indebtedness to Mitchell regarding both indices and quantifiers. However, when he described Mitchell's way of using indices, he deviated from what Mitchell said. Peirce interpreted Mitchell's formula "F_1" as "the proposition F is true of every object in the universe" and formula "F_u" as "the proposition F is true of some object in the universe." For Mitchell, the symbol F was any logical polynomial involving class terms and their negatives. He did not take it to be a proposition, but rather called it a predicate or a description of every or some part of the universe (Mitchell 1883, 75 and 96). Moreover, Peirce used the concept of individual, which Mitchell did not use. Otherwise, it is true that Mitchell had both indices and quantifiers, as Peirce declared.

Mitchell supported the view that objects of thought, in which logic is interested, are either class terms or propositions, but that every proposition expresses a relation among class terms (Mitchell 1883, 73). Because Mitchell thought that, basically, every proposition expresses a relation among class terms, he relied on the Boolean method, which started from concepts and came up with propositions by combining concepts. It is precisely this way of thinking which Frege attacked, as we noted. Hence, even if Mitchell did suggest indices and quantifiers, the new logical language cannot be encountered in his treatment. Peirce's contribution was to take propositions as the starting point of analysis and generate a distinction between relative terms and the names of individuals.

In his article "On a New List of Categories" (1867), which was meant to improve Kant's doctrine of categories, Peirce relied on the subject-predicate form of propositions and assumed that in the aggregate of a subject and a predicate the subject represented what he calls substance (CP, 1.547 and 1.548).

For Peirce, substance was the present in general, hence not an individual. It is not until the 1880s that individuals in the sense of Kant's intuitions appeared in Peirce's logical notation. These observations suggest that between Peirce's "New List" (1867) and his discovery of the new notation (1883, 1885) there was a methodological change, which contributed to his logical discovery. Hence, the decisive insight both for Frege and Peirce was that a judgment is not an aggregate of terms that represent concepts or classes but that its elements have different kinds of roles in their contexts. Two of those basic roles are that of representing relations and that of denoting individuals.

5. Origins of Twentieth-Century Semantics: Frege's Distinction between *Sinne* and *Bedeutungen*

Even if Frege did not have any semantic theory, he expressed views of semantic concepts and had considerations in his works that can be called semantic. For Frege, the *Sinne*, senses, of sentences are thoughts and the *Bedeutungen*, references, of sentences are truth values, the True and the False. Sentences are compounded out of proper names, which refer to objects, and function names, which refer to functions. The *Sinne* of function names are simply parts of thoughts.[26] But what are the *Sinne* expressed by proper names? In "Über Sinn und Bedeutung" (1892), Frege remarked that the sense of a proper name is a way the object to which this expression refers is presented, or a way of "looking at" this object. Furthermore, he stated that the sense expressed by a proper name belongs to the object to which the proper name refers. In other words, for Frege, senses were not primarily senses of names but senses of references. Hence, it is more advisable to speak about senses expressed by names than senses of names. Frege also gave examples of senses, like "the Evening Star" and "the Morning Star" as senses of Venus, and "the teacher of Alexander the Great" and "the pupil of Plato" as senses of Aristotle ("Über Sinn und Bedeutung," KS, 144).

Nonetheless, Frege admitted that we speak meaningfully about entities which do not exist. In his view, a sentence lacks only a truth value—but not a sense—if it contains a name that has no reference ("Über Sinn und Bedeutung," KS, 148). Russell adopted a critical standpoint against this idea, according to which an expression can have a sense although it lacked a reference. In his article "On Denoting" (1905) he argued that a sentence like "The present King of France is bald" should be construed as the sentence "One and only one being has the property of being the present King of France, and that being is bald." The property of being the present king of France does not belong to any being, and therefore the sentence is false. Moreover, according to Russell, the sentence "The present King of France is not bald" is false if it means that there is an entity which is now king of France and is not bald. Russell, however, suggested another analysis for the latter sentence which says that it is false that there is an entity that is now king of France and is bald. On this interpretation, the sentence turns out to be true (Russell 1956, 53).

Frege regarded it as possible for an object to be given to us in a number of different ways. He observed that it is common in our natural language that one single proper name expresses many of those senses which belong to an object. For Frege, to each way in which an object is presented there corresponds a special sense of the sentence that contains the name of that object. The different thoughts that we get from the same sentence have the same truth value. In Frege's view, we must sometimes stipulate that for every proper name there is just one associated manner of presentation of the object denoted by the proper name ("Der Gedanke," 1918, KS, 350). However, he believed that different names for the same object are unavoidable, because one can be led to the object in a variety of ways ("Über den Begriff der Zahl," 1891/92, NS, 95). For Frege, our knowledge of an object determines what sense, or what senses, the name of the object expresses to us. One sense or a number of senses provides us only with one-sided knowledge (*einseitige Erkenntnis*) of an object. Frege argued: "Complete knowledge [*allseitige Erkenntnis*] of the reference would require us to be able to say immediately whether any given sense belongs to it. To such knowledge we never attain" ("Über Sinn und Bedeutung," KS, 144; Frege 1952, 58).[27]

On the basis of Frege's hints, we may conclude that his concept of *Sinn* is thoroughly cognitive. Many of his formulations suggest that *Sinne* are complexes of individual properties of objects, hence, something knowable. If this interpretation of the concept of *Sinn* were correct, it would have been Frege's view that we know an object completely only if we know *all* its properties, which is not possible for a finite human being. It would also follow that according to Frege, each object could in principle have an infinite number of names which would correspond to the modes of presentation of the object. Frege did not hold the position that knowing some arbitrary property or complex of properties of an object constitutes knowing the object completely since, for him, a necessary condition for knowing an object would be knowing all the properties of that object. Nevertheless, on the suggested interpretation he thought that in a weaker sense we know an object precisely by knowing some properties of that object. It is true Frege's weaker sense of knowing an object is not free from problems, either, even if it is more natural than the stronger sense. This is because Frege does not explain *which* properties of an object one must know to know the object.

In Frege's view, we are not able to speak about the senses of proper names *as senses*, for if we start speaking about them, they turn into objects, which, again, have their own senses. But what are these objects in case we speak about the senses expressed by proper names? Frege said that senses can be named ("Über Sinn und Bedeutung," KS, 144–145) and proposed such examples as "the teacher of Alexander the Great" and "the pupil of Plato." But if senses were complexes of the properties that belong to objects, as suggested, their names ought to be such as "being the teacher of Alexander the Great" or "being the pupil of Plato." Frege's examples suggest that when we name a sense of an object, we do not name any new object which would be a complex of individual properties of that object, but we name the original object in a new way. Hence, it follows from these examples that we do not succeed in naming a sense of an object as any new object, after all. Instead, we only name the object itself as considered under the description with which the sense provides us.

There has been much discussion on what Frege's motivation for adopting the distinction between senses and references might have been. When he introduced the distinction, he primarily referred to identity statements. It seems as if the distinction between *Sinn* and *Bedeutung* had, above all, been meant to give an adequate account of the symbol of identity, which Frege wanted to preserve in his language. By making the distinction between *Sinn* and *Bedeutung*, he sought to give a natural reading for identity statements. When introducing the concepts of sense and reference, Frege tried to solve the problems that what we now call intensional contexts caused for what we now call his idea of extensional language. The principle of functionality, which we may call the principle of compositionality in the case of references, is the core of that idea.[28]

Everything worked well according to what we would call truth tables when Frege constructed complex sentences out of simple sentences by means of conditionality (BS, §5). The trouble for Frege was caused by what became later called intensional contexts. Frege tried to deal with those contexts by introducing the concepts of indirect sense and indirect reference, the latter being the same as the normal sense of an expression. Frege claimed that in certain indirect contexts our words automatically switch their references to what normally are their senses. In a letter to Russell, he even recognized the need for using special signs for words in indirect speech (BW, 236). For example, in the complex sentences "A believes that a is P" and "A believes that b is P," "that a is P" and "that b is P" name two different thoughts, since "a" and "b" have different senses. Let us assume that a and b have the same normal reference. Given that the truth value of the complex sentence is considered to be the value of a function whose arguments are the references of the components of the sentence, it does no harm to what we call the principle of functionality even if the complex sentences have different truth values. Since the arguments of the function differ from each other, that is, because a and b have different indirect references, the references of the complex expressions may quite well be different, and the principle of functionality is thus saved.

Frege's theory of *Sinn* and *Bedeutung* was not only a solution offered to the problems that indirect contexts caused to the idea of extensional language, but it was also a direct consequence of his idea of a universal language. As noted, Frege's begriffsschrift, conceptual notation, was meant to be a realization of Leibniz's great idea. Leibniz thought that the terms of our natural language do not correspond to the things of the world in a proper way, and therefore we ought to construct a new language which mirrors correctly the whole universe.[29] He dreamed of a language that speaks about the actual world in the sense of mirroring the individual concepts instantiated in this world. Frege's world differed from that of Leibniz in the sense that for him the actual world was the only world. For Frege, *Sinne* were something that we cannot avoid when we try to reach the world by means of our language. Frege's belief in the inescapability of *Sinne* can thus be considered a special form of the Kantian belief that we must always consider objects through our conceptual systems. In "Ausführungen über Sinn und Bedeutung" (1982–1985) he remarked: "Thus it is *via* a sense and only *via* a sense that a proper name is related to an object" (NS, 135; Frege 1979, 124). Hence, the distinction between senses and references was something that Frege would have accepted in any case because of his belief in the role of conceptual machinery in reaching the world. That observation brings us back to where we started, namely, to how Frege understood the nature of his conceptual notation.[30]

Notes

I have used extracts from my article "Analysis as the Method of Logical Discovery: Some Remarks on Frege and Husserl," *Synthese* 77 (1988), 73–97, with the kind permission of Springer Science+Business Media. The chapter also contains passages from my article "Existence and Propositional Attitudes: A Fregean Analysis," *Logical Analysis and History of Philosophy* 4 (2001), 75–86, which appear here with the kind permission of Mentis, and from my article "Finnish Studies in Phenomenology and Phenomenological Studies in Finland," in Leila Haaparanta and Ilkka Niiniluoto (eds.), *Analytic Philosophy in Finland, Poznań Studies in the Philosophy of the Sciences and the Humanities*, vol. 80 (Rodopi, Amsterdam, 2003), 491–509, which appear here with the kind permission of Rodopi. I have used the manuscripts "Göttinger Vorlesungen über Urteilstheorie" (1905) and "Logik als Theorie der Erkenntnis" (1910–1911) with the kind permission of the Husserl Archives at the University of Leuven.

1. For the debate between psychologists and antipsychologists, see, for example, Kusch (1995).

2. See Friedman (1996, 2000) and Haaparanta (1999a, 2003).

3. See Beaney (2002) and Haaparanta (2007).

4. See Haaparanta (1985, 1999a) and Friedman (1996, 2000).

5. See Gabriel (1986). Cf. Ziehen (1920), 132–240.

6. See, for example, Haaparanta (1985) and Mancosu (1998). Also see Detlefsen (1992) and chapters 9 and 14 in this volume.

7. See Haaparanta (1988, 1999b).

8. See Lohmar (2002a, 2002b).

9. See Becker (1927) and Heyting (1930a, 1930b, 1931).

10. See, for example, Leibniz (1961a), 84 and 192, and Leibniz (1961b), 29, 152, and 283. See, for example, Frege, "Booles rechnende Logik und die Begriffsscrift" (1880/1881), NS, 9–52, "Über den Zweck der Begriffsschrift" (1883), BS (1964), 98, "Über die Begriffsschrift des Herrn Peano und meine eigene" (1896), KS, 227, GGA II, §§56–65, and "Anmerkungen Freges zu: Philip E. B. Jourdain, The development of the theories of mathematical logic and the principles of mathematics" (1912), KS, 341. For the terminological difference between Leibniz and Frege, see Haaparanta (1985), 11, and its references.

11. The idea that Jaakko Hintikka (1979, 1981a, 1981b) has labeled as the idea of the ineffability of semantics and to which Hugly (1973) has also paid attention in Frege's logic is visible at various points in Frege's writings. For example, see Frege's remarks on senses in "Über Sinn und Bedeutung" (1892), KS, 144–145, on functions in "Über Begriff und Gegenstand" (1892), KS, 170, on the concept of identity in "Rezension von: E. G. Husserl, Philosophie der Arithmetik I" (1894), KS, 184, and on the concept of truth in "Der Gedanke" (1918), KS, 344. Also see his informal explanations of the semantics of his conceptual notation, "Darlegung der Begriffsschrift," in GGA I. See Haaparanta (1985), 33, 41–43, 61–62, and 66.

12. See, for example, "Dialog mit Pünjer über Existenz" (before 1884), in NS, GLA, §53, "Über Begriff und Gegenstand," (1891), KS, 173, and Frege's letter to Hilbert 6.1.1900, BW, 75. See also Haaparanta (1985).

13. See Frege's "Vorwort" to GGA I. Also see his article "Der Gedanke" (1918), KS, 342–362.

14. See "Der Gedanke," KS, 351, where Frege discusses the realm of representations (*Vorstellungen*). In Frege's view, representations like sense impressions and feelings need someone who has them. Obviously, acknowledging the truth of a thought, that is, judging, needs someone who acknowledges. Frege's terminology thus suggests that he takes the acts of judging to belong to the realm of our minds.

15. References are to the manuscripts "Göttinger Vorlesungen über Urteilstheorie" (1905) and "Logik als Theorie der Erkenntnis" (1910–1911).

16. See, for example, Kemp Smith (1962), 43–45.

17. See Haaparanta and Korhonen (1996), 40–41. See Crowell (1992) and Friedman (1996), 58–59.

18. See Haaparanta (1985).

19. See Haaparanta and Korhonen (1996), 42.

20. See Haaparanta (1988).

21. See Husserl's biography in Schuhmann (1977). Even if the doctrine of "propositions in themselves" was popular among a number of Husserl's predecessors and contemporaries, Husserl's view can also be interpreted as ensuing from certain internal motives of his philosophy. This kind of reading is suggested by Cooper-Wiele (1989), who emphasizes the role of the idea of a totalizing act in Husserl's thought. See Cooper-Wiele (1989), 11 and 90–108.

22. For Husserl's concept of formal or analytical law, see LU II, A 246–251/B_1 252–256. For Husserl's discussion concerning the relationship between logical laws and the analytic a priori, see, for example, Husserl (1950b), 28.

23. The same problem had also been tackled by Hegel from an opposite point of view. In Hegel's view, Kant's problem was that his critical philosophy tried to study the faculty of knowledge before the act of knowing. Hegel argued that other tools can be studied before they are used, but the use and study of logical tools is one and the same process (Hegel, 1970, §10 and §41, Zusatz 1).

24. See Fries (1819), 8. Also see Fries (1827), 4. For Erdmann's psychologistic interpretation of transcendentalism, see Erdmann (1923), 472–477. For Frege's criticism of Erdmann, see GGA I, "Vorwort," xv–xvi.

25. Kusch (1995) has studied the sociological aspects of the debate on psychologism. My presentation is restricted to those aspects that are internal to the philosophical discussion. For Husserl's criticism of psychologism, also see Willard (1984), 143–166.

26. See Haaparanta (1985). Also see chapter 13 in this volume.

27. The word "complete" is not a good translation for *allseitig*, but it is in any case not so misleading as the word "comprehensive" chosen by Geach and Black. A better expression would, perhaps, be "knowledge from every angle."

28. See Frege's argumentation in "Über Sinn und Bedeutung."

29. See note 10. See also Leibniz (1969), sec. 8.

30. There are a great number of studies in late nineteenth-century and early twentieth-century philosophy, especially Frege and Husserl, that one could recommend for further reading, for example, Beaney (1996), Bilezki and Matar (1998), Dummett (1993), Floyd and Shieh (2001), Glock (1999), Hill (1991), Hill and Rosado Haddock (2000), Kreiser (2001), Macbeth (2005), Mendelsohn (2005), Reck (2002), Schumann (1977), Tieszen (1989, 2004), Tragesser (1977), and Weiner (2004).

References

Beaney, Michael. 1996. *Frege: Making Sense*. London: Duckworth.

Beaney, Michael. 2002. Decompositions and Transformations: Conceptions of Analysis in the Early Analytic and Phenomenological Traditions. *Southern Journal of Philosophy* 43: 53–99.

Becker, Oskar. 1927. *Mathematische Existenz: Untersuchungen zur Logik und Ontologie mathematischer Phänomäne*. Halle: Max Niemeyer.

Bilezki, Anat, and Anat Matar, eds. 1998. *The Story of Analytic Philosophy: Plot and Heroes*. London: Routledge.

Bolzano, Bernard. [1837] 1929–1931. *Wissenschaftslehre in vier Bänden*, hrsg. von W. Schultz. Leipzig: Verlag von Felix Meiner.

Boole, George. [1847] 1965. *The Mathematical Analysis of Logic, being an essay towards a calculus of deductive reasoning*. Oxford: Basil Blackwell.

Boole, George. [1854] 1958. *An Investigation of the Laws of Thought, on which are founded the mathematical theories of logic and probabilities*. New York: Dover.

Brentano, Franz. 1929. Über die Gründe der Entmutigung auf philosophischem Gebiete. In *Über die Zukunft der Philosophie*, ed. O. Kraus, 83–100. Leipzig: Verlag von Felix Meiner.

Burge, Tyler. 1992. Frege on Knowing the Third Realm. *Mind* 101: 633–650.

Carnap, Rudolf, 1931. Überwindung der Metaphysik durch logische Analyse der Sprache. *Erkenntnis* 2: 219–241.

Coffa, J. Alberto. 1991. *The Semantic Tradition from Kant to Carnap. To the Vienna Station*, ed. Linda Wessels. Cambridge: Cambridge University Press.

Cooper-Wiele, J. K. 1989. *The Totalizing Act: Key to Husserl's Early Philosophy*. Dordrecht: Kluwer.

Crowell, Stephen G. 1992. Lask, Heidegger, and the Homelessness of Logic. *Journal of the British Society for Phenomenology* 23: 222–239.

Detlefsen, Michael. 1992. Brouwerian Intuitionism. In *Proof and Knowledge in Mathematics*, ed. Michael Detlefsen, 208–250. London: Routledge.

Dummett, Michael. 1993. *Origins of Analytical Philosophy*. London: Duckworth.

Erdmann, Benno. [1892] 1923. *Logische Elementarlehre*, dritte vom Verfasser umgearbeitete Auflage, hrsg. von E. Becher. Berlin: Walter de Gruyter.

Floyd, Juliet, and Sanford Shieh, eds. 2001. *Future Pasts: The Analytic Tradition in Twentieth-Century Philosophy*. New York: Oxford University Press.

Frege, Gottlob. [1879] 1964. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* In Frege (1964), 1–88.

Frege, Gottlob. [1883] 1964. Über den Zweck der Begriffsschrift. In Frege (1964), 97–106.

Frege, Gottlob. [1884] 1968. *Die Grundlagen der Arithmetik: eine logisch mathematische Untersuchung über den Begriff der Zahl.* In *The Foundations of Arithmetic/Die Grundlagen der Arithmetik*, repr. and trans. by John. L. Austin. Oxford: Basil Blackwell. (Referred to as GLA.)

Frege, Gottlob. 1893. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet, Erster Band*. Jena: Verlag von H. Pohle. (Referred to as GGA I.)

Frege, Gottlob. 1903. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet, Zweiter Band*. Jena: Verlag von H. Pohle. (Referred to as GGA II.)

Frege, Gottlob. 1952. *Translations from the Philosophical Writings of Gottlob Frege*, ed. Peter Geach and Max Black. Oxford: Basil Blackwell.

Frege, Gottlob. 1964. *Begriffsschrift und andere Aufsätze*, hrsg. Ignacio Angelelli. Hildesheim: Georg Olms. (Referred to as BS.)

Frege, Gottlob. 1967. *Kleine Schriften*, hrsg. Ignacio Angelelli. Darmstadt: Wissenschaftliche Buchgesellschaft, und Hildesheim: Georg Olms. (Referred to as KS.)

Frege, Gottlob. 1969. *Nachgelassene Schriften*, hrsg. Hans Hermes, Friedrich Kambartel, und Friedrich Kaulbach. Hamburg: Felix Meiner. (Referred to as NS.)

Frege, G. 1972. *Conceptual Notation and Related Articles*, trans. and ed. by T. W. Bynum. Oxford: Clarendon Press.

Frege, Gottlob. 1976. *Wissenschaftlicher Briefwechsel*, hrsg. Gottfired Gabriel, Hans Hermes, Friedrich Kambartel, Christian Thiel, und Albert Veraart. Hamburg: Felix Meiner. (Referred to as BW.)

Frege, Gottlob. 1979. *Posthumous Writings*, trans. Peter Long and Roger White. Oxford: Basil Blackwell.

Friedman, Michael. 1996, Overcoming Metaphysics: Carnap and Heidegger. In *Origins of Logical Empiricism, Minnesota Studies in the Philosophy of Science, Vol. XVI*, ed. Richard N. Giere and Alan W. Richardson, 45–79. Minneapolis: University of Minnesota Press.

Friedman, Michael. 2000. *A Parting of the Ways: Carnap, Cassirer and Heidegger*. Chicago and La Salle: Open Court.

Fries, Jakob F. [1811a] 1819. *System der Logik*, zweite verbesserte Auflage. Heidelberg: Mohr und Winter.

Fries, Jakob F. [1811b] 1827. *Grundriss der Logik*, dritte Auflage. Heidelberg: Christian Friedrich Winter.

Gabriel, Gottfried. 1986. Frege als Neukantianer. *Kantstudien* 77: 84–101.

Glock, Hans-Johann. 1999. *The Rise of Analytic Philosophy*. Oxford: Blackwell.

Goldfarb, Warren D. 1979. Logic in the Twenties: The Nature of the Quantifier. *Journal of Symbolic Logic* 44: 351–368.

Haack, Susan. 1978. *Philosophy of Logics*, Cambridge: Cambridge University Press.

Haaparanta, Leila. 1985. *Frege's Doctrine of Being. Acta Philosophica Fennica* 39.

Haaparanta, Leila. 1988. Analysis as the Method of Discovery: Some Remarks on Frege and Husserl. *Synthese* 77: 73–97.

Haaparanta, Leila. 1994. Intentionality, Intuition and the Computational Theory of Mind. In *Mind, Meaning and Mathematics: Essays on the Philosophical Views of Husserl and Frege*, ed. Leila Haaparanta, 211–233. Dordrecht: Kluwer.

Haaparanta, Leila. 1995. On the Possibility of Pure Epistemology: A Husserlian Point of View. In *Mind and Cognition: Philosophical Perspectives on Cognitive Science and Artificial Intelligence*, ed. Leila Haaparanta and Sara Heinämaa, 151–167. *Acta Philosophica Fennica* 58: 151–167.

Haaparanta, Leila. 1999a. On the Relations between Logic and Metaphysics: Frege between Heidegger and the Vienna Circle. In *Metaphysics in the Post-Metaphysical Age, Papers of the 22nd International Wittgenstein Symposium, Contributions of the Austrian Ludwig Wittgenstein Society*, vol. 7 (1), ed. Uwe Meixner and Peter Simons, 243–248.

Haaparanta, Leila. 1999b. On the Possibility of Naturalistic and of Pure Epistemology. *Synthese* 118: 31–47.

Haaparanta, Leila. 2003. Finnish Studies in Phenomenology and Phenomenological Studies in Finland. In *Analytic Philosophy in Finland, Poznań Studies in the Philosophy of the Sciences and the Humanities*, vol. 80, ed. Leila Haaparanta and Ilkka Niiniluoto, 491–509. Amsterdam: Rodopi.

Haaparanta, Leila. 2007. The Method of Analysis and the Idea of Pure Philosophy in Husserl's Transcendental Phenomenology. In *The Analytic Turn: Analysis in Early Analytic Philosophy and Phenomenology*, ed. Michael Beaney, 262–274. London: Routledge.

Haaparanta, Leila and Anssi Korhonen. 1996. Kolmannen valtakunnan vieraina—huomautuksia loogisten objektien olemassaolosta. In *Tieto, totuus ja todellisuus*, ed. Ilkka A. Kieseppä, Sami Pihlström, and Panu Raatikainen, 38–44. Helsinki: Gaudeamus.

Hegel, Georg Wilhelm Friedrich. [1830] 1970. *Enzyklopädie der philosophischen Wissenschaften*. In G. W. F. Hegel, *Werke in zwanzig Bänden: Band 8*, ed. E. Moldenhauer und K. M. Michel. Frankfurt: Suhrkamp.

Heidegger, Martin. [1929] 1992. *Was ist Metaphysik?*. Frankfurt: Vittorio Klostermann.

van Heijenoort, Jean. 1967. Logic as Calculus and Logic as Language. *Synthese* 17: 324–330.

Heyting, Arend. 1930a. Sur la logique intuitionniste. *Académie Royale de Belgique, Bulletin* 16: 957–963; the English translation "On Intuitionistic Logic," in Mancosu (1998), 306–310.

Heyting, Arend. 1930b. Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 42–56; English translation "The Formal Rules of Intuitionistic Logic," in Mancosu (1998), 311–327.

Heyting, Arend. 1931. Die intuitionistische Grundlegung der Mathematik. *Erkenntnis* 2: 106–115.

Hill, Claire Ortiz. 1991. *Word and Object in Husserl, Frege, and Russell: The Roots of Twentieth-Century Philosophy*. Athens: Ohio University Press.

Hill, Claire Ortiz and Guillermo E. Rosado Haddock. 2000. *Husserl or Frege? Meaning, Objectivity, and Mathematics*. Chicago: Open Court.

Hintikka, Jaakko. 1979. Frege's Hidden Semantics. *Revue Internationale de Philosophie* 33: 716–722.

Hintikka, Jaakko. 1981a. Semantics: A Revolt against Frege. In *Contemporary Philosophy*, vol. 1, ed. Guttorm Fløistad, 57–82. The Hague: Martinus Nijhoff.

Hintikka, Jaakko. 1981b. Wittgenstein's Semantical Kantianism. *In Ethics, Proceegings of the Fifth International Wittgenstein Symposium*, ed. E. Morscher and R. Stranzinger, 375–390. Vienna: Hilder–Pichler–Tempsky.

Hugly, Peter. 1973. Ineffability in Frege's Logic. *Philosophical Studies* 24: 227–244.

Husserl, Edmund [1891] 1970. *Philosophie der Arithmetik*, mit ergänzenden Texten (1890–1901), *Husserliana XII*, hrsg. L. Eley. The Hague: Martinus Nijhoff.

Husserl, Edmund. [1900, 1913] 1950a. *Logische Untersuchungen I*, *Husserliana XVIII*, Text der 1. (1900) und der 2. (1913) Auflage, hrsg. E. Holenstein. The Hague: Martinus Nijhoff. (Referred to as LU I, A/B.)

Husserl, Edmund. [1901, 1913, 1921] 1984. *Logische Untersuchungen II*, *Husserliana XIX/1–2*, Text der 1. (1901) und der 2. (1913, 1. Teil; 1921, 2. Teil) Auflage, hrsg. U. Panzer. The Hague: Martinus Nijhoff. (Referred to as LU II, A/B_1 and LU II, A/B_2.)

Husserl, Edmund. [1913, 1921] 1970. *Logical Investigations I–II*, trans. J. N. Findlay. New York: Humanities Press. (Referred to as LI.)

Husserl, Edmund. [1910–1911] 1965. *Philosophie als strenge Wissenschaft*, hrsg. R. Berlinger. Frankfurt: Vittorio Klostermann. (Referred to as PsW.)

Husserl, Edmund. [1913] 1950b. *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie I* (1913), *Husserliana III*, hrsg. Walter Biemel. Den Haag: Martinus Nijhoff; trans. W. R. Boyce Gibson, Allen & Unwin, London, and Macmillan, New York, 1931, and F. Kersten, Martinus Nijhoff, The Hague, 1982.

Husserl, Edmund. 1929. *Formale und transzendentale Logik: Versuch einer Kritik der logischen Vernunft*. Halle: Verlag von Max Niemeyer. (Referred to as FTL.)

Husserl, Edmund. [1939] 1964. *Erfahrung und Urteil: Untersuchungen zur Genealogie der Logik*, red. und hrsg. L. Landgrebe. Hamburg: Claassen Verlag. (Referred to as EU.)

Husserl, Edmund. 1969. *Formal and Transcendental Logic*, trans. D. Cairns. The Hague: Martinus Nijhoff.

Kant, Immanuel. [1781, 1987] 1904. *Kritik der reinen Vernunft*, 1781 (A), 2nd ed. 1787 (B). In *Kant's gesammelte Schriften, Band III*, Berlin: G. Reimer; trans. Norman Kemp Smith (1929). Macmillan, London, 1929. (Referred to as KRV.)

Kemp Smith, Norman. 1962. *A Commentary to Kant's 'Critique of Pure Reason,'* 2nd ed. New York: Humanities Press.

Kreiser, Lothar. 2001. *Gottlob Frege: Leben – Werk – Zeit*. Hamburg: Felix Meiner Verlag.

Kusch, Martin. 1995. *Psychologism: A Case Study in the Sociology of Philosophical Knowledge*. London and New York: Routledge.

Lask, Emil. [1911] 1923. Die Logik der Philosophie und die Kategorienlehre. In E. Lask, *Gesammelte Schriften II*, hrsg. Eugen Herrigel, 1–282. Tübingen: J. C. B. Mohr. (Referred to as LP.)

Leibniz, Gottfried Wilhelm. [1890] 1961a. *Die philosophischen Schriften von Gottfried Wilhelm Leibniz, Siebenter Band*, hrsg. von C. I. Gerhardt. Hildesheim: Georg Olms.

Leibniz, Gottfried Wilhelm. [1903] 1961b. *Opuscules et fragments inédits de Leibniz*, ed. Louis Couturat. Hildesheim: Georg Olms.

Leibniz, Gottfried Wilhelm. 1969. Discourse on Metaphysics. In G. W. Leibniz, *Philosophical Papers and Letters*, trans. and ed. L. E. Loemker, 2nd ed., 303–330. Dordrecht: Reidel.

Lohmar, Dieter. 2002a. Elements of a Phenomenological Justification of Logical Principles, including an Appendix with Mathematical Doubts concerning some Proofs of Cantor on the Transfiniteness of the Set of Real Numbers. *Philosophia Mathematica* 10: 227–250.

Lohmar, Dieter. 2002b. The Transition of the Principle of Excluded Middle from a Principle of Logic to an Axiom: Husserl's Hesitant Revisionism in Logic. In *New Yearbook of Phenomenology and Phenomenological Philosophy*, ed. Burt Hopkins and Steven Crowell, 53–68. Madison: University of Wisconsin Press.

Lotze, Hermann. 1874. *System der Philosophie, Erster Teil: Drei Bücher der Logik*. Leipzig: Verlag von G. Hirzel.

Macbeth, Danielle. 2005. *Frege's Logic*. Cambridge, Mass.: Harvard University Press.

Mancosu, Paolo, ed. 1998. *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920s*. New York: Oxford University Press.

Mendelsohn, Richard L. 2005. *The Philosophy of Gottlob Frege*. Cambridge: Cambridge University Press.

Mill, John Stuart. [1843] 1906. *A System of Logic: Ratiocinative and Inductive*. New York: Longmans, Green.

Mitchell, O. H. 1883. On a New Algebra of Logic. In *Studies in Logic* by Members of the Johns Hopkins University, ed. Charles Peirce, 72–106. Boston: Little, Brown.

Mohanty, J. N. 1982. *Husserl and Frege*. Bloomington: Indiana University Press.

Murphey, Murray G. 1961. *The Development of Peirce's Philosophy*. Cambridge, Mass.: Harvard University Press.

Peirce, Charles Sanders. 1931–1935. *Collected Papers of Charles Sanders Peirce* (CP), vols. I–VI, ed. Charles Hartshorne and Paul Weiss. Cambridge, Mass.: Harvard University Press.

Peirce, Charles Sanders. 1958. *Collected Papers of Charles Sanders Peirce* (CP), vols. VII–VII, ed. Arthur Burke. Cambridge, Mass.: Harvard University Press.

Peirce, Charles Sanders. 1976. *The New Elements of Mathematics by Charles S. Peirce* (NE), vols. 1–4, ed. Carolyn Eisele. The Hague: Mouton.

Peirce, Charles Sanders. 1984. *Writings of Charles S. Peirce: A Chronological Edition* (W), vol. 4, ed. E. Moore et al. Bloomington: Indiana University Press.

Reck, Erich H., ed. 2002. *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy*. New York: Oxford University Press.

Russell, Bertrand. 1956. *Logic and Knowledge: Essays 1901–1950*, ed. Robert Charles Marsh. London and New York: Routledge.

Schlick, Moritz. 1918. *Allgemeine Erkenntnislehre*. Wien: Verlag von Julius Springer.

Schlick, Moritz. 1938. *Gesammelte Aufsätze 1926–1936*, Wien: Gerold.

Schlick, Moritz. 1986. *Die Probleme der Philosophie in ihrem Zusammenhang, Vorlesung aus dem Wintersemester 1933/34*, hrsg. von H. Mulder, A. J. Kox, und R. Hegselmann. Frankfurt am Main: Suhrkamp.

Schuhmann, Karl. 1977. *Husserl-Chronik. Denk- und Lebensweg Edmund Husserls*. The Hague: Martinus Nijhoff.

Seebohm, Thomas. 1989. Transcendental Phenomenology. In *Husserl's Phenomenology: A Textbook*, ed. J. N. Mohanty and W. R. McKenna, 345–385. Washington, D.C.: Center for Advanced Research in Phenomenology & University Press of America.

Sluga, Hans D. 1980. *Gottlob Frege*. London: Routledge.

Tieszen, Richard. 1989. *Mathematical Intuition: Phenomenology and Mathematical Knowledge*. Dordrecht: Kluwer.

Tieszen, Richard. 2004. Husserl's Logic. In *Handbook of the History of Logic, vol. 3: From Leibniz to Frege*, ed. Dov M. Gabbay and John Woods, 207–321. Amsterdam: Elsevier.

Tragesser, Robert S. 1977. *Phenomenology and Logic*. Ithaca, N.Y.: Cornell University Press.

Trendelenburg, Adolf. 1867. Über Leibnizens Entwurf einer allgemeinen Charakteristik. In Adolf Trendelenburg, *Historische Beiträge zur Philosophie, Dritter Band: Vermischte Abhandlungen*, 1–47. Berlin: Verlag von G. Bethge.

Weiner, Joan. 2004. *Frege Explained: From Arithmetic to Analytic Philosophy*. La Salle, Ill.: Open Court.

Willard, Dallas. 1984. *Logic and the Objectivity of Knowledge: A Study in Husserl's Early Philosophy*. Athens: Ohio University Press.

Der Wiener Kreis. [1929] 1973. The Vienna Circle of the Scientific Conception of the World (Wissenschaftliche Weltauffassung, Der Wiener Kreis). In *Otto Neurath: Empiricism and Sociology*, ed. Martha Neurath and Robert S. Cohen, 301–318. Dordrecht: Reidel.

Ziehen, Theodor. 1920. *Lehrbuch der Logik auf positivistischer Grundlage mit Berücksichtigung der Geschichte der Logik*. Bonn: A. Marcus & E. Webers Verlag.

8

A Century of Judgment and Inference, 1837–1936: Some Strands in the Development of Logic

Göran Sundholm

Dedicated to Per Martin-Löf on the occasion of his 60th birthday.

"O judgement! thou art fled to brutish beasts,
And men have lost their reason."
—Julius Caesar, Shakespeare

My office in the present chapter is to tell how, within a century, the notions of judgment and inference were driven out of logical theory and replaced by propositions and (logical) consequence. Systematic considerations guide the treatment. My history is unashamedly Whiggish: A current position will be shown as the outcome, or even culmination, of a historical development. No apology is offered, nor, in my opinion, is one needed. Philosophy in general, and the philosophy of logic in particular, treats of conceptual architecture. The logical edifice is an old one and its supporting concepts have a venerable pedigree. Many parts of the building are buried in the past. Thus, the study of conceptual architecture has to be aided by conceptual archaeology. In the light

---

The present chapter is based on lectures that I have given to second-year philosophy students at Leyden since 1990, and also draws on my inaugural lecture (1988). Per Martin-Löf's (1983) Siena lectures were an important source of inspiration, as were innumerable subsequent conversations with him on the history and philosophy of logic. In recent years conversations with my colleagues Maria S. van der Schaar and E. P. Bos have also been helpful. I am also indebted to Dr. Björn Jespersen and Dr. van der Schaar for valuable comments on the penultimate draft and to Dott.ssa Arianna Betti for much appreciated help with word processing. The material has been treated in invited lecture-courses, at the ESSLI Summerschool in Saarbrücken 1991, and at the universities of Siena 1992, Campinas and Rio de Janeiro 1993, Turku 1998, and Amsterdam 1999, as well as in a complete semester-course at Stockholm 1994. I am indebted to hosts and participants alike. My Cracow 1999 *LMPS* 11 lecture, now published as (2002), briefly tells the inference half of the tale. Translations into English are in general my own.

of the many changes that logic underwent during my chosen period, this may even seem rather apposite. Within the philosophy of logic, to understand *what* a present-day position is, it is often essential to understand *how* it became what it is. Furthermore, the systematic philosophical underpinning of present-day logic is not fixed; the balance is not ready to be drawn up. Accordingly, a survey of the historical development that led to the various options is a required aid for an informed choice among contemporary alternatives. In one essential respect, though, mine differs from a Whig history. The final outcome is not necessarily seen as an improvement on earlier but now largely abandoned views. My own preferences go in the direction of anti-realism, but a deliberate attempt has been made to adopt a neutral stance when describing the various positions.

The heroes and villains of my plot, in rough chronological order, are John of St. Thomas, Bernard Bolzano, Franz Brentano, Gottlob Frege, the Ludwig Wittgenstein of the *Tractatus*, Arend Heyting, and Gerhard Gentzen. Minor roles will be played by Immanuel Kant, Johann Gottlieb Fichte, David Hilbert, Bertrand Russell and G. E. Moore, Harold Joachim, and L. E. J. Brouwer.

The treatment will not be exhaustive. In particular, many eminent logicians will not be treated, even though they do belong to the period under consideration, for the simple reason that their contributions did not touch the systematic theme that unifies my exposition. The criteria for inclusion and exposition are based also on systematic considerations. It is my conviction, with respect to our present stage of logical knowledge, both systematic and historical, that this deserves preference above a mere recording of chronological facts. The systematic framework in which such facts are fitted confers coherence and memorability on the unfolding tale. Such a procedure is not without its dangers. They have been faced with great lucidity by Jonathan Barnes:

> On the one hand, no discussion of the ancient theories will have any value unless it is conducted in moderately precise and rigorous terms; and on the other, a rigorous and precise terminology was unknown in the ancient world. If I insist on precision I shall be guilty of anachronism. If I stick to the ancient formulations, I shall be guilty of incoherence. I prefer anachronism.[1]

Barnes's point is well taken and applies with equal force to the nineteenth century. Taking my cue from him, if methodological demands force me into anachronism, I would rather be coherent than (chronologically) right.

The (Oxford English) dictionary explains logic as the *art and science of valid reasoning*. In my chosen century, the central notion of logic is that of *judgment*. Its form and function in *inference* will play a crucial role in the sequel. Changes in the conception of judgment and, concomitantly, of inference, are central here. Other topics, such as the position of the law of the excluded third, its function as a criterion for significance, and its relation to the knowability of truth, also serve to structure the chapter. The (un)definability of truth, as well as the nature of the formal calculus used (if any), will also so serve.

1. L'ancien Régime: The Logic That Was to Be Overturned

The preface to early editions of Quine's *Methods of Logic* opened with the terse observation that "Logic is an old subject, and since 1879 it has been a great one."[2] One would be hard put not to agree with the first part of Quine's quip, but a number of us have taken issue with the second. Surely, logic was great also prior to 1879, the year in which Frege published his *Begriffsschrift*. George Boolos and Hilary Putnam have respectively dated the inauguration of logical greatness to 1847 and 1854 on the strength of the appearance of George Boole's logical works.[3] Contrary to the received Massachusetts wisdom of Harvard and MIT, it seems obvious to me that the year 1837 deserves pride of place within the history of logic as the proper counterpart to 1879.[4]

To grasp the substance and magnitude of the logical revolution, we have to consider in outline the kind of logic that was superseded. To a large extent it was nothing but a latter-day version of traditional logic, with the typical methodological accretions that became common after the *Port Royal Logic*.[5] We do well to remember that traditionally logic was conceived of as more wide-ranging than what is today the case. As a matter of fact, the "sweet Analytics of Aristotle"—*Prior* and *Posterior*—are not addressed to the same problematic. The *Analytica Priora* is devoted to the theory of *consequence*, that is, an answer is offered to the question: What follows from what? The *Analytica Posteriora*, on the other hand, treats of the theory of *demonstration*, where the crucial question is: How does one obtain further knowledge from known premises? Present-day logic restricts itself to the theory of consequence and relegates the theory of demonstration to epistemology. In the nineteenth century, on the other hand, these epistemological concerns constituted a part of *logical* theory.

At the beginning of my chosen period, the traditional patrimony is still very much in charge. The following familiar square offers a convenient starting point for (my description of) the successive revolutions in logic:

The Traditional Structure of Logic:

	Operation of the Intellect	(Mental) Product	(External) Sign
1	*Simple Apprehension*	Concept, Idea, (Mental) Term	(Written/spoken) Term
2	*Judging*, Composition/Division of two terms	Judgment, (Mental) Proposition: S is P.	Assertion, (Written/spoken) Proposition
3	*Reasoning*, Inferring	(Mental) Inference	(Written/spoken) Inference, Reasoning

The diagram[6] employs a conceptual order of priority from left to right, from acts, via products, to signs: Acts of various kinds have mental products that (may) have (outward) linguistic signs, be they spoken or written. But for this "horizontal" order of conceptual priority there is also a "vertical" order of priority among the (act-)kinds, that is, the operations of the mind. The proper subject matter of logic is *reasoning*, that is, the third operation of the mind. Nevertheless, the two other operations have to be included in the domain of logic, since inferences are built from judgments. Judgments, in turn, are formed through the composition, or division, of two concepts ("terms"). In logic, the conceptual order starts with terms, and proceeds via judgment to inference.

The traditional diagram exhibits a characteristic tripartite

$$\frac{\text{act}}{\qquad} \;\; \text{object} \longleftrightarrow \text{sign}$$

structure. Indeed, Johann Gottlieb Fichte went so far as to claim that essentially *there are only two philosophical positions* with respect to its epistemological components

$$\frac{\text{act}}{\qquad} \;\; \text{object.}^{7}$$

Either you give the object through the act, in which case—with Fichte—you are an *idealist*, or you direct the act toward the prior object, in which case you are a *dogmatist*.[8] Under this act/object structure, concepts are objects of acts of grasping ("apprehending"), and similarly the judgments made ("mental propositions") are products of the acts of judging. With respect to the third operation, though, the traditional position is not consequent. To sort this out we note a basic ambiguity in the term *inference*. On the one hand, inference may be taken in the sense of an inference *pattern* (German *Schlussweise*). Such a pattern, or *mode*, of inference can be given by means of a schema I:

$$\frac{J_1 \;\; J_2 \;\; \ldots \;\; J_k}{J},$$

where I deliberately have allowed more than the customary two premises of traditional syllogisms. The mode I of inference corresponds to a rule of inference according to which you have the right to make, that is, to know, the judgment J, provided that you have already made, that is, provided that you already know, the judgments $J_1, J_2, \ldots, J_k$. On the other hand, *inference* can also pertain to an *act* of inference, say, for instance, one made according to the mode I. Such an act has, or perhaps better, *proceeds according to*, the structure

$$\frac{\overset{|}{J_1} \;\; \overset{|}{J_2} \;\; \ldots \;\; \overset{|}{J_3}}{J}.^{9}$$

The product of an act of inference, though, is not an *act* of inference—the act, clearly, does not have *itself* as product—nor is it the *mode* of inference I, according to which the act was carried out; on the contrary, it is the *judgment made* J. The traditional diagram is accordingly in error when it puts (mental) *inference* in the product place of the act of inference (reasoning). What could such a "mental inference" be? No suitable entity seems available for service in the role. The inference *mode* is not an act of inference, nor a product of such an act; it is a blueprint, or manual, for inference acts that have products. An inference act is a *mediate* act of judgment, in which one judgment, the *conclusion*, is known on the basis of certain other judgments, the *premises*, being known. Thus, an act of inference is a particular kind of judging, whence its (mental) product is a judgment made.

Already Kant famously reversed one of the above orders of priority, namely, that between rows 1 and 2. Concepts are no longer held to be prior to judgments: "We can reduce all actions of reason to judgements, so that reason generally can be regarded as a capacity for judgement."[10] This reversal, in one form or other, we shall encounter in most of the thinkers here considered. Also other paradigm shifts in philosophy can be accounted for in terms of the traditional diagram. The most original contribution of twentieth-century philosophy, namely, the abolition of the primacy of the inner mental life that was effected by Wittgenstein,[11] can be seen as nothing but a reversal of the priorities between the second and third columns. The outward sign is no longer conceptually posterior to the inner product.

2. Speech Act *Intermezzo*: A Unified *Linguistic* Account for Some Nineteenth-Century Changes

Traditionally, the linguistic counterpart to the mental judgment made is the assertion. This term, in common with other English *-ion* words, exhibits a process/product ambiguity.[12] It may concern the act of asserting (judging) or the product of such an act, the assertion (judgment) made. The appropriate linguistic tool for assertion is the declarative sentence. In general, when S is a declarative sentence the question

Is it true that S?

may legitimately be put. An assertion that snow is white is readily effected by means of a single utterance of the declarative sentence

Snow is white.[13]

By convention, in the absence of counterindications that it should not be so held, a single utterance of a declarative is an assertion. For instance, the declarative S is not used assertorically in

Consider the example: S.

or

He claimed that S, but I don't know whether it really constitutes so.[14]

Not every use of a declarative sentence is assertoric, but assertoric uses can be recognized as such since the counterquestions:

How do you know that S is true? What are your grounds?[15]

are a legitimate response to an assertoric use of S. The *content* of the assertion effected by means of an assertoric utterance of the declarative

Snow is white,

that is, the assertion that snow is white, is given by means of a nominalized *that clause*,

that snow is white.

In general, a single utterance of this clause alone will not serve to effect an assertion that snow is white.[16] To get back to a declarative, a single utterance of which will so serve, one must either append

is true

or prefix

it is true

to the clause in question. Then we obtain, respectively,

that snow is white is true

and

it is true that snow is white,

single utterances, either of which do suffice for asserting that snow is white. Note that the first of these two formulations admits of the prefix *the content*. It then yields a yet fuller but still equivalent formulation of the judgment made:

The content that snow is white is true.

The second formulation, though, resists the corresponding interpolation, which results in ungrammatical nonsense:

It is true the content that snow is white.

These considerations suggest that

judgable content A is true

is the proper form of judgment, when one prefers a *unary* form of judgment that makes explicit the content judged in the judgment made.[17] The content in question will be given by a that-clause formed from a declarative S. The judgment made in or by the act of judging that is made public through the assertoric utterance of the declarative S accordingly, takes the form

that S is true.[18]

It must be stressed, though, that this path to this unary, content-explicit form of judgment is manifestly *not* language independent, because it draws extensively on linguistic considerations, albeit very simple ones. As such, it would be rejected out of hand by most major figures considered here, even though the unary form of judgment itself is accepted. Different analyses offered by various logicians reach the same result, but different routes are taken. Nevertheless the speech act theory route to the unary form of judgment constitutes as good an example as any of the characteristic—twentieth-century—*linguistic turn in philosophy* that was inaugurated by Frege (1891): ontological and epistemological questions are now answered (while recast in linguistic form) via a detour through language.[19] However, drawing on the traditional conceptual link between judgment and assertion, namely, that between mental object and exterior sign, the above exposition, in spite of its anachronistic (twentieth-century) flavor, explains *why* the (nineteenth-century) unary form of judgment has to take the form it has.

3. Revolution: Bolzano's *Annus Mirabilis*

I postulated that 1837 was a crucial year for logic, no reason being given. However, in this year the four hefty tomes of Bernard Bolzano's *Wissenschaftslehre* made their weighty appearance.[20] This event constitutes the greatest revolution in logical theory since Aristotle, even though the *Wissenschaftslehre* fell stillborn from the press, as far as near-time influence is concerned, owing to clerical and political censorship. Indeed, in the preface to the second edition of his main work (the first edition of which appeared in the year of Bolzano's birth) a very distinguished professor of philosophy could still write: "Since Aristotle, [Logic] has not had to retreat a single step. Also remarkable is that it has not been able to take a single step forward, and thus to all appearance is closed and perfect,"[21] which state of affairs continued until the coming of the second nineteenth-century revolution in logic. Within logical theory, 1879, the year of Quine's choosing, is the counterpart to the *second* revolutionary year 1848. Traditional logic was first and foremost a *term logic*, rather than a propositional logic. In spite of the medieval scholastic achievements concerning the theory of *consequentiae*, and the insights of the—much earlier—Stoic logic, the syllogism, in one version or other, still ruled supreme, which circumstance renders Kant's opinion considerably less farfetched than it might seem today. For instance, his own conception of logic as set out in the *Jäsche Logik* (whether it be truly Kantian or not) is cast entirely in the customary traditional mold.[22]

Bolzano's revolution with respect to the traditional picture is threefold. First, the middle ("product") column of the traditional schema is *objectified*. The mental links are severed, and thus, in particular, the traditional notions *mental* term (concept, idea) and *mental* proposition (judgment) are turned into their ideal, or *Platonist*, counterparts idea-in-itself (*Vorstellung an sich*) and proposition-in-itself (*Satz an sich*).[23] Second, the pivotal middle square of the diagram is altered: The judgment made no longer takes the traditional (S is P) form. Logic is no longer *term* logic. Instead Bolzano uses the propositional, unary form of judgment that was canvassed above, with his *Sätze an sich* taking the role of judgable contents:

The *Satz an sich* S is true.[24]

Third, Bolzano bases his logical theory, not on inference (from judgments known to judgment made), but on (*logical*) *consequence* between propositions.[25] Judgment is dethroned and its *content* now holds pride of place in logical theory.

Needless to say, Bolzano, a priest steeped in the tradition, does not jettison everything traditional: A *Satz an sich*, that is, the judgable *content*, rather than the judgment made, has (or can brought to) the canonical form

V_1 has V_2,

which is very close to the Aristotelian form

S is P.

Instead of the Aristotelian *judgment*

Man is mortal

we find the Bolzanian *content*

Man has mortality.

The precise reasons for this shift from the concrete *mortal* to the abstractum *mortality* need not detain us here; in essence, Bolzano takes the Aristotelian form of judgment and turns it into a form of content, where the contents are objectified denizens of the ideal—Platonic—third realm.[26] Bolzano's key notion is that of proposition-in-itself: The idea-in-itself is explained as a part of a proposition-in-itself that is not a proposition-in-itself.

Bolzano's logical objectivism is a *Platonism*: As already noted, his crucial *an sich* notions are all *ideal*. We are not told very much about what *ideal* means here. Instead, his manner of proceeding is that of a *via negativa*: a list of *non*applicable attributes is offered. Thus, the ideal realm is characterized as *a*temporal, *a*spatial, *inert*, *non*linguistic, *non*mental, *un*changeable, *non*generated.... Furthermore, the propositions-in-themselves serve in various logical roles, in particular as contents of mental acts and declarative sentences.[27]

However, not only propositions and their parts are ideal *an sich* notions: The *truth* of a true proposition-in-itself is truth-*in-itself*. Bolzano's explanation of truth is an interesting one. According to him, all propositions have or can be brought to the logical form (V_1 has V_2), and so truth only has to be explained for propositions of this form:

> The proposition (V_1 has V_2) is true if and only if V_1 really (German *wirklich*) has V_2,

for instance,

> the proposition-in-itself that snow has whiteness is true if and only if snow really has whiteness.[28]

This, virtually "disquotationalist," rendering is compatible with currently fashionable "minimalist" positions concerning truth. Bolzano, however, was on the road toward a more substantial notion of truth when he noted that the following proportion holds concerning truth and a certain kind of existence, namely, that of *instantiation* (German *Gegenständlichkeit*), that is, the higher-order property of an idea-in-itself of being instantiated:

> the similarity between this relation among propositions and . . . that among ideas is obvious. Namely, what holds, concerning ideas, for the circumstance whether indeed a certain object falls under them or not, holds, concerning propositions, for the circumstance whether truth pertains to them or not.[29]

In the form of a proportion:

$$\frac{\text{proposition-in-itself}}{\text{truth}} = \frac{\text{idea-in-itself}}{\text{instantiation}}.$$

Thus, what it is for a proposition-in-itself to be true is what it is for an idea-in-itself to have something falling under it. In other words, applied to my (snow-bound) stock example:

> the proposition-in-itself *that snow is white* is true (or has truth, in the terminology preferred by Bolzano) precisely when the idea-in-itself *the whiteness of snow* has nonemptiness, that is, when some entity falls under *the whiteness of snow*.[30]

Bolzano here anticipates something of considerable importance for the analysis of truth, and we shall have occasion to return to his comparison in the sequel.

Bolzano's apparatus for logical analysis, comprising *propositions*, *ideas*, and *instantiation*, is highly versatile.[31] Thus, for instance, as Leibniz knew, the four categorical Aristotelian judgments are readily cast in the required form. For instance, an E judgment,

$$\text{No } V_1 \text{ are } V_2,$$

is rendered

> the Idea (in-itself) of a V_1 that is V_2 does not have existence (*Gegenständlichkeit*).[32]

The notion of truth *in itself* for propositions-in-themselves is *bivalent*:

> For every proposition-in-itself A, A has truth or A has falsity, also in-itself.

The *an sich* character of the truth of true propositions-in-themselves is one of the pillars on which Bolzano's logical realism rests.[33] Another is the reduction of epistemological matters to the Platonist *an sich* notions. The first instance of this reduction concerns judgment: a judgment of the novel form, that is,

$$\text{proposition-in-itself } A \text{ is true}$$

is correct (*richtig*) if A really is a truth-in-itself.[34]

This reduces the epistemic notion of the correctness for judgments to the Platonist *an sich* notion of truth for propositional contents. Here Bolzano pays a price—in my opinion too high a price—for his iron-hard realism in logic *and* epistemology. Under the Bolzano reduction, a *blind* judgment, a mere guess, without any trace of justification, is a piece of *knowledge* (an *Erkenntnis*).[35] The only thing that matters is the *an sich* truth, whether knowable or not, of the proposition-in-itself that serves as content of the judgment in question.[36] Thus, for instance, according to Bolzano, if, independently of any counting, it happens to hit bull's eye, my unfounded claim that the City Hall at Leyden has 1234 window panes, is simply a piece of *knowledge*. In this I, for one, cannot follow him. Bolzano deserves high praise for his lucid and uncompromising realism. Also antirealists profit from reading him: His version of realism is one of the very best on offer.[37] Admitting blind judgments as *pieces of knowledge*, however, is not just realism but realism run rampant.

Bolzano's transformation of the third and final notion in the traditional picture, namely, that of inference, makes an unmistakably modern impression. The changed form of judgment transforms the inference schema I into I′:

$$\frac{A_1 \text{ is true} \quad A_2 \text{ is true} \quad \ldots \quad A_k \text{ is true}}{C \text{ is true}}.$$

An inference according to I′ is valid if the proposition-in-itself C is a logical consequence of the propositions-in-themselves $A_1, A_2, \ldots, A_k$. Such a *logische Ableitbarkeit*—Bolzano's terminology—holds between the A's and C when each uniform variation V of all nonlogical ideas that makes all the A's true also makes C true.[38] In other—more modern—words, C is *a logical consequence* of $A_1, A_2, \ldots, A_k$ when the proposition-in-itself

$$(A_1 \,\&\, A_2 \,\&\, \cdots \,\&\, A_k) \supset C$$

is not just true but *logically* true, that is, true under all uniform variations of its nonlogical parts.[39]

The notion of an *Ableitbarkeit* provides yet another Bolzano reduction of an epistemic notion to Platonist *an sich* notions. In the same fashion that

Bolzano reduced the (epistemic) correctness (*Richtigkeit*) of the judgment made to the *an sich* truth of its *an sich* content, the validity of an inference is also reduced to, or in this case perhaps better, *replaced by* something on the level of the Platonist contents of judgments. Indeed, the fourth chapter, §§223–268, of the *Wissenschaftslehre* bears the title "Von den Schlüssen," but deals with *Ableitbarkeiten* among propositions-in-themselves, rather than with judgments that are made on the basis of certain other judgments already having been made.[40] Thus the inference is valid or not, irrespective of whether it transmits knowledge from premise judgments to the conclusion judgment, solely depending on the *an sich* truth-behavior of the propositions-in-themselves that serve as contents of the judgments in question, under all variations with respect to suitable in-themselves parts of the relevant propositions. Bolzano's position is accordingly threatened not just by the phenomenon of blind knowledge. Under his account also inference can be *blindly valid*, irrespective of whether it preserves knowability from premise(s) to conclusion.

Logical consequence (*logische Ableitbarkeit*) is a relation that may obtain between any propositions whatsoever, be they true or false. Bolzano also studies another consequence relation among propositions, but now restricted to the field of truths-in-themselves only, that he calls *Abfolge* (grounding). The theory of Bolzano's grounding relation is difficult and as yet not very well explored; it can be seen as yet another reduction of epistemic notions to Platonist ones. Consider the inference I″:

$$\frac{A \text{ is true}}{B \text{ is true}}.$$

When I″ is valid, that is, preserves knowledge from premise to conclusion, and the premise is known, the judgment A *is true* serves to ground the judgment B *is true*. Then a certain relation obtains between the propositions A and B that serve as contents of the judgments in question. *Abfolge* can be seen as a "propositionalization" $\mathbf{Abf}(A, B)$ of that relation: The relation of grounding, *which holds* in the first instance between pieces of knowledge, that is, *between judgments known*, is turned into a propositional relation ("connective") between propositions, that is, contents of judgments. Every truth has a grounding tree that is partially ordered according to the *Abfolge* relation.[41] It can be seen as an ideal proof that shows *why* the true proposition is true, somewhat along the lines of Aristotelian demonstrations διοτι.[42]

In the light of Bolzano's innovations and ensuing reductions, it is important to distinguish between the holding of a consequence, that is, the preservation of truth from antecedent propositions to consequent proposition, and the validity of an inference figure, that is, the preservation of knowability from premise judgments to conclusion judgment.[43] This insight is lost to modern philosophy of logic that largely accepts the Bolzano reduction to such an extent that (validity of) inference and (logical holding of) consequence are identified.

4. Revisionism; the "Novel" Contributions of Brentano

Franz Brentano, in lectures given at the Universities of Würzburg and Vienna, from the early 1870s onward, proposed another revision of traditional doctrine. Because of his distaste for all Platonist notions in logic, such as Bolzano's proposition-in-itself, Brentano rejected the single unary form of judgment that ascribes truth to a Platonist content.[44] Instead, he canvassed the use of *two* unary forms of judgments, namely,

$$\alpha \text{ IS (exists), in symbols } \alpha+,$$

and

$$\alpha \text{ IS NOT (does not exist), in symbols } \alpha-,$$

where α is a (general) concept. Brentano, however, was not the first to note this. Already Bolzano explicitly considered these forms, under the respective guises of

α has Non-Emptiness (*Gegenständlichkeit*)

and

α has Emptiness,

and determined their most important properties. In particular, we already noted, Bolzano knew that the four Aristotelian categorical judgments can be dealt with using these two forms.[45] Credibility might not be stretched to the point of credulity if we surmise that this anticipation provides one of the reasons for Brentano's staggering lack of generosity toward the Great Bohemian:

> When ... I drew attention to Bolzano, this ... in no way, was intended to recommend Bolzano as a teacher and leader to the young people. What they could learn from him, I dare say, they could learn better from me. ...
>
> And ... as I myself never took a single thesis from Bolzano, so I was never able to convince my pupils that they would find there a true enrichment of their philosophical knowledge.[46]

Under the circumstances, "methinks the learned Gentleman doth protest too much!" However, it is not unlikely that also Bolzano's logical objectivism disqualified him as a "teacher and leader" in the eyes of Brentano, who distrusted all kinds of logical Platonism.

Of more lasting value than Brentano's employment—and alleged rediscovery—of the Leibniz–Bolzano reductions are his views on the blind judgment.[47]

These have profound consequences for his formulation of the traditional laws of thought, such as noncontradiction and excluded third, as well as for the relation between truth and evidence. Young man Brentano construed evidence as "experience of truth" (German *Erlebnis der Wahrheit*—Husserl's terminology), whence the order of dependence goes from truth to evidence.[48]

Later, under the pressure from the phenomenon of blind judgment, he reversed this order of priority and held that truth (correctness, German *Richtigkeit*) should be seen as *possibility for evident judgment*:

> Truth pertains to the judgments of he who judges rightly, that is, to the judgements of him who judges what someone would judge who judged with evidence; that is, he who asserts what would be asserted also by someone judging with evidence.[49]

Similarly he is led to a *negative* formulation of the law of excluded middle:

> It is impossible that someone, who rejects something that is wrongly accepted by someone else, rejects it wrongly, as well as that someone who accepts something, that is wrongly accepted by someone, accepts it wrongly, presupposed ... that both judge with the same mode of representation and with the same mode of judgement.[50]

From an antirealist point of view, Brentano is certainly on the right track; he refrains from asserting that a content must be either true or false, in entire independence of whether it is known to be so. His formulation, though, is not entirely correct. Brentano, the great crusader against the blind judgment, here forgets to take it into account. Of course, it is possible that the object A is wrongly accepted by P_1, as well as wrongly rejected by P_2, namely, when P_1 and P_2 *both judge blindly*, that is, without evidence. On the other hand, the corresponding formulation of Noncontradiction is correct: It is impossible that someone rightly rejects what is rightly accepted by someone else.

5. Functions Triumphant: Frege's Account of Judgment and Inference

Frege, *pace* Quine, is generally held to have inaugurated the revolution in logic. From the present perspective though, his contribution is remarkably slender. Logical objectivism, with its novel unary judgment, is present wholesale already in Bolzano, where it is cast in a more perspicuous form. Frege, furthermore, does not treat of logical consequence among propositions, or *Thoughts*, as he called them. For better or worse, Bolzano, with his insistence on replacing inference with the notion of consequence, makes a much more modern impression than Frege, whose traditional views on inference have come in for much criticism. We must not forget, however, that Frege was a mathematician and from the outset his aims were those of a mathematician rather than of a philosopher. His contributions to my topic are all subservient to the aim of providing a secure foundation for mathematical analysis, very much in the style of traditional Aristotelian foundationalism: One seeks a small number of primitive concepts, and basic truths concerning those primitives, in terms of which, at least a very sizable part and preferably all, of mathematics can be formulated, while its truths can be derived by means of primitive inference steps, where the basic axioms and primitive inference steps are made evident from the concepts they contain.[51] In the *Begriffschrift* booklet from 1879 (what turns out to be) a preliminary version of the formal language is given and the basic notions explained. In the *Grundlagen der Arithmetik* from 1884 the program of securing the mathematical theorems by means of reducing the mathematical axioms to logical theorems is spelled out informally. However, Frege was aware of the fact that he had only made plausible the reduction of arithmetic to logic, since, possibly at the instigation of the Brentanist Carl Stumpf, the *Grundlagen* development was informal and not carried out in the begriffsschrift. Thus Frege could not guarantee that his demonstration were really gap-free. The means of demonstration, whether logical or arithmetical, were not explicitly listed. Accordingly, his inferences have not been made evident solely from the concepts employed in them, and so the arithmetical edifice remains shaky. The (considerable) changes in the begriffsschrift that were put into effect around 1890 served to make the formal execution of the logicist program feasible; unfortunately, the project failed owing to the emergence of the Zermelo–Russell paradox in Frege's system.

Thus, when compared to Bolzano, Frege's most important contribution is his begriffsschrift.[52] By creating this formal language, Frege provides a partial realization of the Leibnizian *calculus ratiocinator* project. That an inference step, or axiom, is valid depends on contentual aspects pertaining to the notions from which the step, or axiom, in question has been built.[53] However, once such a step has been explicitly formulated and validated in terms of contents, it is mechanically recognizable as such. No further contentual, "intuitive" considerations are required to determine whether the inference in question is valid; being of the appropriate syntactic *form* suffices and that form is mechanically, or "blindly," recognizable.

As far as the theoretical framework is concerned, Frege's one step over and beyond Bolzano is minute but with enormous consequences. In both early and mature formulations of his theory of judgment, Bolzano's unary form of judgment is retained:

The circumstance that S is a fact

and

> a judgment is not the mere grasping of a Thought, but the acknowledgment of its truth.[54]

Frege, however, by training and profession was a *mathematician*. His teaching activity was mainly devoted to analytical geometry. Through his mentor Ernst Abbe, one of Riemann's few students, he also gained access to the latest developments in the then emerging function theory, that is, that branch of mathematical analysis that deals with analytic functions in the complex plane. His logical revolution draws heavily on the notion of function: Instead of Bolzano's clumsy form of content "A has b", Frege carves up his contents using the versatile form

$$P(a),$$

that is, function P applied to argument a. Frege's logic is mathematized from the outset. It is especially well suited for coping with Weierstraß's rigorous treatment of analysis; indeed, the notation could have been (and probably was) invented for the very purpose.[55] The familiar concepts of pointwise continuity, and its refinement into uniform continuity, illustrate this:

$$(\forall x \in I)(\forall \varepsilon > 0)(\exists \delta > 0)(\forall y \in I)(|x - y| < \delta \supset |f(x) - f(y)| < \varepsilon)$$

and

$$(\forall \varepsilon > 0)(\exists \delta > 0)(\forall x \in I)(\forall y \in I)(|x - y| < \delta \supset |f(x) - f(y)| < \varepsilon).$$

These succinct formulations show how admirably the Fregean quantifier is geared to expressing distinctions involving *multiple generality*.[56] A verbal, natural language treatment would be much harder to take in.

Frege's function-theoretic conception of logic imposed an interesting bifurcation on his views on truth. Mathematicians speak of the value of a function for a certain argument. For instance, $2 + 2$ is the value of the function $x + 2$ for the argument 2. In the first instance, the plus-two function takes numbers into numbers, but owing to Frege's doctrine of universality, it has to be extended into one defined for all objects. One then makes use of what Quine has called a "don't care" argument, for instance,

$$\xi + 2 =_{\text{def}} \begin{cases} r + 2 & \text{if } r \text{ is a number;} \\ \text{the Moon} & \text{otherwise.} \end{cases}$$

Adopting the same perspectives also at the level of sentences, from the complete sentence

Caesar conquered Gaul,

we get the function

ξ conquered Gaul,

which must also be defined for *all* objects, including me, the Moon, and Louis XIV, as well as the number of those grains of sand at Syracuse beach that were not counted by Archimedes when writing the *Sandreckoner*, and plutonium, an element unknown at the time of Frege. A value of the conquering Gaul function will have to be something close to a judgable content, or Thought. It will not, however, *be* a judgable content, because it is not *invariant* under different descriptions of the argument. Frege's by now notorious example concerning the planet Venus makes this clear:

> Venus = Venus, The Morning Star = Venus, and the Evening Star = Venus

are three values of the function

$$\xi = \text{Venus}.$$

Since in all three cases the argument—Venus = The Morning Star = The Evening Star—is the same, the value has to be the same. The Thought expressed is different in all three cases.[57] Accordingly, the value of the function for the argument Venus (under any description) is not a Thought. Instead Frege avails himself of certain ideal objects, the True and the False, that are known as "truth values," and serve as appropriate function-values. By the truth value of a sentence, Frege understands the circumstance that it is true or that it is false.[58] Thus, the common function-value in the three cases above is the truth value the True. Sentences are then seen as *truth value names*.[59] In his elucidation of the revised begriffsschrift Frege lays down, for each regular sentence, under what condition it is a name of the True. The sentence then expresses, or has as its sense (*Sinn*), the Thought that this truth condition is fulfilled.[60] Frege's theory of meaning is a bipartite mediation theory, very much along the lines of early medieval theories of signification: The sign expresses its sense that refers to an entity (called *Bedeutung* by Frege).

In Frege's theory a number of themes are dealt with that were touched on in the section 2. Frege deemed it necessary to include in his begriffsschrift a specific symbol,

$$\vdash$$

that makes explicit the assertoric force that the *Kundgabe* of a judgment made carries. Frege's view of inference has come in for much criticism; an inference is an "act of judgement, which is made, according to logical laws, on the basis of judgements already made."[61] On the symbolic level, this is reflected in the omnipresence of the judgment stroke, both on premises and conclusion, in Frege's formal inference-figures in the *Gg. Modus ponens*, for instance, takes the form

$$\frac{\vdash A \supset B \qquad \vdash A}{\vdash B}.$$ [62]

The sign "⊢" has changed its meaning and in the logic of today it is an ordinary (meta)mathematical predicate applicable to certain (meta)mathematical objects, namely wff's, that is, elements of a free algebra of "expressions" generated over a certain "alphabet." When $\varphi \in$ wff, "$\vdash \varphi$" has the meaning *there exists an inductively defined derivation-tree of wff's with* φ *as end formula*; in particular, the Frege sign does no longer function as a force indicator, but can be negated and occur in an antecedent of an implication.[63] In Frege, however, it is clear that it expresses assertoric force. Thus, both premises and conclusion of inferences are known, since, as we remarked, assertions made do contain claims to knowledge. The practice of drawing inferences from mere hypothesis, however, in particular as embodied in the works of Gerhard Gentzen, is held to refute Frege at this point.[64]

Frege was firmly committed to realism:

> Being true (*Wahrsein*) is something different from being held true (*Fürwahrhalten*), be it by one, be it by many, be it by all, and

> can in no way be reduced to it. It would be no contradiction that something is true that is held false by all.[65]

This is a clear statement of one of the central roles of truth, namely, its metaphysical role. By this I understand the task of truth to hold open the possibility of making mistakes. It is a minimum requirement on any viable epistemological position that it must allow for the possibility of mistaken acts of knowledge: "What is true is independent of our acknowledgement. We can make mistakes."[66]

The opposite, "Protagorean" position would make man the measure of all things and would equate truth with truth-for-us. It would constitute an epistemological nihilism, where anything goes, along the lines of moral nihilism within ethics: "If God is dead, everything is permitted." Mistaken deeds, be they logico-epistemic or ethical, presuppose a *norm*. Frege avails himself of the required norm via the notion of truth for judgable contents. He then reduces the rightness (Latin *rectitudo*) of epistemic acts, that is, the notion that is needed, strictly speaking, to uphold metaphysical realism, to that of the correctness of the judgment made in such an act, and that correctness finally to the truth of the Thought that serves as content of the judgment made. Truth for Thoughts, finally, is *bivalent*: Any Thought is either true or false, come what may. Frege secures this via his doctrine of sharp concepts. Thoughts are the result of applying concepts, that is, functions from objects to truth values, to objects. Concepts have to be sharply defined on *all* objects:

> The Law of Excluded Third is really the requirement that concepts be sharply delineated in another guise. An arbitrary object Δ either falls under the concept Φ, or it does not fall under it: *tertium non datur*.[67]

Thus Frege, and before him Bolzano, secures the metaphysical role of truth, namely, that of providing the notion of rightness for epistemic acts, via the bivalence of truth for judgmental contents. This is not the only way to secure the notion of rightness for acts; Brentano, for instance, rejected the notion of proposition (Thought) and instead used product-correctness as the basic, absolute notion. Wittgenstein, on the other hand, did not take propositional truth as the basic notion, the way Frege and Bolzano did, but reduced it to the ontological notion of obtaining with respect to states of affairs. One can also take the notion of rightness as a primitive notion sui generis, which is my own preferred option.[68]

Frege throughout his career held the view that truth (for propositions) is sui generis and indefinable. Since the Thought that S is the same as the Thought that it is true that S, every Thought contains (the notion of) truth and so there is no neutral ground left from which to formulate a definition: Every putative definiens irreducibly contains the definiendum in question. Frege's realism, just like Bolzano's, is a logical one: There is no attempt at a further ontological reduction of propositional truth. For Frege, a fact is nothing but a true Thought and correspondence theories of truth are firmly rejected.[69] His ontology is very sparse: objects, functions, and that is all—no facts, no states of affairs, no tropes, or what have you. Frege held the wheel at the first bend of the linguistic turn. His only category distinction is that between saturated and unsaturated entities, and this ontological distinction draws on the linguistic distinction between expressions with and without gaps into which other expression may be fitted.

In spite of his thoroughgoing realism, Frege appears committed to the view that every true proposition can be known as such:

> The most secure demonstration is obviously the purely logical, which, abstracting from the particular character of the things, rests only on the laws on which all knowledge depends. We then divide all truths that require a justification into two kinds, in that for the one, the demonstration can proceed purely logically, for the other has to be based on facts of experience.[70]

Truths are then divided into those that need justifications and those that do not; the former are split into those that have purely logical demonstrations and those whose demonstrations rest on experiential facts. Thus, in either case, it appears that if the truth is one that stands in need of justification, then there is a demonstration. Thus all truths can be known: If it needs no justification, it can be known from itself, whereas truths that do need justification can be known through a demonstration, be it logical or empirical.

6. Truth Made: The Correspondence Theory Strikes Back

Half a decade after Frege's *Hochleistungen*, G. E. Moore inaugurated his realist apostasy from the Hegelianism of his philosophical apprenticeship by adopting something very much like Bolzano's theory of propositions with an *an sich* notion of simple truth. In this he was soon followed by Bertrand Russell.[71] Russell and Moore were not crystal clear (to put it mildly). The best formulation of their novel theory was offered by a staunch upholder of the old order, the idealist H. H. Joachim, whose aptly titled (1906) book *The Nature of Truth* has a chapter TRUTH AS A QUALITY OF INDEPENDENT ENTITIES. His characterization of the *an sich* theory of truth is a powerful one:

> "Truth" and "Falsity," in the only strict sense of the terms, are characteristics of "Propositions." Every Proposition, in itself in an entire independence of mind, is true or false; and *only* Propositions can be true or false. The truth or falsity of a Proposition is, so to say, its *flavor*, which we must recognize, if we recognize it at all, immediately: much as we appreciate the flavor of pineapple or the taste of gorgonzola.[72]

Joachim also articulated clearly the possibility of unknowable truths on the *an sich* reading of truth: "The independent truth will be and remain entirely in itself, unknown and unknowable."[73] In an oblique way Russell had already admitted of the possibility of unknowable truths:

> Now, for my part, I see no possible way of deciding whether propositions of infinite complexity are possible or not; but this at least is clear, that all the propositions known to us (and it would seem, all propositions that we *can* know) are of finite complexity.[74]

In philosophy, claims that something cannot be done are dangerous and invariably tend to provoke attempts to achieve what has been denied. Frege's view that truth is sui generis and cannot be defined was challenged even before it had been published:[75] After yet another decade of logico-semantical soul-searching Moore and Russell were veering toward the correspondence theory of truth.[76] Both gave *reductions* of truth in ontological terms by means of a *truth-maker*[77] analysis in the form

$$\text{proposition } A \text{ is true} = \text{there exists a truth-maker for } A.$$

In a truth-maker analysis, to each proposition there is related a suitable notion of truth-maker and also a suitable notion of existence with respect to such truth-makers. Moore chose "facts" as his truth-makers and Russell used "complexes." For Moore, a proposition is true if it corresponds to an existing fact, and for Russell it is true if the complex to which it corresponds exists. The intricacies of their respective ontologies of facts and complexes need not detain us here; both were superseded by Wittgenstein's *Tractatus* and are now merely of historical interest.

The *Tractatus* rests on three main pillars, to wit (i) Wittgenstein's famous *picture theory* of linguistic representation; (ii) the doctrine of *logical atomism*, according to which every proposition is a truth-function of elementary propositions; and (iii) the *Saying/showing* doctrine. Of these the picture theory serves to structure the work.[78] In a brief attempt at an exposition, I treat the proposition

(∗) Peter is the father of John

as if it were a Tractarian elementary proposition.[79] Thus, our example (∗) is an elementary proposition of the form

$$aRb.$$

Hence, it must (?) immediately (?) strike us as a picture and indeed even one that obviously resembles its subject matter (4.12). How can we make sense of this?

On the ontological side, in the world, we have the state of affairs that Peter and John stand in the father-son relation. We now have to construe the propositional sign used to express the proposition (∗) as a fact that serves to present this state of affairs. The two structures—linguistic and ontological—have to be, in mathematical parlance, *isomorphic*.

Language	World
"*Peter*"	Peter
"*John*"	John
$Q(a, b)$	father-son relation
Q("*Peter*", "*John*")	Peter and John's standing in the father-son relation

Our task to ensure isomorphism between language and world amounts to finding an appropriate Q-relation. Obviously the field of such a relation must consist of expressions and this is the key to Wittgenstein's solution:

> $Q(\alpha, \beta) =_{\text{def}}$ the expressions α and β stand, respectively, immediately to the left and to the right of the sign-array "*is the father of.*"[80]

Hence, "that 'Peter' stands in a certain relation, namely the Q-relation, to 'John', says that Peter and John stand in the father-son relation" (3.1432). Using the Q-relation, the sentence-sign $(*)$ is (or can be viewed as) a *fact*, since the two proper names do stand in the Q-relation. This *syntactic* fact in turn serves to present the state of affairs that Peter is the father of John. When this state of affairs exists (or *obtains*), it is a fact, and the proposition is true. In this case the proposition is a picture *of* the fact.

According to the picture theory, every atomic, or *elementary*, proposition E presents a state of affairs (*Sachverhalt*) SE that may or may not obtain (4.21).[81] Accordingly, if the presented state of affairs SE *obtains* the elementary proposition is true and *depicts* (what is then) the *fact* SE (4.25, 2). States of affairs are logically independent of each other; from the obtaining of one nothing can be concluded about the obtaining of another (2.062). A point (*Wahrheitsmöglichkeit*) v in logical space LS is an assignment of + (obtains) and − (does not obtain) to each state of affairs (4.3); in other words, a point in logical space is a function v from states of affairs to $\{+, -\}$. Thus, $LS = \{+, -\}^{SV}$, that is, the collection of functions from the collection SV of *Sachverhalte* to $\{+, -\}$.

A situation (*Sachlage*) σ in logical space is a partition of LS into two parts σ^+ and σ^- (2.11). Points in the positive part σ^+ are compatible and those in the negative part σ^- are incompatible with σ. A proposition A is a *truth-functional combination* of elementary propositions (5).[82] The truth-functional composition of the proposition A determines whether A is true or false with respect to or *at* a point v in LS. A point $v \in LS$ induces a $\{\mathbf{T}(\text{rue}), \mathbf{F}(\text{alse})\}$-valuation $\mathbf{v}$ on truth-functional propositions in the following way:

For an elementary proposition E,

$$\mathbf{v}(E) = \mathbf{T} \quad \text{if } v(S_E) = +;$$
$$\mathbf{v}(E) = \mathbf{F} \quad \text{if } v(S_E) = -.$$

Thus, an elementary proposition is true at a point v if v is compatible with the state of affairs that the elementary proposition presents.

For the proposition $A = \mathbf{N}(\bar{\xi})$,

$$\begin{aligned} \mathbf{v}(A) = \mathbf{T} &\quad \text{if } v(B) = \mathbf{F} \text{ for every proposition } B \text{ in the range } \xi; \\ \mathbf{v}(A) = \mathbf{F} &\quad \text{otherwise.} \end{aligned}$$

[83]

Thus, a generalized (joint) negation is true only if all the negated propositions are true (6, 5.5ff.). A proposition C is a *logical consequence* of a class Γ of propositions if for every $v \in LS$ such that $\mathbf{v}(A) = \mathbf{T}$ for every $A \in \Gamma$ also $\mathbf{v}(C) = \mathbf{T}$ (5.11, 512).

Because every proposition is obtained through repeated applications of the **N**-operation to (suitably presented) ranges of propositions, the explanation determines fully whether a proposition is true or false at a point in logical space (5.501–3). The sense (*Sinn*) of the proposition A is a certain *Sachlage* σ_A in LS (4.021, 4.2).[84] The positive part of the sense of A is given by $\{ v \in LS \mid \mathbf{v}(A) = \mathbf{T} \}$, and similarly for the negative part, of course. The thesis of truth-functionality then ensures that the *Sachlage* σ_A , that is, the sense of A can be "computed" from the symbol A.

From this epitome it should be clear that the Tractarian logical theory is a realism of the kind that was inaugurated by Bolzano.[85] However, Wittgenstein carries the logical realism of Bolzano and Frege to a fitting conclusion: The *logical* realism of Bolzano is here replaced by an ontological realism. Propositional truth, the primitive *an sich* notion of logical realism, is reduced one step further to a prior ontological notion, namely, the obtaining of states of affairs. Neither Bolzano nor Frege ignored epistemological issues; in fact, they were of an all-encompassing importance for Frege's logicist project. Wittgenstein, on the other hand, deliberately eschews epistemic concerns in logic, for instance, the Frege–Russell assertion sign (4.442). Also the epistemic notion of inference is eliminated in favor of logical consequence by means of the Bolzano reduction (5.132).

Nevertheless, concerning the deployment of logic, Wittgenstein held that it must be possible to compute mechanically from the symbols alone whether one proposition follows from another (5.13, 6.126, 6.1262). He was wrong in this. In general, the "computation" cannot be executed, owing to its infinitary character. When he wrote the *Tractatus*, Wittgenstein was not aware of the unsolvability of the general *Entscheidungsproblem* for the predicate calculus. It was discovered—by Church and Turing—only in 1936, and poses an insuperable *technical* obstacle for the Tractarian philosophy of logic and language. Thus, Wittgenstein's vision that everything important concerning logic could be read off mechanically *am Symbol allein* was rendered illusory.

Wittgenstein was certainly aware of the fact that reasoning presupposes a correctness norm, because otherwise correct (*right*) and correct-for-me coincide, in which case there is no possibility for mistakes anymore. However, rather than taking rightness (*rectitudo*) of acts as a primitive notion, he adopts an ontological reduction of rightness. The order of explanation runs as follows: The rightness of the act of inference is reduced to the correctness of the product of the act, that is the judgment made, or knowledge obtained, which notion, in its turn, is reduced to the truth of the content. The truth of the content, finally, is reduced to an ontological notion, namely, that of the *obtaining* of states of affairs.[86] If objectivity is guaranteed at that level, say, in the form of bivalence for states of affairs—a state of affairs either obtains or does not obtain—it can be exported back to other levels, whence the possibility for mistakes is held open. In diagram form:

The truth-maker reduction in Wittgenstein's *Tractatus*

(4)
(2) {content of object} | ← act of knowledge
↓
S_C obtains ↔ [{Proposition C} is true]
↑ ↑
(1) state of affairs [object of the act] (3)
= [asserted statement, statement known].

From an epistemological point of view, the rightness notion for acts of knowledge is the most crucial one. It is enough to uphold the difference between appearance and reality, and, as such, constitutes the minimum requirement on a viable epistemology.[87]

The need for an ultimate correctness-norm for acts of knowledge, Wittgenstein certainly knew and accepted. Whereas I prefer to take it as primitive, Wittgenstein in the *Tractatus* reduces the rightness of the act to the correctness of the assertion made, and that in turn to the truth of the propositional content, which, finally, is reduced to the obtaining (and nonobtaining) of the corresponding state of affairs. Committed realists, when challenged, often reduce the norm of rightness one step further, from the notion of obtaining for states of affairs, to "reality itself," which accordingly has to provide for the obtaining and nonobtaining of states of affairs. When this reduction is coupled with the idea that "reality itself" is the sum total of all (material) objects and the wish to treat also reality itself as a material object, conceptual confusion results. However, without being a *transcendent* notion, reality cannot fulfill its required role as norm. It certainly cannot be subject to contingent facts the way material objects are, because such facts are responsible to the norm, whence it cannot be a material object. Wittgenstein had thought harder about these issues than most and such confusion is certainly avoided in the *Tractatus*: "Reality is the obtaining and non-obtaining of states of affairs" (2.06). On such a view, the notion of obtaining (and nonobtaining) of states of affairs can (pleaonastically) be reduced to reality itself. On the other had, "reality itself" thus construed is in no way less transcendent a notion than that of the obtaining of states of affairs.

7. Constructive Proofs of Propositions, the Traditional Form of Judgment Resurfaces

Moore and Russell continue, or perhaps rediscover, the realist stance in logic that had been advocated by Bolzano. Classical, bivalent logic is upheld for *an sich* bearers of truth—the "propositions"—either by means of a primitive sui generis notion of *an sich* truth (Bolzano, Frege) or by means of an ontological reduction of truth via a truth-maker analysis (Moore, Russell, Wittgenstein). One would not expect such metaphysical generosity concerning truth to come cheap. The currency in which the price has to be paid is, however, epistemological rather than metaphysical: Unknowable truths cannot be ruled out. The issue is by now a familiar one, owing to the works of Michael Dummett, who has challenged realist accounts of truth on meaning-theoretical grounds: Bivalent truth cannot serve as a key concept in an adequate theory of meaning, owing to the occurrence of propositions with undecidable truth-conditions.[88] However, Dummett was not the first to challenge unreflective realism. Already in the 1880s, the Berlin mathematician Leopold Kronecker and his pupils, among whom was Jules Molk, challenged the automatic use of realist logic:

> *Definitions should be algebraic and not merely logical.* It is not enough just to say: "something either is or is not." Being and non-being have to be set forth with respect to the particular domain within which we operate. Only in this way do we take a step forward. If we define, for instance, an irreducible function as a function that is not reducible, that is to say, that is not decomposable into other functions of a fixed kind, we do not give an algebraic definition at all, we only enunciate what is but a simple logical truth. *In Algebra*, for it to be rightful to give this definition, it must be preceded by the indication of a method that permits one, with the aid of finitely many rational operations, to obtain the factors of a reducible function. Such a method only confers an algebraic sense on the words *reducible* and *irreducible*.[89]

In other words, the following "definition" is not a permissible one:

$$f(x) =_{\text{def}} \begin{cases} 1 & \text{if the Riemann hypothesis is true;} \\ 0 & \text{if the Riemann hypothesis is false.} \end{cases}$$

When the definition is read classically (or "logically"), the function f is constant and therefore, trivially, a computable function. However, at the moment of writing, we are unable to compute the "computable" function in question. On the "logical" view, $f(14)$, say, is a natural number, but its numerical value cannot be ascertained. Definitions by means of undecided cases do not admit the effective substitution of *definiens* for *definiendum*. They contravene the canon for definitions that has been with us for three centuries, ever since

Pascal.[90] This is the price that a mathematician has to pay for unrestricted use of classical logic. His language will then contain *nonprimitive* terms that *cannot be eliminated* in favor of primitive vocabulary: Accordingly, there is no guarantee that meaning has been conferred on the terms in question.

The Kronecker criticism, in my opinion rightly, rules out definition of functions by means of undecidable separation of cases. Possibly a classical mathematician could live happily without these contrived functions. However, Dirichlet's famous definition of the function that is 1 on rational real numbers and 0 on irrational real numbers provoked a change in the conception of what a function is and can hardly be dismissed for want of mathematical interest. It also proceeds by an undecided separation of cases. Many proofs in classical analysis make use of this method. For instance, the standard "bisection of intervals" proof of the Bolzano–Weierstraß theorem that every bounded infinite set of real numbers has an accumulation point proceeds in exactly this fashion.[91] Again, these are mathematical matters and perhaps the classical *logician*, rather than the classical mathematician, need not be worried. Alas, this hope turns out to be forlorn: We only have to notice that Frege's explanation of the classical quantifier is cast in the form of an undecided separation of cases for matters to become more serious. Quantifier(phrase)s are function(expression)s that take (expressions for) propositional functions and yield (expressions for) propositions. Propositions, for the mature Frege, are ways of specifying truth values, and it seems advisable to make explicit also the relevant domain of quantification.[92] Accordingly, we consider a truth value valued function

$$A[x] \in \{\text{The True}, \text{The False}\}, \text{ provided that } x \in D.$$

Frege then defines the universal quantifier by means of the following explanation:

$$(\forall x \in D)A[x] =_{\text{def}} \begin{cases} \text{The True}, & \text{if } A[a/x] = \text{The True, provided } a \in D; \\ \text{The False}, & \text{otherwise.} \end{cases}$$

However, when the domain D is infinite, unsharp, or otherwise undecidable, the separation of cases cannot be carried out and the defined quantifier cannot be eliminated. Uncharitably put "the classical logician *literally* does not know what he is talking about." To my mind, this is the strongest way to marshal undecidability considerations against classical logic. The law of excluded middle is not the real issue.[93] Already *the classical rules of quantifier formation are unsound*: They do not guarantee that "propositions" formed accordingly actually do have content.

Until 1930, *content* was a very live issue. Work on the foundations of mathematics was dominated by the wish to secure a foundation for the practice of mathematical analysis after the ε-δ fashion of Weierstraß that satisfies the following conditions:

(**A**i) a formal system is given, in a syntactically precise way,

(**A**ii) with meaning explanations that endow the well-formed expression of its formal language with content,

(**A**iii) in such a way that its axioms and rules of inference are made evident, and

(**B**) classical logic is validated.

Frege's *GGA* was the first substantial attempt to meet the double desiderata of contentual formalization (**A**) *and* classical logic (**B**), but it foundered on the Zermelo–Russell paradox: Somewhere in Frege's §§29–31 there is an error, since otherwise every regular expression would have a *Bedeutung* and every derivable expression would be a name of the True. Whitehead and Russell also failed in their attempted *Principia Mathematica* execution of the foundationalist program: Their meaning explanations do not suffice to make evident the "Axioms" of Infinity, Choice, and Reducibility. Similarly, Wittgenstein's *Tractatus* provides (an attempt at) a semantic superstructure for the formal languages designed by Frege and Peano (as modified by Whitehead and Russell), as does the work of Frank Ramsey (1926). By 1930, faith in the project is waning: Carnap (1931) represents logicism's last stand.

The metamathematical Hilbert program (1926) was an attempt to secure the unlimited use of classical logic, *at the price of giving up content*, by means of an application of positivist philosophy of science to mathematics. The use of classical logic and impredicative methods are all fine as long as "the verifiable consequences," that is, those theorems that do have content, actually "check out."[94] Passing content by, this means that every free-variable equation between numerical terms that is derivable using also ideal axioms without content has to be correct, when read with content. Hilbert discovered that this holds if the ideal system is *consistent*, that is, does not derive, say, the formula $0 = 1$. In a way, this would have been an ideal approach to the foundations of mathematics for the working mathematician. The conceptual analysis required for foundational work, at which a mathematician does not necessarily excel, is replaced by a clear-cut *mathematical* issue, to be resolved by a (meta)mathematical proof, just like any other mathematical problem. Alas, it was too good to be true: With the appearance of Gödel (1931) all hope ended here, but the mathematical study of languages without content, which Hilbert had introduced in pursuit of a certain philosophical program, stayed on as an mathematical research program even when the philosophical position had collapsed.

Shortly after 1930, the first wave of (meta)mathematical results come in: Tarski and Lukasiewicz (1930), the already mentioned Gödel (1931), and Tarski (1933a, 1933b). Under the influence of these (meta)mathematical successes, even Carnap, the last logicist diehard, jettisons content and anything goes:

> Up to now, in constructing a language the procedure *has* usually *been*, first assign a meaning to the fundamental mathematico-logical

> symbols, and then to consider what sentences and inferences are seen to be logically correct in accordance with this meaning. . . .
>
> The connection will only become clear when approached from the opposite direction: let any postulates and rules be chosen arbitrarily; then this choice, whatever it may be, will determine what meaning is to be assigned to the fundamental logical symbols.[95]

The weakness of the position to which Carnap converted is obvious: If anything goes, what guarantee is there that content can be assigned? After all, there had been a few attempts at securing analysis already, meaning explanations and all, that had foundered on inconsistencies in the underlying formalisms. In such a calculus, demonstrably, no content can be had. Carnap's novel gospel is an extremely liberal one:

> *Principle of Tolerance*. It is not our business to set up prohibitions, but to arrive at conventions. . . .
>
> *In logic, there are no morals*. Everyone is at liberty to build up his own logic . . . as he wishes.[96]

A quarter of a century earlier, at the same time when, in Cambridge, Russell and Moore bit the bullet of unknowable truths, Carnapian licentiousness was rejected on the other side of the North Sea in the (1907) doctoral dissertation of a young Amsterdam mathematician who took over the torch of mathematical constructivism from Kronecker. L. E. J. ("Bertus") Brouwer (1881–1966) claimed that language use was responsible to the mathematical deed of construction and not the other way round:

> In the edifice of mathematical thought thus erected, language plays no part other than that of an efficient, but never infallible or exact, technique for memorizing mathematical constructions, and for communicating them to others so that mathematical language by itself can never create new mathematical systems. But because of the highly logical nature of mathematical language the following question naturally presents itself. *Suppose that, in mathematical language, trying to deal with an intuitionist mathematical operation, the figure of an application of one of the principles of classical logic is, for once, blindly formulated. Does this figure of language then accompany an actual languageless mathematical procedure in the actual mathematical system concerned?*[97]

In particular, laws of whatever theoretical logic have no validity on their own, but have to be applied in such a fashion that they do ensure proper content. A year after his thesis, Brouwer reaches the conclusion that the law of excluded middle cannot guarantee that the required deed of construction can be executed, whence it has to be rejected not as false but as unfounded.[98] Thus he refrains from asserting that "$A \vee \neg A$ is true."[99] Also the method of proofs by means of *nonconstructive dilemma* that proceeds by obtaining the conclusion that C is true from both the assumption that A is true as well as from the opposite assumption that A is false and concludes that C is true, is rejected, as is the method of *indirect* (or *apagogic*) *proof*, when one assumes a negative claim, obtains a contradiction and concludes a positive claim from this contradiction. Reductio ad absurdum proofs, on the other hand, are perfectly acceptable to constructivists: In these one proves a *negative* claim from a positive assumption that yields a contradiction. Here a method is provided for obtaining a contradiction from an assumption that constitutes a construction for the negation.

Mere formulation or postulation does not automatically confer validity on the rules in question. Formulation alone is not enough to secure preservation of content at the level of the mathematical deed of construction. In this, surprisingly enough, Brouwer resembles Frege who, at roughly the same time, severely criticized formalist accounts of mathematics for their lack of content.[100] For Frege, however, it was a commonplace that the contents expressed by declaratives have to be *bivalent* propositions, *tertium non datur*. Frege hoped to secure this by making the bond between propositions and truth values a tight one: A proposition is a means of presenting a truth value. Owing to lack of effectiveness in some of the chosen means of presentation, for example, quantification with respect to an infinite domain via an undecidable separation of cases, an operational want of content is the result. Accordingly, Brouwer, as well as other mathematical constructivists who insist on the constructional deed in mathematics, will have to provide for another notion of proposition than that of (a mode of presentation of) a truth value, if the formal logical calculi shall not be void of content. This Brouwer did only by precept in his mathematical work: With a lifelong love-hate relationship to language, he never took to formalization and the emerging symbolic calculi of logic.[101]

It was left to others, to wit Hermann Weyl, one of few first-rate mathematicians with a sympathy for intuitionism, and Brouwer's pupil Arend Heyting, to formulate the required notions explicitly. Brouwer's style of exposition in his intuitionistic writings was not to everybody's taste and Weyl, who deftly wielded a polemical pen, took over the early propaganda work, at which he excelled. From his study at Göttingen, Weyl had firsthand knowledge of Husserl's phenomenology, and this influence can be seen in his writings around 1920.[102] It was left to him, possibly drawing on work of Schlick and Pfänder, to formulate explicitly the required notion of constructive existence to be applied in a constructive truth-maker analysis:

> *An existential proposition*—for instance, "there is an even number" —*is not at all a proper judgement that expresses a state of affairs*; existential states of affairs are an empty invention of logicians. "2 is an even number": that is a real judgement that expresses a state of affairs; "there is an even number" is only a judgement-abstract that has been obtained from this judgement.[103]

Here we have a novel form of judgment, namely,

$$\alpha \text{ exists},$$

where α is a general concept. Its assertion condition is given by the rule

$$\frac{a \text{ is an } \alpha}{\alpha \text{ exists}},$$

whence one is entitled to assert that α exists only if one already knows an α.[104]

The contribution of Heyting is twofold. First, he gave an explicit formulation of the proper intuitionistic rules of logic.[105] Second, he intervened decisively in a confused debate whether logic according to intuitionists would need a third truth value: true, false, and undefined, thereby leading to a law of the "excluded fourth," and so on.[106] In his intervention Heyting formulated explicitly a constructivist notion of proposition that admits of a truth-maker analysis:

> A proposition p, for example, "Euler's constant is rational" expresses a problem, or better still, a certain expectation (that of finding two integers a and b such that $C = a/b$) that may be realized or disappointed.[107]

Here the intuitionistic novelty is introduced: proofs of propositions, that is, judgable contents, rather than judgments. All previous proving in the history of logic and mathematics had been at the level of judgment and not at that of their contents. These proofs of propositions are not epistemic but ontological in character; inspection of the examples given by Heyting and Brouwer reveals that they are common or garden mathematical objects: functions, ordered pairs, and so on. A proposition A is given by a certain set Proof(A) of proof-objects for the proposition in question. Many alternative formulations have been offered:

Proposition	Proof	Heyting (1934)
Intention Expectation	Fulfillment	Heyting (1930), (1931)
Problem	Solution	Heyting (1930), Kolmogorov (1932)
Type	Object	Howard (1980)
Set	Element	Martin-Löf (1982)
Specification	Program	Martin-Löf (1982)

The explanation of the standard logical constants then take the following form:

$\bot$ There are no proofs for $\bot$.

& When a is a proof for A and b is a proof for B, $\langle a, b\rangle$ is a proof for $A \mathbin{\&} B$.

$\vee$ When a is a proof for A, $i(a)$ is a proof for $A \vee B$.
When b is a proof for B, $j(a)$ is a proof for $A \vee B$.

$\supset$ When b is a proof for B, provided x is a proof for A, $\lambda x.b$ is a proof for $A \supset B$.

$\forall$ When D is a set, P is a proposition, when $x \in D$, and b is a proof for P, when $x \in D$, $\lambda x.b$ is a proof for $(\forall x \in D)P$.

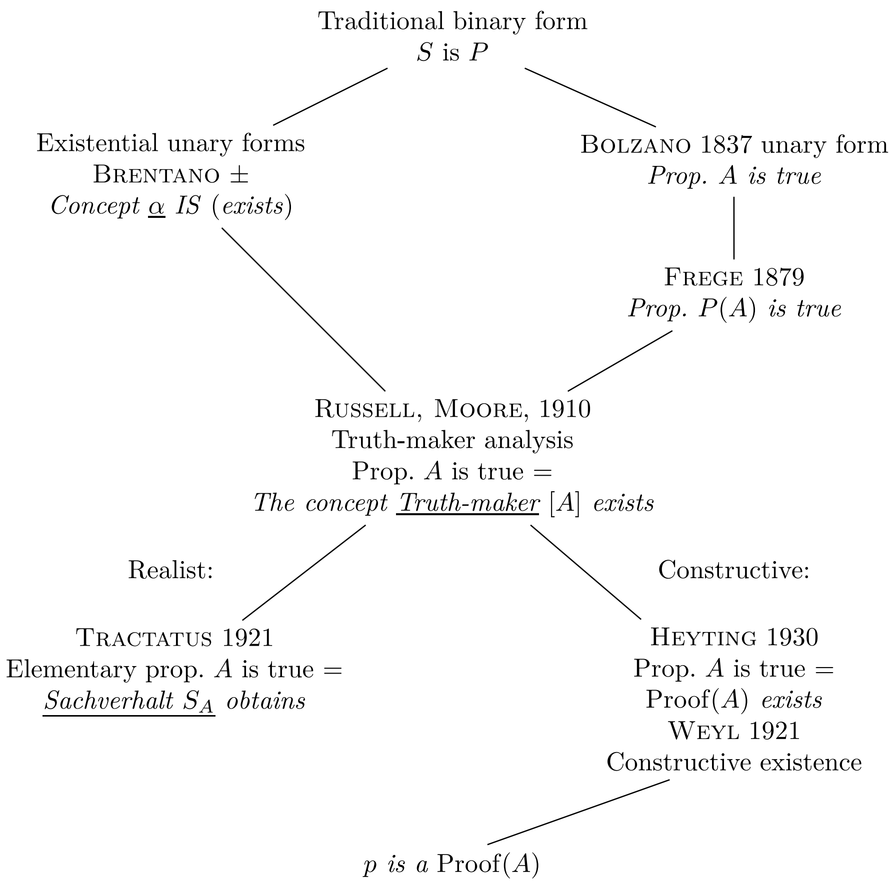
$\exists$ When D is a set, $a \in D$, P is a proposition, when $x \in D$, and b is a proof for $P[a/x]$, $\langle a, b \rangle$ is a proof for $(\exists x \in D)P$.[108]

The constructivist truth-maker analysis then takes the form

$$\text{proposition } A \text{ is true} = \text{Proof}(A) \text{ exists},$$

where the notion of existence is the constructive (Brouwer–)Weyl existence already explained.[109] The wheel has come full circle: A judgment made that ascribes truth to a proposition is elliptic for another judgment in the fully explicit form: a is a Proof(A), which is nothing but a judgment of the traditional form: S is P.[110]

Transformation of the form of judgment



8. Inference versus Consequence: How Gentzen Had It Both Ways

The interpreted formal systems of Frege, Whitehead and Russell, and Heyting were all axiomatic. These systems (are meant to) have an intended interpretation in terms of the respective meaning explanations. In such systems, a formal derivation is or can be read as a proof that shows that its conclusion formula, when read according to its interpretation, does express a truth. In the modern metamathematical systems of propositional and predicate logic, on the other hand, the end-formula has *no* intended interpretation, but has to be true under any truth value assignment or set-theoretic interpretation, respectively. Frege, furthermore, explicitly held that one can only draw inferences from *known* premises. This claim has been controverted, most famously by Gentzen, who created another kind of formalism in his 1933 Göttingen dissertation.[111] The derivable objects are still formulae, but may depend on assumptions, and several rules serve to discharge open assumptions. A derivation takes the general form:

$$(D) \qquad \begin{array}{c} A_1, \quad A_2, \quad \ldots, \quad A_k \\ \vdots \\ C \end{array}$$

where $A_1, \ldots, A_k$ are the undischarged assumption on which the end-formula C depends. The rules of inference are divided into two groups of introduction and elimination rules. The conjunction introduction rule (&I), say, allows you to proceed to the conclusion $A \,\&\, B$, given two derivations of A and B, respectively, that depend on open assumptions in the lists Γ and Δ, respectively. The derivation of $A \,\&\, B$ depends on open assumptions in the joint list Γ, Δ. The rule (&E) of conjunction elimination, on the other hand, allows you to obtain the conclusion A from the premise $A \,\&\, B$, and also the conclusion B from the same premise, while the open assumptions remain unchanged. The rule (⊃I) of implication introduction allows one to proceed to $A \supset B$ from the premise B that has been derived from assumption formulae in the list Γ, while discharging as many premises of the form A as one wants—one, many, or none. The derivation of $A \supset B$ depends on assumptions in the list Γ_1, where Γ_1 coincides with Γ, except possibly for some deleted occurrences of the assumption formula A. The system is convenient to work with when one actually has to find the derivations in question.

Michael Dummett put the case for Gentzen's natural deduction as follows:

> Frege's account of inference allows no place for $a[n]$ … act of supposition. Gentzen later had the highly successful idea of formalizing inference so as to leave a place for the introduction of hypotheses.

Indeed, "it can be said of Gentzen that it was he who showed how proof theory should be done."[112] However, Dummett's comparison between Frege and Gentzen is not entirely fair, since it does not take the metamathematical paradigm shift into account. For Frege, the formal system was a tool in the epistemological analysis of mathematics: it was actually used for for proving theorems. For Gentzen, (meta)mathematician, or *Beweistheoretiker*, the formal system was Hilbertian, that is, an object of study, without content, *about which one proves* (meta)mathematical *theorems*, such as, for instance, his famous (1936, 1938) consistency theorem by means of ε_0-induction. For a fair comparison, the respective formal systems of Frege and Gentzen accordingly have to be placed on an equal footing: We either divest Frege systems of their content and treat them as if they were metamathematical, or we supply meaning explanations for the key notions in Gentzen's systems, so as to endow its object "language" with content. The present chapter is devoted to the notion of judgment, and an inference is nothing but a judgment of a particular (mediate) kind. However, without content no judgment, so it is to the second of these alternatives that we have to turn. Our task is to give a reformulation, call it *Gentzen*, of Gentzen, at the same level of interpretation as that provided by Frege. The early stages of the conversion present no difficulties: It is clear that the wff's in the formal language, say, of first-order arithmetic, can be interpreted as propositions. The syntactic terms are readily turned into numerical expressions, and the predicates $<$ and $=$ obviously lend themselves for interpretation as the computable numerical relations *less than* and *identity*, respectively. So far so good; with respect to elementary syntax and semantics, Frege and *Gentzen* march in step.

The difficulties arise when we turn to the pragmatic dimension that is involved in Frege's use of the turnstile as an assertion sign, that is, as an explicit force indicator. Gentzen does not use a turnstile, but if he had it would undoubtedly have been used as a Kleene–Rosser theorem predicate; Gentzen was a (meta)mathematician. Here we see a first difficulty for *Gentzen*: Gentzen (and with him other metamathematicans) used his wff's in two roles. Wff's are fed to connectives, that is, Frege's *Gedankengefüge*, to build other, more complex wff's: Accordingly, for *Gentzen* they are propositions. On the other hand, Gentzen also used wff's as end formulae of derivation trees: Accordingly, for *Gentzen*, the wff's also have to be turned into theorems, that is, assertions (judgments made) that propositions are true. Here *Gentzen* confronts a potentially damaging ambiguity. However, we must allow *him* the same leeway as that offered to Frege: He can make use of the turnstile as an assertion sign, and also other force indicators, should he want to do so. The obvious option for *Gentzen* is to use *two* force indicators, one for *assertion* ($\vdash$) and another for *assumption* ($\dashv$). Finally, *Gentzen* also has to interpret the derivation trees of Gentzen. The Gentzen derivation D will be interpreted by means of the following procedure:

i. Append "is true" to each wff that stands on its own at a node in D (rather than as part of another wff);

ii. prefix transformed wff's "A is true" of assumption formulae by the assumption sign "$\dashv$" and prefix transformed "A is true" of conclusion formulae by the assertion sign "$\vdash$";

iii. interpret all wff's in D as propositions.

The result is the tree D'':

$$(D'') \qquad \begin{array}{c} \dashv A_1 \text{ is true}, \quad \dashv A_2 \text{ is true}, \quad \ldots, \quad \dashv A_k \text{ is true}, \\ \cdot \qquad\qquad\qquad \cdot \\ \cdot \qquad\quad \cdot \\ \cdot \\ \vdash C \text{ is true}. \end{array}$$

We may have some hope that D'' will serve as a flow-chart for a proof-act that yields the knowledge that proposition C is true. However, this simpleminded approach does not work: The interaction of the two kinds of force—assumption and assertion—is more involved. Consideration of an example, in which proof-theoretical experts will recognize one of Dag Prawitz's reduction figures, makes this clear:[113]

$$(d) \qquad \dfrac{\dfrac{\begin{array}{c}[A]\\ \vert \\ B\end{array}}{A \supset B} \quad \begin{array}{c}\vert\\ A\end{array}}{B}$$

This tree is dressed according to the procedure and transformed into the tree

$$(d') \qquad \dfrac{\dfrac{\begin{array}{c}\dashv A \text{ is true (1)}\\ \vert \\ B \text{ is true (2)}\end{array}}{\vdash A \supset B \text{ is true (3)}} \quad \begin{array}{c}\vert\\ \vdash A \text{ is true (4)}\end{array}}{\vdash B \text{ is true (5)}}$$

The force apparatus is almost equal to its task: the proposition A occurs as part of an assertion (4), of an assumption (1), and as an unasserted part of an assertion (3). The notation is rich enough to distinguish these cases clearly. With respect to the proposition B matters are less fortunate, though. For the proposition B, assertion (5) and unasserted part (3) are coped with, in the same way as for the proposition A. The premise that B is true of the ($\supset$I) rule (2), however, is neither assumed nor asserted, and its force cannot be expressed with the two force indicators at hand. There one asserts that the proposition B is true, *provided* that the proposition A is true.[114] One must note, though, that it is not the assertion that is hypothetical or conditional; the assertion is categorical, whereas the notion of truth has been made conditional. We no longer ascribe outright truth to the proposition B, but only the constrained notion

$$\ldots \text{ is true, provided that } A \text{ is true.}$$

Thus we have an unconditional, categorical assertion that conditional truth pertains to the proposition B. Strictly speaking, this is a novel form of judgment. The derivation tree D' above, where the assumptions

$$A_1 \text{ is true, } A_2 \text{ is true, } \ldots, A_k \text{ is true}$$

are still open, or *undischarged*, does not allow for the ascription of outright truth to the proposition C, but only of truth on condition that A_1 is true, A_2 is true, ..., A_k is true. The general case of the weakened, *conditional* truth in question will then be:

$$\ldots \text{ is true } (A_1 \text{ is true, } A_2 \text{ is true, } \ldots, A_k \text{ is true}).$$

Accordingly nodes in derivation trees are not covered with statements of the form

$$A \text{ is true,}$$

but with statements of the conditional form. Effecting this transformation, the derivation tree D ultimately takes the form D''':

$$\begin{array}{lc} & A_1 \text{ is true } (A_1 \text{ is true}),\ A_2 \text{ is true } (A_2 \text{ is true}),\ \ldots,\ A_k \text{ is true } (A_k \text{ is true}) \\ (D''') & \vdots \\ & C \text{ is true } (A_1 \text{ is true, } A_2 \text{ is true, } \ldots, A_k \text{ is true}). \end{array}$$

The relevant notion of assertion is still categorical, but the truth that is asserted of various proposition may be weakened. We must distinguish between the two statements:

i. proposition $A \supset B$ is true,

ii. B is true, provided that A is true, or its (synonymous) variant,

ii′. if A is true, then B is true.

The statement (i) is explained classically via truth-making of atomic propositions and then inductively via the truth tables, say, and constructively in terms of an assertion-condition demanding a (canonical) proof-object, as in section 7.

From the constructive point of view, an assertion of the final statement in D''', that is,

$$(*) \qquad C \text{ is true } (A_1 \text{ is true, } \ldots, A_k \text{ is true}),$$

demands a dependent proof-object:

(∗∗) c is a proof of C, provided that
x_1 is a proof of A_1, …, x_k is a proof of A_k.

Accordingly, the conditional statement (∗) represents a novel form of judgment, with the assertion condition (∗∗). This suggests how natural deduction derivations should be interpreted: They are notations for *dependent* proof-objects.

Gentzen did not have only one format for natural deduction derivations but two. Sometimes they are considered as mere notational variants.[115] In the present context their differences are significant. In 1936 he used a *sequential* format for the derivations.[116] The derivable objects are no longer well-formed formulae, but *sequents.* A sequent

$$A_1, A_2, \ldots, A_k \Rightarrow C$$

lists all the open assumptions on which C depends. Derivations have no assumptions, but *axioms* only of the form

$$A \Rightarrow A,$$

with the *Gentzen* interpretation

$$A \text{ is true, provided that } A \text{ is true,}$$

indeed, something undeniably correct, albeit not very enlightening. Consideration of the tree D''' shows that its top formulae are axioms of this kind and that the conditional statements at the nodes in the tree are nothing but *sequents in another notation.*

Because there are no acts of assumption, no discharge of assumptions takes place, but antecedent (assumption-)formulae can get struck out; for instance, the rule (⊃I) takes the form

$$\frac{A, \Gamma \Rightarrow B}{\Gamma \Rightarrow A \supset B}.$$

Conjunction introduction (&I) will be

$$\frac{\Gamma \Rightarrow A \quad \Delta \Rightarrow B}{\Gamma, \Delta \Rightarrow A \,\&\, B}.$$

On the *Gentzen* interpretation the sequent $A_1, \ldots, A_k \Rightarrow C$ is interpreted as

$$C \text{ is true, on condition that } A_1 \text{ is true, } \ldots, A_k \text{ is true,}$$

and the *Gentzen* sequent should properly include the truth ascriptions:

$$A_1 \text{ is true, } \ldots, A_k \text{ is true} \Rightarrow C \text{ is true.}$$

Derivations in the sequential format of natural deduction describe, or are, blueprints for proof-acts that certain propositions are conditionally true. Properly speaking, we have here a treatment of *consequence relations* among propositions: The consequent proposition is true when the antecedents are all true. One should also note that the statement

$$\text{proposition } C \text{ is true}$$

is a special case of the sequent, when the number k of antecedent propositions $= 0$.

The sequents can also be read as closed sequents

$$(A_1, \ldots, A_k) \Rightarrow C.$$[117]

Just as the *Gentzen* sequents represent a novel form of judgment, so do these closed sequents, and their *Gentzen* interpretation should be

$$\text{the sequent } (A_1, \ldots, A_k) \Rightarrow C \text{ holds.}$$

To have the right to assert that a closed sequent holds we must give a verifying object. This is a function f that takes proofs $a_1, \ldots, a_k$ of the antecedent propositions into a proof

$$f(a_1, \ldots, a_k) \text{ of the consequent proposition } C.$$

We must distinguish between three equiassertible statements:

the proposition $A \supset B$ is true
(demands a proof of $A \supset B$);

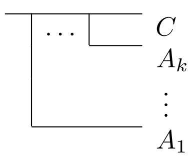
the conditional statement (open sequent) A true $\Rightarrow B$ true
(demands a dependent proof b of B, provided that x is a proof of A);

the closed sequent $(A) \Rightarrow B$ holds
(demands a function from Proof(A) into Proof(B)).

The assertion condition is different in all three cases, but one can be met only if the other two can also be met. Furthermore, one reason that these notions are not always kept apart is that all three are refutable by the same counterinstance, namely, a proof-object a of A and a dependent proof-object c for the open sequent B true $\Rightarrow \bot$ true.

With this distinction, my treatment of the sequent calculus comes to an end. At the level of assertion, there is apparently little to choose between Gentzen and Frege.

It only seems fair to let Frege have the last word: The repeated Frege conditional



is commonly read as the implication

$$(A_1 \supset (A_2 \supset (\cdots \supset (A_k \supset C) \cdots))).$$

However, rotating the Frege conditional 90° clockwise, while altering the notation only slightly, produces another familiar result as the late Pavel Tichý (1988, pp. 248–252) observed, namely

$$A_1, A_2, \ldots, A_k \Rightarrow C.$$ [118]

The correspondence between the calculi of Frege and Gentzen operates even with respect to the fine structure of the rules, sometimes even exhibiting a surprising(?) resemblance of terminology. Thus, the following question acquires some urgency: Did Gentzen read Frege's *Grundgesetze* prior to 1933, the year in which his dissertation was composed?[119]

Be that as it may. Bolzano gave us a coherent theory of (logical) consequence between propositions. Frege was more right about inference from judgments made to judgment than he is given credit for. However, only in Gentzen's sequential natural deduction do we have a theory that treats of both consequence as well as inference.

Brief Biographical Notes

1. Bernard Bolzano, 1781–1848[120] A Bohemian priest of Italian origin, who held the chair of Philosophy of Religion at the Charles University in Prague from 1805 until 1820, when he was summarily dismissed, as well as barred from public teaching and preaching, for holding too liberal views concerning matters both spiritual and temporal, gave fundamental contributions to mathematical analysis ("Bolzano–Weierstraß theorem"). A wholly admirable man, he led a retiring life with friends in the Bohemian countryside, devoting himself to logical and mathematical researches. The magnitude of Bolzano's contribution to logical theory, as well as to philosophy in general, can hardly be overestimated. Being censored, it was left unrecognised, thereby retarding logical progress by half a century. Appreciation is mounting with the growing volume of the *Gesamtausgabe*, and Bolzano might yet receive the credit that is so amply his due: "the greatest philosopher of the nineteenth century, bar none."[121]

2. Franz Brentano, 1838–1917 Brentano belonged to a prominent German cultural family. Ordained a priest, he held, after impressive Aristotle studies, a (Catholic) extraordinary chair in Philosophy at Würzburg. Misgivings over the definition of Papal Infallibility in 1870 led him to renounce the priesthood and change his chair for one in Vienna, where his lectures aquired cult status as society happenings. The Concordat between Austria and the Vatican allegedly forbade Austrian ex-priests to marry and led him to resign also this chair and his Austrian citizenship in 1880; having taken Saxon citizenship he then married, confidently expecting reappoinment. This never happened, reputedly

at the personal instigation of the emperor; after 15 years as *Privatdozent*, the former Professor Brentano left Austria and settled in Florence. Blindness darkened his last decade, to which belong important brief essays on truth as well as a wide-ranging correspondence with utterly devoted pupils. Italy's entry into World War I forced a final move to Switzerland, where Brentano died in 1917. Brentano, an outstandingly successful lecturer and supervisor, though not devoid of dictatorial leanings, had highly able doctoral students, who bitterly disappointed him by not speaking in unison with their master's voice. Nevertheless, his influence ranges wide, not only among devoted Brentanists, but also in two major schools of twentieth-century philosophy, to wit, the Lvow-Warsaw school under Twardowski, the first to introduce analytical techniques, and Husserl's phenomenology.

3. Gottlob Frege, 1848–1925 A German mathematician at Jena who taught (mainly) analytical geometry several hours a week, never reached the rank of *Ordinarius* but gave fundamental contributions to the foundations of logic and mathematics. In the *Begriffsschrift* and *Die Grundlagen der Arithmetik* the program of reducing arithmetic to logic, as well as the logic to which it was to be reduced, are set out with great lucidity. His attempt at a fully rigorous execution of his foundationalist program, in the *Grundgesetze der Arithmetik*, proved to be irredeemably flawed owing to the emergence of the Zermelo–Russell paradox within the system. Three important essays from the early 1890s provide a philosophical underpinning for the *Grundgesetze*. Of these, *Über Sinn und Bedeutung* is commonly regarded as the origin of modern philosophy of language. Frege founded no school, and, for a long time was only known through and for his influence on major figures such as Russell, Carnap, and Wittgenstein. A deeply conservative man in matters cultural and political, Frege died forgotten in the Weimar Republic to which he could not relate. His contributions to logic, its philosophy, and the philosophies of mathematics and language are now recognized in their own right, and not only as an influence on others, whereby Frege rightly emerges as a major thinker of the nineteenth century.

4. Ludwig Wittgenstein, 1889–1951 The youngest son of Karl Wittgenstein, a main architect of the Industrial Revolution in Austria, as well as one the wealthiest men in Europe, was educated at the *Oberrealshule* in Linz, where Adolf Hitler was a fellow pupil, and subsequently at the *Technische Hochschule*, Berlin-Charlottenburg, and Manchester University, prior to settling at Cambridge, where his work on the foundations of logic ripened in close contact with Bertrand Russell, in relation to whom Wittgenstein went through the stages of pupil, co-worker, and implacable critic. At the outbreak of World War I, Wittgenstein volunteered for the Austrian army, and during the war he refined and deepened his views on logic that were published in the aphoristic *Logisch-Philosophische Abhandlung*, now universally known as *Tractatus* (*Logico-Philosophicus*). After the Great War, Wittgenstein gave away his in-

herited fortune and became a schoolteacher in the hinterlands of lower Austria. A Vienna lecture by L. E. J. Brouwer, March 1928, rekindled his interest in philosophy, and led him to return to England. Keynes wrote to his wife: "God is in England. I met him at the 5.15 train." Elected a fellow of Trinity, and from 1939 professor in succession to G. E. Moore, Wittgenstein developed an entirely novel conception of philosophy, on which he published nothing. A stern man, more unsparing of himself than of others, Wittgenstein died in 1951, his last words being: "Tell them I have had a wonderful life."

Notes

1. Barnes (1988, p. 48). I am indebted to my Leyden colleague Dr. J. van Ophuijsen for drawing my attention to this marvelous passage.

2. (1950), p. vii.

3. Putnam (1982), Boolos (1994). Boolos also canvasses 1858, the year in which Dedekind cut the rationals, as a candidate.

4. I have argued as much, in and out of print, since 1988 (p. 4).

5. Thus texts would comprise three main parts, or "books": *Of Terms*, *Of Judgement*, and *Of Reasoning*, and possibly a final part treating *Of Method*. Kant's *Jäsche Logik* is a good case in point.

6. The diagram draws on a similar one in Maritain (1946, p. 6) but is reasonably standard. Maritain's source, and also that of virtually all other neo-Thomists, is the splendid *Ars Logica* by John of St. Thomas.

7. Fichte (1797) (which bears the title *Wissenschaftslehre*). The convenient representation of the act/object distinction was introduced in Martin-Löf (1987).

8. A committed anti-antirealist, or the unbiased reader, might prefer the less pejorative *realist* for the other alternative.

9. Here the vertical bars above the judgments $J_1, J_2, \ldots, J_k$ represent acts that yield, respectively, the judgment in question.

10. *KdrV*, A69. "Wir können alle Handlungen des Verstandes auf Urteile zurückführen, so daß der Verstand überhaupt als ein Vermögen zu urteilen vorgestellt werden kann."

11. And by Heidegger, or so I have been told.

12. As does *judgment*: act of judging versus judgment made.

13. Recent scholarly tradition associates this familiar example with Tarski (1944). It is, nevertheless, considerably older than so. We find it in Boole (1854, p. 52), as well as in Hilbert and Ackermann (1928, p. 4) (who undoubtedly have it from Boole). The latter is cited by Tarski in *Der Wahrheitsbegriff*. However, the ultimate source for the present-day logical obsession with arctic meteorology might well be Aristotle's *Prior Analytics*, Book A, ch. iv, where we find a discussion of things—among them *snow*—that admit the predication of "white."

14. Here the letter S serves as a schematic letter for declaratives.

15. For lack of space, in what is after all an inquiry into the recent history of *logic*, I must here leave well-known (*spät*) Wittgensteinian claims to the contrary without due consideration. I simply register my conviction that they do not present an insurmountable obstacle, and that my view as given in the text is essentially correct.

16. Sometimes a single utterance of the nominalization will, nevertheless, effect the required assertion, for example, when responding to the question: "Which of the two alternatives is the true one?"

17. The apt term judgable content we owe to Frege's *Bs.* His *beurtheilbarer Inhalt* has been variously rendered into English as (i) possible content of judgment (Geach), (ii) content that can be judged (Van Heijenoort), (iii) judicable content (Jourdain), and (iv) judgable content (Dummett). Of these, the last deserves preference over the second for the sake of brevity, while the first is likely to cause serious confusion, owing to its pointing in the direction of modal logic.

18. This is not the full story. The right to ask for grounds, when faced with an assertion made, shows that there is an implicit claim to knowledge contained in the assertoric force with which the sentence has been uttered, and which sometimes comes to the fore, for instance, in Moore's paradoxical assertion of

It is raining but I do not believe it.

Thus,

I know that snow is white,

which is in the *performative* first person, might have been a more felicitous form to use for the assertion made by means of my assertoric utterance of the declarative "Snow is white," were it not for the fact that it is prone to be conflated with the *third-person use* of the first person, which, when applied to me, is synonymous with

Göran Sundholm knows that snow is white.

19. This *linguistic turn in philosophy* was so named by Gustav Bergmann (1964, p. 177). The term gained wide currency after Richard Rorty (1967) chose it for his title.

20. *Weighty* in every sense of the word; its four volumes add up to a total of close to 2500 pages.

21. Kant, *KdrV*, B VIII (my translation).

22. The Kantian authority for Jäsche's text is not undisputed, see Boswell (1988).

23. The English rendering of Bolzano's *Satz an sich* is a matter of some delicacy. The modern, Moore-Russell notion of proposition, being an English counterpart of the Fregean Thought (German *Gedanke*), really is an *an sich* notion, and, for our purposes, essentially the same as Bolzano's *Satz an sich.* Thus, *proposition-in-itself* is pleonastic: The in-itself component is already included in the proposition. Furthermore, the mental propositions and their linguistic signs, that is, written or spoken propositions, as explained, carry assertoric force, whereas Bolzano's *Sätze an sich* manifestly do not, serving, as they do, in the role of judgmental *content.* Accordingly, it might be better to use *Sentence in itself*, which does not seem to carry the presumption of assertoric force. However, as Ockham and other medieval thinkers noted, the *propositio mentalis*, and its matching exterior signs, can be further analyzed into *propositio judicationis*, which does carry assertoric force, and *propositio apprehensionis*, which does not. Ockham has *Quodlibetal Questions* with congenial titles: Questio V:vi "Is an act of apprehending really distinct from an act of judging?" and Q iv:16 "Does every act of assenting presuppose an act of apprehending with respect to the same object?" So from this point of view, Bolzano's proposition-in-itself is obtained by severing the (mental) links that tie the *propositio apprehensionis* to its mental origin and its linguistic signs.

24. WL, §34.

25. Occasionally I shall permit myself to drop the "in-itself" idiom in the interest of perspicuity and readability and speak just of "propositions."

26. Contrary to a common misapprehension, Frege, who employed the term *third realm* in 1918, is not its progenitor. It was in general use in neo-Kantian circles. Gottfried Gabriel's lemma *Reich, drittes* in the *Historisches Wörterbuch der Philosophie* tells the full story.

27. In some theories, for instance, that of Frege, propositions (thoughts) are explained as the meanings of declarative sentences. This is not Bolzano's way of proceeding, the sui generis, absolutely mind- and language-independent propositions-in-themselves, are there in their own right, so to say, and they are capable of fulfilling various logical offices, among them that of serving as sentence meanings.

28. The reader will please note long shadows being cast forward toward Tarski (1944). Bolzano (§28) has a discussion of whether *really* is really necessary in the right side of this explanation.

29. WL, §154 (4). "Auch leuchtete jedem die Ähnlichkeit ein, die zwischen diesem Verhältnisse unter den Sätzen und zwischen jenem, welches ... unter *Vorstellungen* ... obwaltet. Was nähmlich bei Vorstellungen den Umstand gilt, ob ein gewisser Gegenstand durch sie in der Tat vorgestellt werde, das gilt bei Sätzen der Umstand, ob ihnen Wahrheit zukomme oder nicht."

30. When snow is white, the idea-in-itself *the whiteness of snow* is instantiated and the idea *the blackness of snow* is not.

31. Sebestik (1992) offers a beautiful *précis* of Bolzano's framework.

32. WL, §138. The categorical judgments of the A and I forms and O are treated of, respectively, at §225 Anmerkung, and §171. The treatment of an O judgment (Some α is not β) follows the pattern of the I judgments: The idea-in-itself of an α that is not β has *Gegenständlichkeit*.

33. The falsity of false propositions-in-themselves is also *an sich*.

34. WL, §34. The status of my chosen form of judgment [A is true] is very delicate with respect to Bolzano's system. On the one hand, it has to be a proposition-in-itself, since the iteration of ... is *true* is the key step in Bolzano's non-apagogic "proof" that there are infinitely many true propositions-in-themselves (WL, §32). On the other hand, propositions-in-themselves are supposed to be *sentence meanings*, as well as the bearers of truth and falsity, as is clearly documented by the following passage (cited from Mark Textor [1996], p. 10) in Bolzano's *Von der mathematischen Lehrart* (my translation): "not what grammarians call a proposition, namely the linguistic expression, but rather the *sense* of this expression, that must always be only one of true or false, is for me a *proposition in itself* or an *objective* proposition." The sense (*Sinn*) of the declarative sentence (grammatical proposition) "Snow is white" is *that snow is white*. Furthermore, that-clauses are what yield grammatical declarative sentences when saturated with "is true" or "is false." Accordingly, that-clauses seem to be the appropriate linguistic counterparts to propositions-in-themselves. But then, the ascription of truth to a proposition-in-itself is not a proposition-in-itself, since the declarative "the proposition-in-itself A is true" isn't a that-clause, that is, does not have the required form for being (the linguistic counterpart of) a proposition-in-itself. This note attempts to answer Wolfgang Künne, who objected, after my Cracow lecture, that for Bolzano, [A is true] is just another proposition-in-itself, but not a judgment. In spite of the considerations, on balance, I am inclined to think that

this might be an instance where I let my systematic preferences override historical subtleties, and that some injustice is done to Bolzano.

35. Franz Brentano, for whom the problem of the blind judgment became a major issue, and whose views will be considered in the sequel, propagated this apt terminology. As Per Martin-Löf has pointed out to me, the notion and *term* might ultimately derive back to Plato, *The Republic*, 506c: "opinions divorced from knowledge, are ugly things[.] The best of them are blind. Or do you think that those who hold some correct opinion without evidence differ appreciably from blind men who go the right way?"

36. WL, §36. Fairness bids me to report that Bolzano was aware of a certain awkwardness in his doctrine at this point. In WL, §314, with the telling title "Are there Definite Limits to Our Capacity for Knowledge?," he notes:

> For since every judgment that agrees with the truth is a piece of knowledge, even if that agreement is only accidental and had come about only by way of previous errors, it can very well be seen that the limits of our capacity for knowledge, if we were able to abide by such a broad definition, would fluctuate everywhere, since mere chance and even a mistake could contribute to its enlargement.

In this connection we should further note that according to Bolzano, every truth is knowable, since God knows the truth of every true proposition-in-itself, whence it can be known: *ab esse ad posse valet illatio.*

37. In much the same way, Bolzano profited immeasurably from having Kant, the foremost idealist of the age, as his main target. That can be seen by comparing the pristine clarity of Bolzano's work with the murkiness of early Moore and Russell 60 years later. Their realism was the result of an apostasy from and battle with a much inferior version of idealism, namely the British Hegelianism of Bradley, Green, and Bosanquet. Another example of the same phenomenon is provided by Wittgenstein, who, according to Geach (1977, vi), held Frege's *Der Gedanke* in low esteem: "it attacked idealism on its weak side, whereas a worthwhile criticism of idealism would attack it just where it was strongest." The destructive side of a philosophical position seems to gain in quality with the target it attacks.

38. Apparently Bolzano was unable to give a material criterion for what it is to be a (non)logical idea, but then so were his successors, who offered virtually identical accounts of logical truth and consequence a century later.

39. The foregoing brief formulations do not do perfect justice to Bolzano on a number of scores. (i) Logical consequence (in the modern sense) is a two-place relation between antecedent and succedent propositions, whereas *Ableitbarkeit*, be it logical or not, is a *three*-place relation between antecedent proposition(s), consequent proposition(s), and idea(s) (that occur in at least one antecedent or consequent proposition), where the ideas indicate the places where the variation takes place. Logical *Ableitbarkeit* considers variation with respect to the collection of *all* nonlogical ideas that occur in the antecedent and consequent propositions. As a limiting case (possibly one rejected by Bolzano) one might consider merely material *Ableitbarkeit* that consists in the preservation of truth under variation with respect to no ideas, and which holds between A_1, A_2, ..., A_k and C, when the implication

$$(A_1 \mathbin{\&} A_2 \mathbin{\&} \cdots \mathbin{\&} A_k) \supset C$$

is just true, but not necessarily *logically* true. (ii) With respect to the consequence (or *sequent* in the terminology of Hertz and Gentzen)

$$A_1, A_2, \ldots, A_k \Rightarrow B_1, B_2, \ldots, B_m,$$

Bolzano demands that the A's and the B's be compatible, that is, there must be some suitable variation that makes them all true. Furthermore (iii), modern theory holds that the multiple-succedent sequent should be read as

$$A_1 \mathbin{\&} A_2 \mathbin{\&} \cdots \mathbin{\&} A_k \supset B_1 \vee B_2 \vee \cdots \vee B_m;$$

the sequent is valid when this implication is a logical truth. Bolzano, however, uses another meaning for the sequent, namely,

$$A_1 \mathbin{\&} A_2 \mathbin{\&} \cdots \mathbin{\&} A_k \supset B_1 \mathbin{\&} B_2 \mathbin{\&} \cdots \mathbin{\&} B_m;$$

according to Bolzano an *Ableitbarkeit* with many succedent propositions holds when every variation that makes all antecedent propositions true also makes *every* (and not just at least one) succedent proposition true.

In Siebel (1996) Bolzano's theory of *Ableitbarkeit* is studied in depth and related (with due consideration for significant differences) to a number of well-known modern topics, such as Russell's theory of propositional functions, the Quine-Ajdukiewicz-Tarski account of logical truth and consequence, and the relevance logic of Anderson and Belnap. Nevertheless, in spite of the sometimes considerable differences, it is proper to regard Bolzano as the founder of the modern theory of (logical) consequence among propositions; he is the first to reduce the validity of inference (from judgment to judgment) to a matching relation among propositions (-in-themselves) that serve as contents of the relevant premise and conclusion judgments, respectively.

40. In *WL*, part III (*"Erkenntnislehre"*), ch. II (*"Von den Urtheilen"*), §300 ("mediation of a judgement through other judgements"), Bolzano considers also inferences proper, that is, mediate acts of judgments, and not only their Platonist simulacra, namely, consequence relations (*Ableitbarkeiten*) among the respective judgmental contents. Lack of space prevents me from developing this theme any further.

41. A true proposition-in-itself can stand in the relation of *Abfolge* to more than one grounding proposition.

42. WL, §220. See Aristotle, *An. Post.*, I:13.

43. Validity of an inference figure must be distinguished from that of validity (rightness) of an act of inference. An act of inference, that is, a mediate act of judgment, is valid (right, real, or true) if its axioms, that is, according to Frege's GLA, §3, p. 4, characterization, judgments neither capable of nor in need of demonstration, really are correct, and the inference-figures employed therein really are valid, that is, do preserve knowability.

44. Compare, for instance, the fragments reprinted in part III and appendix 2 of Brentano (1930), with titles such as "Against so-called Judgmental Contents" and "On the Origin of the Erroneous Doctrine of the *entia irrealia*."

45. Furthermore, these reductions were well known already to Leibniz, for instance in the *General Inquisitions*, §§146–151. Franz Schmidt provides a list of 28 (!) different Leibnizian reductions of the four categorical judgments in Leibniz (1960, pp. 524–529). For instance, the singular affirmative judgment "Some A are B" is rendered alternatively as "AB is," "AB is a thing," and "AB has existence," by reductions 17,

22, and 27. Here Bolzano's German *hat Gegenständlichkeit* is matched by Leibniz's Latin *est* or *est Ens*.

46. Letter from Brentano to Hugo Bergmann, June 1, 1909, quoted from Bergmann (1968, pp. 307–308) (my translation).

> Wenn ich ... auf Bolzano aufmerksam machte, so geschah dies, ..., keineswegs, um den jungen Leuten Bolzano als Lehrer und Führer zu empfehlen. Was sie von ihm, das dürfte ich mich sagen, konnten sie besser von mir lernen. ...
>
> Und wie gesagt, wie ich selbst von Bolzano nie auch nur einen einzigen Satz entnommen habe, so habe ich auch niemals meinen Schülern glaubhaft gemacht, daß sie dort eine wahre Bereicherung ihrer philosophischen Erkenntnis gewinnen würden.

47. Brentano (1889, Anm. 27, pp. 64–72) and the fragments in Brentano (1930, part IV) are important here. Also relevant is his *Versuch über die Erkenntnis*, that is, Brentano's (1903) attempt at an *Essay* on human knowledge after the fashion of Locke and Leibniz. For instance, its first part bears the grandiloquent title: DESTROY PREJUDICE! AN APPEAL TO THE CONTEMPORARY AGE, THAT IT FREE ITSELF, IN THE SPIRIT OF BACON AND DESCARTES, FROM ALL BLIND APRIORI.

48. Note that this use of the term *evidence* is different from its use within current analytical philosophy of science and the Anglo-Saxon common-law legal systems. ("My lord, I beg leave to enter exhibit 4 into evidence.") There one is concerned with supporting evidence *for* a claim. Brentano's use is concerned with that which is evident (known). Evidence is the quality that pertains to what is evident.

49. Brentano (1930, p. 139).

50. Brentano (1956, p. 175, §39). According to the editor, this negative formulation of the Law of Excluded Third derives from an unpublished fragment "Über unsere Axiome" from 16.2.1916.

51. Scholz (1930) remains the standard treatment of Aristotelian foundationalism.

52. Frege (1879) is a book with the title *Begriffsschrift*, whereas "begriffsschrift" is an English (loan-)word for the eponymous formal language developed in that work. This is not an ideal solution to the title/notion ambiguity of the German term. Using either of the two standard English renderings—*ideography* and *concept*(*ual*) *notation*—seems a worse option, though.

53. This is how it *ought* to be; regarding Frege, his *GGA* axiom 5 concerning *Werthverläufe* and the use of classical second-order impredicative quantification remain unjustified. In place of *notions* we could speak of *terms* or *concepts* here. Either choice runs the risk of being taken in too narrow a sense, though. Today a *term* is a syntactic entity only, often associated with the formation rules of first-order predicate calculus, whereas for Frege a *concept* is confined to a certain kind of function.

54. The first formulation combines passages from (1879, pp. 2, 4): "Der Umstand, dass..." and "...ist eine Thatsache." Apparently the form of judgment is (*...ist eine Thatsache*), where the blank has to be filled with an "*Umstand*." Accordingly the form of the judgment made through an assertoric utterance of "Snow is white" is:

The circumstance that snow is white is a fact

In place of circumstance, Frege also allows for "*Satz*." In (1918, p. 74, n. 8) he identifies *fact* with *true Thought*, which yields the final reformulation:

The proposition that snow is white is true.

The second formulation is taken from 1892 (p. 34, n. 7):

> Ein Urteil ist mir nicht das bloße Fassen eines Gedankens, sondern die Anerkennung seiner Wahrheit.

55. Frege (1880/81, pp. 36 ff.) attempts to sell his begriffsschrift to the mathematicians by treating of the standard notions pertaining to continuity, but to no avail, alas. The mathematicians, and among them to his shame Felix Klein, did not rise to the occasion and the paper was rejected by, for instance, the *Mathematische Annalen*. In desperation, Frege sought refuge with the philosophers at the *Zeitschrift für Philosophie und philosophische Kritik*, but they proved equally cold-hearted. In spite of it being unpublished, Frege's piece must be give full evidentiary value since it was written for publication and repeatedly submitted. Later Frege established very good relations with the *Zeitschrift* where some of this very best papers appeared, in 1882 and 1892.

56. That is, iterated combinations of "for all/there is" and "there is/for all." The passage from continuity to uniform continuity provides a clear example of the shift from "$\forall\exists$" to "$\exists\forall$". The (linear) logical notation employed here is reasonably standard, using inverted A and E for *Alle* (all) and *Es gibt* (there is). It is due to Gerhard Gentzen (1934–35), but derives in essence from Peano, via mediation through Whitehead. Frege's own begriffsschrift is *two*-dimensional and has great versatility, as well as a strange beauty of its own. It was never able, however, to gain proselytes, and so it perished with its progenitor in the early stages of the mounting metamathematical revolution in the late 1920s.

57. Frege's checkered struggle toward an identity criterion for propositions (his Thoughts) is long and fascinating; see Sundholm (1994c).

58. Frege (1892, p. 34): "Ich verstehe unter dem Wahrheitswerte eines Satzes den Umstand, daß er wahr ist oder daß er falsch ist."

59. Frege's notion of a proper name (*Eigenname*), following the German translation of John Stuart Mill's *System of Logic*, comprises not just grammatical proper names but singular terms in general.

60. *GGA*, I, §32. Note that this formulation admit the equation of proposition (Thought) with truth-condition,

> the Thought that snow is white = the Thought expressed in "snow is white" = the truth-condition of "snow is white,"

whence for a declarative sentence S:

> S = that S is true = the truth-condition of "S" is fulfilled = the proposition expressed in "S" is true.

Thus also,

> the Thought that S = the thought that the truth-condition of "S" is fulfilled.

61. Frege (1906b, II, p. 387): "eine Urteilsfällung, die auf Grund schon früher gefällter Urteile nach logischen Gesetzen vollzogen wird."

62. *Bs*, §6, and *GGA*, I, §14. The change from Frege's two-dimensional notation to a one-dimensional Gentzen notation is not always anodyne, but here, where the concern is assertoric force, rather than the specific contents, it seems innocent enough.

63. The ins and outs of Frege's assertion sign are treated very well in Stepanians (1998, chs. 1–5). Concerning the origin of its use as a theorem predicate, see Kleene (1952, p. 88, p. 526).

64. Dummett (1973, p. 309, p. 435). I beg to differ and will return to the issue in section 8.

65. *GGA*, I, preface, pp. xv–xvi:

> Wahrsein ist etwas anderes als Fürwahrgehalten werden, sei es von Einem, sei es von Vielen, sei es von Allen, und es ist in keiner Weise darauf zurückzuführen. Es ist kein Widerspruch, dass etwas wahr ist, was von Allen für falsch gehalten wird.

This marvelous credo is embedded in a passage pp. xv–xvii that is highly germane to the realism issue.

66. *Nachlass*, p. 2. (The *Logik* of the 80s): "Was wahr ist, ist unabhängig von unser Anerkennung. Wir können irren." It is not required that there be mistaken acts of knowledge, but only that their possibility is not ruled out.

67. *GGA*, II, p. 69:

> Das Gesetz des ausgeschlossenen Dritten ist ja eigentlich nur in anderer Form die Forderung, dass der Begriff scharf begrenzt sei. Ein beliebiger Gegenstand Δ fällt entweder unter den Begriff Φ, oder er fällt nich unter ihn: *tertium non datur*.

68. The crucial primacy of the *sui generis* notion of *rightness* was noted by Martin-Löf (1987, 1991). In the light of this, Sundholm (2004) spells out various interrelations between different roles of truth.

69. *Locus classicus*: "Der Gedanke" (1918).

70. Bs, preface, p. IX:

> Die festeste Beweisführung ist offenbar der rein logische, welche, von der besonderen Beschaffenheit der Dinge absehend, sich allein auf die Gesetze gründet, auf denen alle Erkenntnis beruht. Wir theilen danach alle Wahrheiten, die einer Begründung bedürfen, in zwei Arten, idem der Beweis bei den einen rein logisch vorgehen kann, bei den andern sich auf Erfahrungsthatsachen stützen muss.

GLA, §§3–4, contains a further elaboration of this theme into an account of the distinctions analytic/synthetic, a priori/a posteriori. Frege's considerations here, successively stepping from a known truth to its grounds seeking the ultimate laws of justifications, are strongly reminiscent of Bolzano's use of his grounding trees with respect to *Abfolge*.

71. Moore (1898, 1902) and Russell (1903, appendix A, §477, 1904). Cartwright (1987) treats of their early theory in some depth. The notion of proposition is here essentially the same as in Bolzano, and, to some extent, Frege. The grave responsibility for mistranslating the Fregean *Gedanke* (Thought) into *proposition* rests on Moore and Russell: Moore (1898, p. 179) introduced the terminology: "We have approached

the nature of the *proposition or judgment*. A proposition is composed not of words, nor yet of thoughts, but of concepts." Russell (1903, appendix A, §477) completes the error by coupling Frege's *Gedanke* with his own proposition. Through *PM* this misidentification eventually became standard throughout all of modern logic. In its original sense from the tradition, a proposition was either the (mental) judgment made, or its outward announcement in language, whereas after Russell and Moore it is turned into the *content* of a proposition in the original sense.

72. Joachim (1906, p. 37), who apparently wrote this marvelous passage in ignorance of Bolzano, drawing only on what he could find in Russell and Moore, for instance:

> [There] is no problem at all in truth or falsehood; that some propositions are true and some false, just as some roses are red and some white; that belief is a certain attitude towards propositions, which is called knowledge when they are true, error when they are false. (Russell 1904, p. 523)

Note how Russell adheres to the Bolzano reduction of knowledge to the mere truth of its content. Wittgenstein (*Tractatus* 6.111) also took notice of this passage from Russell.

73. Joachim (1906, p. 39). Russell and Moore both responded to Joachim's book in *Mind*. Moore's response is particularly interesting: "That *some* facts are facts, and *some* truths true, which never have been, are not now, and never will be experiences *at all*, and which are not timelessly expressed either" (1907, p. 231). What Moore countenances here are propositions that will remain unknown at all times; that, though, does not make them unknowable. The opposite view presupposes what Lovejoy (1936) called the principle of *Plenitude*, namely, that all potentialities will eventually become actual. (Martin-Löf 1991 rejects the application of Plenitude to knowability: what is knowable need never be known.) Only a year later did Moore commit himself in a review of William James: "It seems to me, then, that very often we have true ideas which we cannot verify; true ideas, which in all probability no man will ever be able to verify" (Moore 1907–08, p. 103).

74. Russell (1903, p. 145). (I am indebted to Prof. Peter Hylton for drawing my attention to this passage.) Since infinitely complex propositions have to be unknowable, one way of deciding the issue concerning their existence is to deny that there are unknowable propositions. Because Russell is unable to pronounce on the issue, this means that he does not want to rule out unknowable truths. Also this passage was noticed by Wittgenstein, see *Tractatus* 4.2211. Russell's is the earliest position (known to me) that allows for unknowable truths. Frege, as we saw, rejects them; every truth either is knowable in itself or has a *Begründung*, that is, a proof.

75. The undefinability of truth was claimed in print only in *Der Gedanke* (1918, p. 60). In *Nachlass*, p. 140 (*Logik* 1897), Frege had made the same points almost verbatim. They in turn go back *in nuce* to the *Logik* of the 80s.

76. Moore in the lecture course from 1910–11 that was published later (1953). Russell in a number of places, for example, 1912 (p. 74) and PM, p. 43.

77. The notion was explicitly formulated by Mulligan, Simons, and Smith (1984).

78. The classic Stenius (1960) remains eminently readable. Hacker (1981) offers the best presentation of the theory and its difficulties.

79. It most certainly is not; Peter and John, assuming they are empirical subjects, are complexes (5.541–5.421) composed of thoughts (3), that is, picture-facts (2.16),

and will be analyzed in terms of the propositions that describe the complexes in question (3.24). The *transcendent* subject (5.63–5.641), on the other hand, "thinks out" the sentence-senses, which constitute the method of projection to the world (3.11–3.13), whereas the empirical subject is composed of sentential signs, that is, thought-facts.

80. The definition of the Q-relation reminds one of the ways that Frege formed unsaturated expressions. It is clear, I think, that this is one of very many places where the influence of "the great works" of Frege (see the preface to the *Tractatus*) can be felt.

81. In the next few paragraphs I use expository devices from the metatheory of the propostional calculus to survey the logico-semantical doctrines of the *Tractatus*. References to the *Tractatus* are by thesis number. Enderton (1972) contains the relevant model theory.

82. Wittgenstein's notion of proposition (*Satz*) is not that of Bolzano–Frege–Russell (*Satz an sich*/*Gedanke*/*proposition*). In the *Tractatus* a proposition is a meaningful sentence in use and the *Sätze* might well better be rendered *sentences* in English translation.

83. Here **N** is Wittgenstein's generalized Sheffer-stroke that negates every member of the range ξ of propositions.

84. The Fregean proposition is a sense, whereas the Tractarian proposition (sentence) has sense.

85. Jan Sebestik (1990) suggests that Robert Zimmerman's *Gymnasium* textbook *Philosophische Propädeutik*, which is replete to the point of plagiarism with material taken from Bolzano's *Wissenschaftslehre*, might be the missing link between Bolzano and Wittgenstein.

86. After the metamathematical revolution around 1930, Wittgenstein's ontological notion, *obtaining of the states of affairs makes the elementary proposition true*, is transformed into the model-theoretic:

$$A \models \varphi,$$

that is, the set-theoretical structure A satisfies the wff φ (Tarski and Vaught 1957). See also Sundholm (1994b).

87. The crucial epistemological role of rightness in upholding the distinction between appearance and reality was noted and stressed by Martin-Löf (1987).

88. Dummett (1976) is the *locus classicus*, while Dummett (1991) offers a book-length treatment. The secondary literature on Dummett's argument has reached the proportions of an avalanche. Sundholm (1986) is an early survey, and Sundholm (1994a) approaches Dummett's position from a more severely constructivist standpoint.

89. Molk (1885, p. 8):

> *Les définitions devront être algébraiques et non pas logiques seulement.* Il ne suffit pas de dire: "Une chose est ou et non pas." Il faut montrer ce ques veut dire être et ne pas être, dans le domaine particulier dans lequel nous nous mouvons. Alors, seulement nouns faisons un pas en avant. Si nous définissions, par exemple, une fonction irréductible comme une fonction qui n'est pas réductible, c'est a dire quie n'est pas décomposable en d'autres fonctions d'une nature déterminée, nous ne donnons point de définition algébraique, nous n'énonçons qu'une simple vérité logique.

> Pour qu'*en Algèbre*, nous soyons en droit de donner cette définition, il faut qu'elle soit précédé de l'exposé d'une méthode nous permettant d'obtenir a l'aide d'un nombre fini d'operations rationelles, les facteurs d'une fonction réductible. Seule cette méthode donne aux mots *réductible* et *irréductible*.

90. See the *Port-Royal Logic*, Arnauld and Nicole (1662, part IV, ch. III).

91. The ∞ set $D \subseteq \Re$, being bounded, is contained in a closed real interval I. Define

$$I_0 =_{\text{def}} I =_{\text{def}} [a_0, b_0]$$

$$I_{k+1} =_{\text{def}} [a_k, a_k + b_k/2] \quad \text{if this left half of } I_k \text{ contains } \infty \text{ many points from } D;$$

$$I_{k+1} =_{\text{def}} [a_k + b_k/2, b_k] \quad \text{otherwise.}$$

(NB. Here we cannot decide whether a half has ∞ many points from D.)

Hence, each of the nested intervals I_k contains ∞ many points from D, and length $(I_k) \to 0$, when $k \to \infty$. Thus, $\bigcap_k I_k$ contains exactly one point that is the required accumulation point for D.

92. Frege does not include the set D, the domain of quantification, since he quantifies over all individual objects.

93. Martin-Löf (1983, p. 33) hints at this way of understanding Brouwer's criticism. It was noted explicitly by Aarne Ranta (1994, p. 38). See also Sundholm (1998), where also Poincaré's criticism of impredicability is cast in the same mold.

The law of excluded middle does not only serve as a principle of reasoning. It is also used meaning-theoretically to delimit the notion of proposition. Thus, for Frege, a proposition is a method for determining one of the truth values True and False. Similarly, every proposition implies itself and something which is not a proposition implies nothing, Russell (1903, §16) notes, and goes on to use "$P \supset P$" as an explanation of what it is for something P to be a proposition. But an assertion that $P \supset Q$ is true is equivalent to is an assertion that P is false or Q is true. Thus an assertion that P is a proposition amounts to an assertion that P is false or P is true. The issue resurfaces in the *Cambridge Letter* R 12 from Wittgenstein to Russell, June 1913, where " '*aRb.v.∼aRb*' must follow directly *without the use of any other premiss*." Also Cantor's explanation of a well-defined set (1882, p. 114) makes meaning-theoretical use of the law of excluded middle.

94. Appropriately enough, free-variable equations between computable terms, with only true numerical substitution instances, are called *verifizierbar* (verifiable) in the canonical exposition Hilbert and Bernays (1934, p. 237).

95. Carnap (1934, p. xv).

96. Carnap (1934, pp. 51–52).

97. Brouwer (1981, p. 5). This formulation, albeit late, expresses Brouwer's lifelong view.

98. Brouwer (1908).

99. One does not, of course, claim that $A \vee \neg A$ is false, that is, that $\neg(A \vee \neg A)$ is true, because the latter claim is refutable outright: Assume that $\neg(A \vee \neg A)$ is true. Assume further that A is true. Under this assumption, $A \vee \neg A$ is also true. Therefore, the assumption that A is true leads to a contradiction. Therefore, A is false, now only under the sole assumption that $\neg(A \vee \neg A)$ is true. Hence $\neg A$ is true,

still under the same assumption. But then, under the same assumption, also $A \vee \neg A$ is true. Thus the assumption that $\neg(A \vee \neg A)$ is true leads to the conclusion that also $A \vee \neg A$ is true, which is a contradiction. Therefore the assumption is wrong and $\neg(A \vee \neg A)$ is false. Thus, $\neg\neg(A \vee \neg A)$ is true.

100. *GGA*, II, §§87–147, as well as his undignified diatribes (1906a) and (1906b), (1908), against Korselt and Thomae, respectively. The need for content in mathematical sign-languages is a theme that Frege pursued from his earliest writings; see for instance, the long *Nachlass* paper on Boole's logic and his own begriffsschrift (1880/81) and above all (1882).

101. Even after World War II—Brouwer lectured regularly at Cambridge from 1946 to 1951—he would proclaim, apparently with a deadpan face, that "Absurdity of absurdity of absurdity is equivalent to absurdity," rather than use the pellucid $\neg\neg\neg A \leftrightarrow \neg A$. See Brouwer (1981, p. 12).

102. Most clearly perhaps in the introduction to Weyl (1918a), but also in the treatment of logic in (1918b).

103. Weyl (1921, p. 54):

> *Ein Existentialsatz*—etwa "es gibt eine gerade Zahl"—*ist überhaupt kein Urteil im eigentlichen Sinne, das einen Sachverhalt behauptet*; Existential-Sachverhalte sind eine leere Erfindung der Logiker. "2 ist eine gerade Zahl": das ist ein wirkliches, einem Sachverhalt Ausdruck gebendes Urteil; "es gibt eine gerade Zahl" ist nur ein aus diesem Urteil gewonnenes *Urteilsabstrakt.*

104. The novel form of judgment and the explicit formulation of the rule that provides its assertion-condition are both due to Per Martin-Löf (1994).

105. Heyting (1930a).

106. The debate in question is treated in Thiel (1988) and Franchella (1994).

107. Heyting (1930b, p. 958):

> Une proposition p, comme, par example, "la constante d'Euler est rationelle"; exprime un problème, ou mieux encore une certaine attente (celle de trouver deux entiers a et b tels que $C = a/b$), qui pourra être réalisée ou déçue.

108. This table is based on a streamlined formulation offered by Per Martin-Löf (1984), and, in each case, lays down what a *canonical* proof-object is for the proposition in question. For the significance of *canonical* in this context, see Sundholm (1997), where a full exposition of the intuitionistic meaning explanations is offered.

109. It should be stressed that these meaning explanations for the logical constants, and the ensuing truth-definition, are neutral with respect to the underlying logic; in fact the framework can be viewed as a Tarskian truth definition—another neutral account. If we allow nonconstructive existence claims, also classical logic holds under the proof-object semantics.

We have to show, reasoning nonconstructively, that $\text{Proof}(A \vee \neg A) \neq \emptyset$.

Assume that $\text{Proof}(A) \neq \emptyset$. Let $a \in \text{Proof}(A)$; then $i(a) \in \text{Proof}(A \vee \neg A)$.

Assume that $\text{Proof}(A) = \emptyset$. Then $\lambda x.x \in \text{Proof}(A) \to \text{Proof}(\bot) = \text{Proof}(\neg A)$, and so $j(\lambda x.x) \in \text{Proof}(A \vee \neg A)$.

Hence, in either case, $\text{Proof}(A \vee \neg A) \neq \emptyset$, so the proposition $A \vee \neg A$ is true. Q.E.D.

110. In Martin-Löf's constructive type theory (1982, 1984) the elliptic form of judgment A *is true* is replaced by the explicit p *is a* Proof(A); Martin-Löf (1983) makes clear that this constitutes a return to the traditional S *is* P form of judgment.

111. Published as Gentzen (1934–35).

112. Dummett (1973, p. 309, and p. 435, respectively).

113. Prawitz (1965, p. 37). Prawitz prefers the opposite order between the two premises of the ($\supset$I) rule, but this is of no importance for the present point.

114. Provided that, given that, on condition that, under the assumption that, under the hypothesis that.... Many variations in the wording are possible here.

115. For instance by Prawitz (1971, remark 1.6, p. 243), Dummett (1977, pp. 121–122, and 1991, p. 248), as well as Sundholm (1983).

116. The *sequential form of natural deduction* uses both introduction rules and elimination rules. It must not be confused with the *sequent calculus* of Gentzen (1934–35) that uses no elimination rules but has both left and right *introduction* rules, on both sides of the sequent arrow.

117. Gentzen did not consider closed sequents; the exploration of their theory is due to Peter Schroeder-Heister (1981, 1984, 1987), half a century after Gentzen.

118. Kutschera (1996) and Schroeder-Heister (1999) both discuss the matter in apparent unawareness of Tichý's explicit treatment. Tichý's remarkable chapter 13—*Inference*—merits attention, as does his paper "On Inference" (1999).

119. The *Übersicht* (1934–35, p. 176) does mention Frege, Russell, and Hilbert as particularly important for the formalization of logical inference, but the remark does not presuppose familiarity with the details of Frege's formalization. The introduction to Hilbert and Ackermann (1928, p. 2), which Gentzen did know, makes similar mention of the same authors.

120. Full biographies are available for a number of authors treated of in the present chapter: Frege (Kreiser 2001), Wittgenstein (McGuinness 1988; Monk 1990), Brouwer (Van Dalen 1999), and Gentzen (Menzler-Trott 2001).

121. Simons (1999, p. 115).

References

Arnauld, Antoine, and Pierre Nicole. 1662. *La Logique ou L'Art de Penser*.

Barnes, Jonathan. 1988. Meaning, Saying, and Thinking. In *Dialektiker und Stoiker*, eds. Klaus Döring and Theodor Ebert, 47–61. Stuttgart: Franz Steiner Verlag.

Bergmann, Gustav. 1964. *Logic and Reality.* Madison: University of Wisconsin Press.

Bergmann, S. H. 1968. Bolzano and Bertano. *Archief für Geschichte der Philosophie* 48: 306–311.

Bolzano, Bernard. [1837] 1929–1931. *Wissenschaftslehre in vier Bänden*, hrsg. von W. Schultz. Leipzig: Verlag von Felix Meiner. (*WL*)

Boole, George. 1854. *The Laws of Thought.* London: Macmillan.

Boolos, George. 1994. 1879? In *Reading Putnam*, eds. Bob Hale and Peter Clark, 31–48. Oxford: Basil Blackwell.

Boswell, Terry. 1988. On the textual authenticity of Kant's logic. *History and Philosophy of Logic* 9: 193–203.

Brentano, Franz. 1889. *Vom Ursprung sittlichher Erkenntnis* (*Philosophische Bibliothek* 55). Hamburg: Felix Meiner, 1955.

Brentano, Franz. 1903. *Versuch über die Erkenntnis* (*Philosophische Bibliothek* 194). Leipzig: Felix Meiner, 1925.

Brentano, Franz. 1930. *Wahrheit und Evidenz* (*Philosophische Bibliothek* 201). Hamburg: Felix Meiner, 1974.

Brentano, Franz. 1956. *Lehre vom richtigen Urteil*, ed. F. Mayer-Hillebrand. Bern: Francke Verlag.

Brouwer, L. E. J. 1981. *Cambridge Lectures on Intuitionism*, ed. Dirk van Dalen. Cambridge: Cambridge University Press.

Brouwer, Luitzen Egbertus Jan. 1908. De onbetrouwbaarheid der logische pricipes. *Tijdschrift voor wijsbegeerte* 2: 152–158.

Cantor, Georg. 1882. Über unendliche lineare Punktmannigfaltigkeiten, III. *Mathematische Annalen* 20: 113–121.

Carnap, Rudolf. 1931. Die logizistische Grundlegung der Mathematik. *Erkenntnis* 2: 91–105. English translation, Logicist Foundations of Mathematics. In *Philosophy of Mathematics*, eds. Paul Benacerraf and Hilary Putnam, 31–41. Englewood Cliffs, N.J.: Prentice Hall, 1964.

Carnap, Rudolf. 1934. *Logische Syntax der Sprache*. Wien: Springer. Cited after the English translation by the Countess von Zeppelin, *Logical Syntax of Language*, London: Routledge and Kegan Paul, 1937.

Cartwright, Richard. 1987. A Neglected Theory of Truth. In *Philosophical Essays*, 71–93. Cambridge, Mass.: MIT Press.

Dummett, M. A. E. 1977. *Elements of Intuitionism*. Oxford: Oxford University Press.

Dummett, Michael. 1973. *Frege. Philosophy of Language*. London: Duckworth.

Dummett, Michael. 1976. What Is a Theory of Meaning? (II). In *Truth and Meaning*, eds. Gareth Evans and John McDowell, 67–137. Oxford: Clarendon Press.

Dummett, Michael. 1991. *The Logical Basis of Metaphysics*. London: Duckworth.

Enderton, H. B. 1972. *A Mathematical Introduction to Logic*. New York: Academic Press.

Fichte, Johann Gottlieb. 1797. Erste Einleitung in die Wissenschaftslehre. *Philosophisches Journal* 5: 1–47.

Franchella, M. 1994. Heyting's contribution to the change in research into the foundations of mathematics. *History and Philosophy of Logic* 15: 149–172.

Frege, Gottlob. 1879. *Begriffsschrift*. Halle: Louis Nebert. (*Bs*)

Frege, Gottlob. 1880/81. Boole's rechnende Logik und die Begriffsschrift. In Frege (1983), 9–52.

Frege, Gottlob. 1882. Über die wissenschaftliche Berechtigung einer Begriffsschrift. *Zeitschrift für Philosophie und philosophische Kritik* 81: 48–56.

Frege, Gottlob. 1884. *Die Grundlagen der Arithmetik*. Breslau: Koebner. (*GLA*)

Frege, Gottlob. 1891. *Funktion und Begriff*. Jena: Hermann Pohle.

Frege, Gottlob. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100: 25–50.

Frege, Gottlob. 1906a. Antwort auf die Ferienplauderei des Herrn Thomae. *Jahresbericht der deutschen Mathematiker-Vereinigung* 15: 586–590.

Frege, Gottlob. 1906b. Grundlagen der Geometrie I, II, III. *Jahresbericht der deutschen Mathematiker-Vereinigung* 15: 293–309, 377–403, 423–430.

Frege, Gottlob. 1908. Die Unmöglichkeit der Thomaeschen formalen Arithmetik aufs neue nachgewiesen. *Jahresbericht der deutschen Mathematiker-Vereinigung* 17: 52–55. And *Schlußbemerkung* on p. 56.

Frege, Gottlob. 1918. Der Gedanke. *Beiträge zur Philosophie des deut/-schen Idealismus* 1: 58–77.

Frege, Gottlob. 1983. *Nachgelassene Schriften*, eds. Hans Hermes, Friedrich Kambartel, and Friedrich Kaulbach. Hamburg: Felix Meiner.

Geach, Peter. 1977. Preface. In Gottlob Frege, *Logical Investigations*, vii–ix. Oxford: Basil Blackwell.

Gentzen, Gerhard. 1934–35. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift* 39: 176–210, 405–431.

Gentzen, Gerhard. 1936. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen* 112: 493–565.

Gentzen, Gerhard. 1938. Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie. In *Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften* (N.F.) 4, 19–44. Leipzig: Felix Meiner.

Gödel, Kurt. 1931. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik* 38: 173–198.

Hacker, P. M. S. 1981. The Rise and Fall of the Picture Theory. In *Perspectives on the Philosophy of Wittgenstein*, ed. Irving Block, 85–109. Oxford: Blackwell.

Heyting, Arend. 1930a. Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der preussischen Akademie von Wissenschaften* Phys.–math. Klasse, 42–56. English translation in Mancosu (1998), 311–327.

Heyting, Arend. 1930b. Sur la logique intuitionniste. *Acad. Roy. Belgique, Bull. Cl. Sci.* 5, 16: 957–963. English translation in Mancosu (1998), 306–310.

Heyting, Arend. 1931. Die intuitionistische Grundlegung der Mathematik. *Erkenntnis* 2: 106–115. English translation in P. Benacerraf and H. Putnam, *Philosophy of Mathematics* (2nd ed.), Cambridge: Cambridge University Press, 1983, 52–61.

Heyting, Arend. 1934. *Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie* (Ergebnisse der Mathematik, vol. 3:4). Berlin: Springer.

Hilbert, David. 1926. Über das Unendliche. *Mathematische Annalen* 95: 161–190.

Hilbert, David and Paul Bernays. 1934. *Grundlagen der Mathematik*, vol. I. Berlin: Springer.

Hilbert, David, with W. Ackermann. 1928. *Grundzüge der theoretischen Logik*. Berlin: Springer.

Howard, W. A. 1980. The Formulae-as-Types Notion of Construction. In *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, eds. J. Seldin and R. Hindley, 479–490. London: Academic Press.

Joachim, H. H. 1906. *The Nature of Truth*. Oxford: Clarendon Press.

Kant, Immanuel. [1781, 1987] 1904. *Kritik der reinen Vernunft*, 1781 (A), 2nd ed. 1787 (B). In *Kant's gesammelte Schriften, Band III*, Berlin: G. Reimer; trans. Norman Kemp Smith (1929). Macmillan, London, 1929. (*KdrV*)

Kleene, Stephen Cole. 1952. *Introduction to Metamathematics*. Amsterdam: North-Holland.

Kolmogoroff, A. N. 1932. Zur Deutung der intuitionistischen Logik. *Mathematische Zeitschrift* 35: 58–65. English translation in Mancosu (1998), 328–334.

Kreiser, Lothar. 2001. *Gottlob Frege. Leben–Werk–Zeit*. Hamburg: Felix Meiner.

Kutschera, Franz von. 1996. Frege and Natural Deduction. In *Frege: Importance and Legacy*, ed. M. Schirn, 301–304. Berlin: De Gruyter.

Leibniz, Gottfried Wilhelm. 1960. *Fragmente zur Logik*. Berlin: Akademie-Verlag. Edited, selected, and translated by Franz Schmidt.

Lovejoy, Arthur. 1936. *The Great Chain of Being*. Cambridge, Mass.: Harvard University Press.

Mancosu, Paolo, ed. 1998. *From Brouwer to Hilbert*. Oxford: Oxford University Press.

Maritain, Jaques. 1946. *An Introduction to Logic*. London: Sheed and Ward.

Martin-Löf, Per. 1982. Constructive Mathematics and Computer Programming. In *Logic, Methodology and Philosophy of Science VI, Hannover 1979*, ed. L. J. Cohen et al., 153–175. Amsterdam: North-Holland.

Martin-Löf, Per. 1983. On the meanings of the logical constants and the justifications of the logical laws, lectures delivered in Sienna, first distributed in 1985, and printed in *Nordic Journal of Philosophical Logic* I (1996): 11–60. Available at http://www.hf.uio.no/filosofi/njpl.

Martin-Löf, Per. 1984. *Intuitionistic Type Theory*. Naples: Bibliopolis.

Martin-Löf, Per. 1987. Truth of a proposition, evidence of judgement, validity of a proof. *Synthese* 73: 191–212.

Martin-Löf, Per. 1991. A Path from Logic to Metaphysics. In *Atti del Congresso Nuovi Problemi della Logica e della Filosofia della Scienza, Viareggio, 8–13 gennaio, 1990*, eds. G. Corsi and G. Sambin, 141–149. Bologna: CLUEB.

Martin-Löf, Per. 1994. Analytic and Synthetic Judgements in Type Theory. In *Kant and Contemporary Epistemology*, ed. P. Parrini, 87–99. Dordrecht: Kluwer.

McGuinness, Brian. 1988. *Wittgenstein, a Life: Young Ludwig, 1889–1921*. London: Duckworth.

Menzler-Trott, E. 2001. *Gentzens Problem: Mathematishe Logik im nationalsozialistischen Deutschland*. Basel: Birkhäuser.

Molk, Jules. 1885. Sur une notion qui comprend celle de la dvisibilité et sur la théorie générale de l'élimination. *Acta Mathematica* 6: 1–166.

Monk, Ray. 1990. *Wittgenstein: The Duty of Genius*. London: Jonathan Cape.

Moore, G. E. 1898. The nature of judgement. *Mind* 8: 176–193.

Moore, G. E. 1902. Truth. In *Encyclopaedia of Philosophy and Psychology*, ed. J. M. Baldwin, vol. 2, 716–718. London: Macmillan.

Moore, G. E. 1907. Mr. Joachim's *Nature of Truth*. *Mind* 16: 229–235.

Moore, G. E. 1907–08. Professor James' "Pragmatism". *Proc. Arist. Soc.* 8. Cited from the reprint in *Some Problems of Philosophy*. London: Routledge and Kegan Paul, 1922, 97–146.

Moore, G. E. 1953. *Some Main Problems of Philosophy*. London: George Allen and Unwin.

Mulligan, Kevin, Peter Simons, and Barry Smith. 1984. Truth-makers. *Philosophy and Phenomenological Research* 44: 287–321.

Prawitz, Dag. 1965. *Natural Deduction*. Uppsala: Almqvist & Wiksell.

Prawitz, Dag. 1971. Ideas and Results in Proof Theory. In *Proceedings of the Second Scandinavian Logic Symposium*, ed. J. E. Fenstad, 235–307. Amsterdam: North-Holland.

Putnam, Hilary. 1982. Peirce the logician. *Historia Mathematica* 9: 290–301.

Quine, Willard Van Orman. 1950. *Methods of Logic*. New York: Holt.

Ramsey, Frank Plumpton. 1926. The foundations of mathematics. *Proceedings of the London Mathematical Society* 25(2): 338–384.

Ranta, Aarne. 1994. *Type-Theoretical Grammar*. Oxford: Clarendon Press.

Rorty, Richard. 1967. *The Linguistic Turn*. Chicago: University of Chicago Press.

Russell, Bertrand. 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press.

Russell, Bertrand. 1904. Meinong's theory of complexes and assumptions. *Mind* 13: 204–219, 336–354, 509–524.

Russell, Bertrand. 1912. *The Problems of Philosophy.* Oxford: Oxford University Press. 1982.

Scholz, Heinrich. 1930. Die Axiomatik der Alten. *Blätter für deutsche Philosophie* 4: 259–278. Cited from Scholz (1961), 27–45. English translation by Jonathan Barnes, Ancient Axiomatic Theory. In *Articles on Aristotle* 1: *Science*, eds. J. Barnes, M. Schofield, and R. Sorabji, 1975, 50–60. London: Duckworth.

Scholz, Heinrich. 1961. *Mathesis Universalis*, eds. Hans Hermes, Friedrich Kambartel, and Joachim Ritter. Basel: Benno Schwabe Verlag.

Schroeder-Heister, Peter. 1981. Untersuchungen zur regellogischen Deutung von Aussagenverknüpfungen. Diss., Bonn.

Schroeder-Heister, Peter. 1984. A natural extension of natural deduction. *Journal of Symbolic Logic* 49: 1284–1300.

Schroeder-Heister, Peter. 1987. *Structural Frameworks with Higher-Level Rules.* Habilitationsschrift. University of Konstanz, Department of Philosophy.

Schroeder-Heister, Peter. 1999. Gentzen-Style Features in Frege. In *Abstracts of the 11th International Congress of Logic, Methodology, and Philosophy of Science* (Cracow, August 1999), 499. Cracow.

Sebestik, Jan. 1990. The Archeology of the Tractatus: Bolzano and Wittgenstein. In *Wittgenstein—Towards a Re-Evaluation, Proc. 14th Int. Wittgenstein Symp., Kirchberg am Wechsel, 13–20 August 1989, Vol. I*, eds. Rudolf Haller and J. Brandl, 112–128. Wien: Verlag Hölder-Pichler-Tempsky.

Sebestik, Jan. 1992. The Construction of Bolzano's Logical System. In *Bolzano's Wissenschafstlehre 1837–1987. International Workshop Firenze, 16–19 settembre 1987*, 163–177. Firenze: Leo S. Olschki.

Siebel, Mark. 1996. *Der Begriff der Ableitbarkeit bei Bolzano.* St Augustin: Academia Verlag.

Simons, Peter. 1999. Bolzano, Brentano, and Meinong: Three Austrian Realists. In *German Philosophy since Kant* (Royal Institute of Philosophy Supplement 44), ed. Anthony O'Hear, 109–136. Cambridge: Cambridge University Press.

Stenius, Erik. 1960. *Wittgenstein's Tractatus: A Critical Expostion of Its Main Line of Thought.* Oxford: Basil Blackwell.

Stepanians, Markus. 1998. *Frege und Husserl über Urteilen und Denken.* Paderborn: Schöningh.

Sundholm, B. G. 1983. Systems of Deduction. In *Handbook of Philosophical Logic* I, eds. D. Gabbay and F. Guenthner, 133–188. Dordrecht: Reidel.

Sundholm, B. G. 1986. Proof Theory and Meaning. In *Handbook of Philosophical Logic*, eds. D. Gabbay and F. Guenthner, vol. III, 471–506. Dordrecht: Reidel.

Sundholm, B. G. 1988. *Oordeel en Gevolgtrekking: Bedereigde Species?* (Judgement and Inference. Endangered Species?). Inauguaral Lecture, Leyden University, September 9, 1988. Published in pamphlet form by the university.

Sundholm, B. G. 1994a. Vestiges of Realism. In *The Philosophy of Michael Dummett*, eds. Brian McGuinness and G. Oliveri, 137–165. Dordrecht: Kluwer.

Sundholm, B. G. 1994b. Ontologic versus Epistemologic: Some Strands in the Development of Logic, 1837–1957. In *Logic and Philosophy of Science in Uppsala*, eds. Dag Prawitz and Dag Westerståhl, 373–384. Dordrecht: Kluwer.

Sundholm, B. G. 1994c. Proof-theoretical semantics and Fregean identity-criteria for propositions. *Monist* 77: 294–314.

Sundholm, B. G. 1997. Implicit epistemic aspects of constructive logic. *Journal of Logic, Language and Information* 6: 191–212.

Sundholm, B. G. 1998. Intuitionism and Logical Tolerance. In *Alfred Tarski and the Vienna Circle* (Vienna Circle Institute Yearbook), eds. Jan Wolenski and E. Köhler, 135–149. Dordrecht: Kluwer.

Sundholm, B. G. 2002. A Century of Inference: 1837–1936. In *Logic, Methodology and Philosophy of Science 11, Cracow 1999*, eds. Jan Wolenski and K. Placek. Dordrecht: Kluwer.

Sundholm, B. G. 2004. Antirealism and the Roles of Truth. In *Handbook of Epistemology*, eds. I. Niiniluoto, M. Sintonen, and J. Wolenski, 437–466. Dordrecht: Kluwer.

Tarski, Alfred. 1933a. *Pojece prawdy wjezykach deukcyjnynch, Travaux de la Société des Sciences et des Letttres de Varsovie, Classe III Sciences Mathématiques et Physiques*, vol. 34. (German translation by Leopold Blaustein: Tarski 1935).

Tarski, Alfred. 1933b. Einige Betrachtungen über die Begriffe der ω-Widerspruchsfreiheit und der ω-Vollständigkeit. *Monatshefte für Mathematik und Physik* 40: 97–112.

Tarski, Alfred. 1935. Der Wahrheitsbegriff in den formalisierten Sprachen. In *Studia Philosophica* I (1936), 261–405. Lemberg: Polish Philosophical Society. Offprints in monograph form dated 1935.

Tarski, Alfred. 1944. The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research* 4: 341–375.

Tarski, Alfred. 1956. *Logic, Semantics, Metamathematics*. Oxford: Clarendon Press.

Tarski, Alfred, with Jan Lukasiewicz. 1930. Untersuchungen über den Aussagenkalkül. *Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie, Classe III* 23: 30–50. English translation in Tarski (1956, pp. 38–59).

Tarski, Alfred, with Robert Vaught. 1957. Arithmetical extensions of relational systems. *Compositio Mathematicae* 13: 81–102.

Textor, Mark. 1996. *Bolzanos Propositionalismus*. Berlin: De Gruyter.

Thiel, Christian. 1988. Die Kontroverse um die intuitionistische Logik vor ihre Axiomatisierung durch Heyting im Jahre 1930. *History and Philosophy of Logic* 9: 67–75.

Tichý, Pavel. 1988. *Frege's Foundations of Logic*. Berlin: De Gruyter.

Tichý, Pavel and Jindra Tichý. 1999. On Inference. In *The LOGICA Yearbook 1998*, Philosophical Institute, Czech Academy of Science, 73–85. Prague: FILOSOFIA Publishers.

Van Dalen, Dirk. 1999. *Mystic, Geometer, and Intuitionist: The Life of L. E. J. Brouwer*, vol. 1. *The Dawning Revolution*. Oxford: Clarendon Press.

Weyl, Hermann. 1918a. *Raum–Zeit–Materie*. Berlin: Julius Springer.

Weyl, Hermann. 1918b. *Das Kontinuum*. Leipzig: Veit.

Weyl, Hermann. 1921. Über die neue Grundlagenkrise der Mathematik. *Mathematische Zeitschrift* 10: 39–79. English translation in Mancosu (1998), pp. 86–118.

Wittgenstein, Ludwig. 1921. Logisch-philosophische Abhandlung. *Annalen für Naturphilosophie* 14: 198–262. Reprinted with English translation in Wittgenstein (1922).

Wittgenstein, Ludwig. 1922. *Tractatus logico-philosophicus*. London: Kegan Paul.

9

The Development of Mathematical Logic from Russell to Tarski, 1900–1935

Paolo Mancosu, Richard Zach, and Calixto Badesa

The following nine itineraries in the history of mathematical logic do not aim at a complete account of the history of mathematical logic during the period 1900–1935. For one thing, we had to limit our ambition to the technical developments without attempting a detailed discussion of issues such as what conceptions of logic were being held during the period. This also means that we have not engaged in detail with historiographical debates which are quite lively today, such as those on the universality of logic, conceptions of truth, the nature of logic itself, and so on. While of extreme interest, these themes cannot be properly dealt with in a short space, as they often require extensive exegetical work. We therefore merely point out in the text or in appropriate notes how the reader can pursue the connection between the material we treat and the secondary literature on these debates. Second, we have not treated some important developments. While we have not aimed at completeness, our hope has been that by focusing on a narrower range of topics our treatment will improve on the existing literature on the history of logic. There are excellent accounts of the history of mathematical logic available, such as, to name a few, Kneale and Kneale (1962), Dumitriu (1977), and Mangione and Bozzi (1993). We have kept the secondary literature quite present in that we also wanted to write an essay that would strike a balance between covering material that was adequately discussed in the secondary literature and presenting new lines of investigation. This explains, for instance, why the reader will find a long and precise exposition of Löwenheim's (1915) theorem but only a short one on Gödel's incompleteness theorem: Whereas there is hitherto no precise presentation of the first result, accounts of the second result abound. Finally, the treatment of the foundations of mathematics is quite restricted, and it is ancillary to the exposition of the history of mathematical logic. Thus, it is not meant to be the main focus of our exposition.[1]

Page references in citations are to the English translations, if available, or to the reprint edition, if listed in the bibliography. All translations are the authors', unless an English translation is listed in the references.

1. Itinerary I. Metatheoretical Properties of Axiomatic Systems

1.1. Introduction

The two most important meetings in philosophy and mathematics in 1900 took place in Paris. The First International Congress of Philosophy met in August and so did, soon after, the Second International Congress of Mathematicians. As symbolic, or mathematical, logic has traditionally been part of both mathematics and philosophy, a glimpse at the contributions in mathematical logic at these two events will give us a representative selection of the state of mathematical logic at the beginning of the twentieth century. At the International Congress of Mathematicians, Hilbert presented his famous list of problems (Hilbert 1900a), some of which became central to mathematical logic, such as the continuum problem, the consistency proof for the system of real numbers, and the decision problem for Diophantine equations (Hilbert's tenth problem). However, despite the attendance of remarkable logicians like Schröder, Peano, and Whitehead in the audience, the only other contributions that could be classified as pertaining to mathematical logic were two talks given by Alessandro Padoa on the axiomatizations of the integers and of geometry, respectively.

The third section of the International Congress of Philosophy was devoted to logic and history of the sciences (Lovett 1900–1901). Among the contributors of papers in logic we find Russell, MacColl, Peano, Burali-Forti, Padoa, Pieri, Poretsky, Schröder, and Johnson. Of these, MacColl, Poretsky, Schröder, and Johnson read papers that belong squarely to the algebra of logic tradition. Russell read a paper on the application of the theory of relations to the problem of order and absolute position in space and time. Finally, the Italian school of Peano and his disciples—Burali-Forti, Padoa, and Pieri—contributed papers on the logical analysis of mathematics. Peano and Burali-Forti spoke on definitions, Padoa read his famous essay containing the "logical introduction to any theory whatever," and Pieri spoke on geometry considered as a purely logical system. Although there are certainly points of contact between the first group of logicians and the second group, already at that time it was obvious that two different approaches to mathematical logic were at play.

Whereas the algebra of logic tradition was considered to be mainly an application of mathematics to logic, the other tradition was concerned more with an analysis of mathematics by logical means. In a course given in 1908 in

Göttingen, Zermelo captured the double meaning of mathematical logic in the period by reference to the two schools:

> The word "mathematical logic" can be used with two different meanings. On the one hand one can treat logic mathematically, as it was done for instance by Schröder in his Algebra of Logic; on the other hand, one can also investigate scientifically the logical components of mathematics. (Zermelo 1908a, 1)[2]

The first approach is tied to the names of Boole and Schröder, the second was represented by Frege, Peano, and Russell.[3] We will begin by focusing on mathematical logic as the logical analysis of mathematical theories, but we will return later (see itinerary IV) to the other tradition.

1.2. Peano's School on the Logical Structure of Theories

We have mentioned the importance of the logical analysis of mathematics as one of the central motivating factors in the work of Peano and his school on mathematical logic. First of all, Peano was instrumental in emphasizing the importance of mathematical logic as an artificial language that would remove the ambiguities of natural language, thereby allowing a precise analysis of mathematics. In the words of Pieri, an appropriate ideographical algorithm is useful as "an instrument appropriate to guide and discipline thought, to exclude ambiguities, implicit assumptions, mental restrictions, insinuations and other shortcomings, almost inseparable from ordinary language, written as well as spoken, which are so damaging to speculative research" (Pieri 1901, 381). Moreover, he compared mathematical logic to "a microscope which is appropriate for observing the smallest difference of ideas, differences that are made imperceptible by the defects of ordinary language in the absence of some instrument that magnifies them" (382). It was by using this "microscope" that Peano was able, for instance, to clarify the distinction between an element and a class containing only that element and the related distinction between membership and inclusion.[4]

The clarification of mathematics, however, also meant accounting for what was emerging as a central field for mathematical logic: the formal analysis of mathematical theories. The previous two decades had in fact seen much activity in the axiomatization of particular branches of mathematics, including arithmetic, algebra of logic, plane geometry, and projective geometry. This culminated in the explicit characterization of a number of formal conditions for which axiomatized mathematical theories should strive. Let us consider first Pieri's description of his work on the axiomatization of geometry, which had been carried out independently of Hilbert's famous *Foundations of Geometry* (1899). In his presentation to the International Congress of Philosophy in 1900, Pieri emphasized that the study of geometry is following arithmetic in becoming more and more "*the study of a certain order of logical relations*; in freeing itself little by little from the bonds which still keep it tied (although weakly) to *intuition*, and in displaying consequently the form and quality of *purely deductive, abstract and ideal science*" (Pieri 1901, 368). Pieri saw in this abstraction from concrete interpretations a unifying thread running through the development of arithmetic, analysis, and geometry in the nineteenth century. This led him to a conception of geometry as a hypothetical discipline (he coined the term "hypothetico-deductive"). In fact he goes on to assert that the primitive notions of any deductive system whatsoever "must be capable of arbitrary interpretations in certain limits assigned by the primitive propositions," subject only to the restriction that the primitive propositions must be satisfied by the particular interpretation. The analysis of a hypothetico-deductive system begins then with the distinction between primitive notions and primitive propositions. In the logical analysis of a hypothetico-deductive system it is important not only to distinguish the derived theorems from the basic propositions (definitions and axioms) but also to isolate the primitive notions, from which all the others are defined. An ideal to strive for is that of a system whose primitive ideas are irreducible, that is, such that none of the primitive ideas can be defined by means of the others through logical operations. Logic is here taken to include notions such as, among others, "individual," "class," "membership," "inclusion," "representation," and "negation" (383). Moreover, the postulates, or axioms, of the system must be independent, that is, none of the postulates can be derived from the others.

According to Pieri, there are two main advantages to proceeding in such an orderly way. First of all, keeping a distinction between primitive notions and derived notions makes it possible to compare different hypothetico-deductive systems as to logical equivalence. Two systems turn out to be equivalent if for every primitive notion of one we can find an explicit definition in the second one such that all primitive propositions of the first system become theorems of the second system, and vice versa. The second advantage consists in the possibility of abstracting from the meaning of the primitive notions and thus operate symbolically on expressions which admit of different interpretations, thereby encompassing in a general and abstract system several concrete and specific instances satisfying the relations stated by the postulates. Pieri is well known for his clever application of these methodological principles to geometrical systems (see Freguglia 1985; Marchisotto 1995). Pieri refers to Padoa's articles for a more detailed analysis of the properties connected to axiomatic systems.

Alessandro Padoa was another member of the group around Peano. Indeed, of that group, he is the only one whose name has remained attached to a specific result in mathematical logic, that is, Padoa's method for proving indefinability (see the following). The result was stated in the talks Padoa gave in 1900 at the two meetings mentioned at the outset (Padoa 1901, 1902). We will follow the "Essai d'une théorie algébrique des nombre entiers, précédé d'une introduction logique a une théorie déductive quelconque." In the Avant-Propos (not translated in van Heijenoort 1967a) Padoa lists a number of notions that he considers as belonging to general logic such as class ("which corresponds to the words: *terminus* of the scholastics, *set* of the mathematicians, *common noun* of ordinary language"). The notion of class is not defined but assumed with its informal meaning. Extensionality for classes is also assumed: "a class is completely known when one knows which individuals belong to it." However, the notion of ordered class he considers as lying outside of general logic. Padoa then states that all symbolic definitions have the form of an equality $y = b$ where y is the new symbol and b is a combination of symbols already known. This is illustrated with the property of being a class with one element. Disjunction and negation are given with their class interpretation. The notions "there is" and "there is not" are also claimed to be reducible to the notions already previously introduced. For instance, Padoa explains that given a class a to say "there is no a" means that the class not-a contains everything, that is, not-a = (a or not-a). Consequently, "there are a['s]" means: not-$a \neq$ (a or not-a). The notion of transformation is also taken as belonging to logic. If a and b are classes and if for any x in a, ux is in b, then u is a transformation from a into b. An obvious principle for transformations u is: if $x = y$ then $ux = uy$. The converse, Padoa points out, does not follow.

This much was a preliminary to the section of Padoa's paper titled "Introduction logique a une théorie déductive quelconque." Padoa makes a distinction between general logic and specific deductive theories. General logic is presupposed in the development of any specific deductive theory. What characterizes a specific deductive theory is its set of primitive symbols and primitive propositions. By means of these, one defines new notions and proves theorems of the system. Thus, when one speaks of indefinability or unprovability, one must always keep in mind that these notions are relative to a specific system and make no sense independently of a specific system. Restating his notion of definition he also claims that definitions are eliminable and thus inessential. Just like Pieri, Padoa also speaks of systems of postulates as a pure formal system on which one can reason without being anchored to a specific interpretation, "for what is necessary to the logical development of a deductive theory is not *the empirical knowledge of the properties of things, but the formal knowledge of relations between symbols*" (1901, 121). It is possible, Padoa continues, that there are several, possibly infinite, interpretations of the system of undefined symbols which verify the system of basic propositions and thus all the theorems of a theory. He then adds:

> The system of undefined symbols can then be regarded as the *abstraction* obtained from all these interpretations, and the generic theory can then be regarded as the *abstraction* obtained from the *specialized theories* that result when in the generic theory the system of undefined symbols is successively replaced by each of the interpretations of this theory. Thus, by means of just one argument that proves a proposition of the generic theory we prove implicitly a proposition in each of the specialized theories. (1901, 121)[5]

In contemporary model theory, we think of an interpretation as specifying a domain of individuals with relations on them satisfying the propositions of the system, by means of an appropriate function sending individual constants to objects and relation symbols to subsets of the domain (or Cartesian products of the same). It is important to remark that in Padoa's notion of interpretation something else is going on. An interpretation of a generic system is given by a concrete set of propositions with meaning. In this sense the abstract theory captures all of the individual theories, just as the expression $x + y = y + x$ captures all the particular expressions of the form $2 + 3 = 3 + 2$, $5 + 7 = 7 + 5$, and so on.

Moving now to definitions, Padoa states that when we define a notion in an abstract system we give conditions which the defined notion must satisfy. In each particular interpretation the defined notion becomes individualized, that is, it obtains a meaning that depends on the particular interpretation. At this point Padoa states a general result about definability. Assume that we have a general deductive system in which all the basic propositions are stated by means of undefined symbols:

> We say that *the system of undefined symbols is **irreducible** with respect to the system of unproved propositions* when *no symbolic definition of any undefined symbol can be deduced from the system of unproved propositions*, that is, when we cannot deduce from the system a relation of the form $x = a$, where x is one of the undefined symbols and a is a sequence of other such symbols (and logical symbols). (1901, 122)

How can such a result be established? Clearly one cannot adduce the failure of repeated attempts at defining the symbol; for such a task, a method for demonstrating the irreducibility is required. The result is stated by Padoa as follows:

> To prove that the system of undefined symbols is irreducible with respect to the system of unproved propositions it is necessary and sufficient to find, for any undefined symbol, an interpretation of the system of undefined symbols that verifies the system of unproved propositions and that continues to do so if we suitably change the meaning of only the symbol considered. (1901, 122)[6]

Padoa (1902) covers the same ground more concisely but also adds the criterion of compatibility for a set of postulates: "To prove the compatibility of a set of postulates one needs to find an interpretation of the undefined symbols which verifies simultaneously all the postulates" (1902, 249). Padoa applied his criteria to showing that his axiomatization of the theory of integers satisfied the condition of compatibility and irreducibility for the primitive symbols and postulates.

We thus see that for Padoa the study of the formal structure of an arbitrary deductive theory was seen as a task of general logic. What can be said about these metatheoretical results in comparison to the later developments? We have already pointed out the different notion of interpretation which informs the treatment. Moreover, the system of logic in the background is never fully spelled out, and in any case it would be a logic containing a good amount of set-theoretic notions. For this reason, some results are taken as obvious that would actually need to be justified. For instance, Padoa claims that if an interpretation satisfies the postulates of an abstract theory, then the theorems obtained from the postulates are also satisfied in the interpretation. This is a soundness principle, which nowadays must be shown to hold for the system of derivation and the semantics specified for the system. For similar reasons the main result by Padoa on the indefinability of primitive notions does not satisfy current standards of rigor. Thus, a formal proof of Padoa's definability theorem had to wait until the works of Tarski (1934–1935) for the theory of types and Beth (1953) for first-order logic (see van Heijenoort 1967a, 118–119, for further details).

1.3. Hilbert on Axiomatization

In light of the importance of the work of Peano and his school on the foundations of geometry, it is quite surprising that Hilbert did not acknowledge their work in the *Foundations of Geometry*. Although it is not quite clear to what extent Hilbert was familiar with the work of the Italian school in the last decade of the nineteenth century (Toepell 1986), he certainly could not ignore their work after the 1900 International Congress in Mathematics. In many ways Hilbert's work on axiomatization resembles the level of abstractness also emphasized by Peano, Padoa, and Pieri. The goal of *Foundations of Geometry* (1899) is to investigate geometry axiomatically.[7] At the outset we are asked to give up the intuitive understanding of notions like point, line, or plane and consider any three system of objects and three sorts of relations between these objects (lies on, between, congruent). The axioms only state how these properties relate the objects in question. They are divided into five groups: axioms of incidence, axioms of order, axioms of congruence, axiom of parallels, and axioms of continuity.

Hilbert emphasizes that an axiomatization of geometry must be complete and as simple as possible.[8] He does not make explicit what he means by completeness, but the most likely interpretation of the condition is that the axiomatic system must be able to capture the extent of the ordinary body of geometry. The requirement of simplicity includes, among other things, reducing the number of axioms to a finite set and showing their independence. Another important requirement for axiomatics is showing the consistency of the axioms of the system. This was unnecessary for the old axiomatic approaches to geometry (such as Euclid's) because one always began with the assumption that the axioms were true of some reality and thus consistency was not an issue. But in the new conception of axiomatics, the axioms do not express truths but only postulates whose consistency must be investigated.

Hilbert shows that the basic axioms of his axiomatization are independent by displaying interpretations in which all of the axioms except one are true.[9] Here we must point to a small difference with the notion of interpretation we have seen in Pieri and Padoa. Hilbert defines an interpretation by first specifying what the set of objects consists in. Then a set of relations among the objects is specified in such a way that consistency or independence is shown. For instance, for showing the consistency of his axioms, he considers a domain given by the subset of algebraic numbers of the form $\sqrt{1+\omega^2}$ and then specifies the relations as being sets of ordered pairs and ordered triples of the domain. The consistency of the geometrical system is thus discharged on the new arithmetical system: "From these considerations it follows that every contradiction resulting from our system of axioms must also appear in the arithmetic defined above" (29).

Hilbert had already applied the axiomatic approach to the arithmetic of real numbers. Just as in the case of geometry, the axiomatic approach to the real numbers is conceived in terms of "a framework of concepts to which we are led of course only by means of intuition; we can nonetheless operate with this framework without having recourse to intuition." The consistency problem for the system of real numbers was one of the problems that Hilbert stated at the International Congress in 1900:

> But above all I wish to designate the following as the most important among the numerous questions which can be asked with regard to the axioms: *To prove that they are not contradictory, that is, that a finite number of logical steps based upon them can never lead to contradictory results.* (1900a, 1104)

In the case of geometry, consistency is obtained by "constructing an appropriate domain of numbers such that to the geometrical axioms correspond analogous relations among the objects of this domain." For the axioms of arithmetic, however, Hilbert required a direct proof, which he conjectured could be obtained by a modification of the arguments already used in "the theory of irrational numbers."[10] We do not know what Hilbert had in mind, but in any case, in his new approach to the problem (1905b), Hilbert made considerable progress in conceiving how a direct proof of consistency for arithmetic might proceed. We will postpone treatment of this issue to later (see itinerary VI) and go back to specify what other metatheoretical properties of axiomatic systems were being discussed in these years. By way of introduction to the next section, something should be said here about one of the axioms, which Hilbert in his Paris lecture calls axiom of integrity and later completeness axiom. The axiom says that the (real) numbers form a system of objects which cannot be extended (Hilbert 1900b, 1094). This axiom is in effect a metatheoretical statement about the possible interpretations of the axiom system.[11] In the second and later editions of the *Foundations of Geometry*, the same axiom is also stated for points, straight lines and planes:

> (Axiom of completeness) It is not possible to add new elements to a system of points, straight lines, and planes in such a way that the system thus generalized will form a new geometry obeying all the five groups of axioms. In other words, the elements of geometry form a system which is incapable of being extended, provided that we regard the five groups of axioms as valid. (Hilbert 1902, 25)

Hilbert commented that the axiom was needed to guarantee that his geometry turn out to be identical to Cartesian geometry. Awodey and Reck (2002) write, "what this last axiom does, against the background of the others, is to make the whole system of axioms categorical. ... He does not state a theorem that establishes, even implicitly, that his axioms are categorical; he leaves it ... without proofs" (11). The notion of categoricity was made explicit in the important work of the "postulate theorists," to which we now turn.

1.4. Completeness and Categoricity in the Work of Veblen and Huntington

A few metatheoretical notions that foreshadow later developments emerged during the early years of the twentieth century in the writings of Huntington and Veblen. Huntington and Veblen are part of a group of mathematicians known as the American postulate theorists (Scanlan 1991, 2003). Huntington was concerned with providing "complete" axiomatizations of various mathematical systems, such as the theory of absolute continuous magnitudes (positive real numbers) (1902) and the theory of the algebra of logic (1905). For instance, in 1902 he presented six postulates for the theory of absolute continuous magnitudes, which he claims to form a complete set. A *complete* set of postulates is characterized by the following properties:

1. The postulates are consistent;
2. They are sufficient;
3. They are independent (or irreducible).

By consistency he means that there exists an interpretation satisfying the postulates. Condition 2 asserts that there is essentially only one such interpretation possible. Condition 3 says that none of the postulates is a "consequence" of the other five.

A system satisfying the conditions (1) and (2) we would nowadays call "categorical" rather than "complete." Indeed, the word "categoricity" was introduced in this context by Veblen in a paper on the axiomatization of geometry (1904). Veblen credits Huntington with the idea and Dewey for having suggested the word "categoricity." The description of the property is interesting:

> Inasmuch as the terms *point* and *order* are undefined one has a right, in thinking of the propositions, to apply the terms in

> connection with any class of objects of which the axioms are valid propositions. It is part of our purpose however to show that there is *essentially only one* class of which the twelve axioms are valid. In more exact language, any two classes K and K' of objects that satisfy the twelve axioms are capable of a one-one correspondence such that if any three elements A, B, C of K are in the order ABC, the corresponding elements of K' are also in the order ABC. Consequently any proposition which can be made in terms of points and order either is in contradiction with our axioms or is equally true of all classes that verify our axioms. The validity of any possible statement in these terms is therefore completely determined by the axioms; and so any further axiom would have to be considered redundant. [Note: Even were it not deducible from the axioms by a finite set of syllogisms] Thus, if our axioms are valid geometrical propositions, they are sufficient for the complete determination of Euclidean geometry.
>
> A system of axioms such as we have described is called *categorical*, whereas one to which it is possible to add independent axioms (and which therefore leaves more than one possibility open) is called *disjunctive*. (Veblen 1904, 346)

A number of things are striking about the passage just quoted. First of all, we are used to define categoricity by appealing directly to the notion of isomorphism.[12] What Veblen does is equivalent to specifying the notion of isomorphism for structures satisfying his 12 axioms. However, the fact that he does not make use of the word "isomorphism" is remarkable, as the expression was common currency in group theory already in the nineteenth century. The word "isomorphism" is brought to bear for the first time in the definition of categoricity in Huntington (1906–1907). There he says that "special attention may be called to the discussion of the notion of isomorphism between two systems, and the notion of a sufficient, or categorical, set of postulates." Indeed, on p. 26 (1906–1907), the notion of two systems being isomorphic with respect to addition and multiplication is introduced. We are now very close to the general notion of isomorphism between arbitrary systems satisfying the same set of axioms. The first use of the notion of isomorphism between arbitrary systems we have been able to find is Bôcher (1904, 128), who claims to have generalized the notion of isomorphism familiar in group theory. Weyl (1910) also gives the definition of isomorphism between systems in full generality.

Second, there is a certain ambiguity between defining categoricity as the property of admitting only one model (up to isomorphism) and conflating the notion with a consequence of it, namely, what we would now call semantical completeness.[13] Veblen, however, rightly states that in the case of a categorical theory, further axioms would be redundant even if they were not deducible from the axioms by a finite number of inferences.

Third, the distinction hinted at between what is derivable in a finite number of steps and what follows logically displays a certain awareness of the difference between a semantical notion of consequence and a syntactical notion of derivability and that the two might come apart. However, Veblen does not elaborate on the issue.

Finally, later in the section Veblen claims that the notion of categoricity is also expressed by Hilbert's axiom of completeness as well as by Huntington's notion of sufficiency. In this he reveals an inaccurate understanding of Hilbert's completeness axiom and of its consequences. Baldus (1928) is devoted to showing the noncategoricity of Hilbert's axioms for absolute geometry even when the completeness axiom is added. It is, however, true that in the presence of all the other axioms, the system of geometry presented by Hilbert is categorical (see Awodey and Reck 2002).

1.5. Truth in a Structure

These developments have relevance also for the discussion of the notion of truth in a structure. In his influential paper (1986), Hodges raises several historical issues concerning the notion of truth in a structure, which can now be made more precise. Hodges is led to investigate some of the early conceptions of structure and interpretation with the aim of finding out why Tarski did not define truth in a structure in his early articles. He rightly points out that algebraists and geometers had been studying "Systeme von Dingen" (systems of objects), that is, what we would call structures or models (on the emergence of the terminology, see itinerary VIII). Thus, for instance, Huntington in (1906–1907) describes the work of the postulate theorist in algebra as being the study of all the systems of objects satisfying certain general laws: "From this point of view our work becomes, in reality, much more general than a study of the system of numbers; it is a study of any system which satisfies the conditions laid down in the general laws of §1."[14] Hodges then pays attention to the terminology used by mathematicians of the time to express that a structure A obeys some laws and quotes Skolem (1933) as one of the earliest occurrences where the expression "true in a structure" appears.[15]

However, here we should point out that the notion of a proposition being true in a system is not unusual during the period. For instance, in Weyl's (1910) definition of isomorphism, we read that if there is an isomorphism between two systems, "there is also such a unique correlation between the propositions true with respect to one system and those true with respect to the other, and we can, without falling into error, identify the two systems outright" (Weyl 1910, 301). Moreover, although it is usual in Peano's school and among the American postulate theorists to talk about a set of postulates being "satisfied" or "verified" in a system (or by an interpretation), without any further comments, sometimes we are also given a clarification that shows that they were willing to use the notion of truth in a structure. A few examples will suffice.

Let us look at what might be the first application of the method for providing proofs of independence. Peano in "Principii di geometria logicamente esposti" (1889) has two signs, 1 (for point) and $c \varepsilon ab$ (c is a point internal to the segment ab). Then he considers three categories of entities with a relation defined between them. Finally he adds:

> Depending on the meaning given to the undefined signs 1 and $c \varepsilon ab$, the axioms might or might not be satisfied. If a certain group of axioms is verified, then all the propositions that are deduced from them will also be true, since the latter propositions are only transformations of those axioms and of those definitions. (Peano 1889, 77–78)

In 1900, Pieri explains that

> the postulates, just like all conditional propositions *are neither true nor false:* they only express conditions that can sometimes be verified and sometimes not. Thus for instance, the equality $(x + y)^2 = x^2 + 2xy + y^2$ is *true*, if x and y are real numbers and *false* in the case of quaternions (giving for each hypothesis the usual meaning to $+$, $\times$, etc.). (Pieri 1901, 388–389)

In 1906, Huntington:

> The only way to avoid this danger [of using more than is stated in the axioms] is to think of our fundamental laws, not as axiomatic propositions about numbers, but as *blank forms* in which the letters a, b, c, etc. may denote any objects we please and the symbols $+$ and $\times$ any rules of combination; such a blank form will become a proposition only when a definite interpretation is given to the letters and symbols—indeed a true proposition for some interpretations and a false proposition for others... From this point of view our work becomes, in reality, much more general than a study of the system of numbers; it is a study of any system which satisfies the conditions laid down in the general laws of §1. (Huntington 1906–1907, 2–3)[16]

In short, it seems that the expression "a system of objects verifies a certain proposition or a set of axioms" is considered to be unproblematic at the time, and it is often read as shorthand for a sentence, or a set of sentences, being true in a system. Of course, this is not to deny that in light of the philosophical discussion emerging from non-Euclidean geometries, a certain care was exercised in talking about "truth" in mathematics, but the issue is resolved exactly by the distinction between axioms and postulates. Whereas the former had been taken to be true tout court, the postulates only make a demand, which might be satisfied or not by particular system of objects (see also on the distinction, Huntington 1911, 171–172).

2. Itinerary II. Bertrand Russell's Mathematical Logic

2.1. From the Paris Congress to the *Principles of Mathematics* 1900–1903

At the time of the Paris congress, Russell was mainly familiar with the algebra of logic tradition. He certainly knew the works of Boole, Schröder, and Whitehead. Indeed, the earliest drafts of *The Principles of Mathematics* (1903; POM for short) are based on a logic of part-whole relationship that was closely related to Boole's logical calculus. He also had already realized the importance of relations and the limitations of a subject-predicate approach to the analysis of sentences. This change was a central one in his abandonment of Hegelianism[17] and also led him to the defense of absolute position in space and time against the Leibnizian thesis of the relativity of motion and position, which was the subject of his talk at the International Congress of Philosophy, held in Paris in 1900. However, he had not yet read the works of the Italian school. The encounter with Peano and his school in Paris was of momentous importance for Russell. He had been struggling with the problems of the foundation of mathematics for a number of years and thought that Peano's system had finally shown him the way. After returning from the Paris congress, Russell familiarized himself with the publications of Peano and his school, and it became clear to him that "[Peano's] notation afforded an instrument of logical analysis such as I had been seeking for years" (Russell 1967, 218). In Russell's autobiography, he claims that "the most important year of my intellectual life was the year 1900 and the most important event in this year was my visit to the International Congress of Philosophy in Paris" (1989, 12). One of the first things Russell did was to extend Peano's calculus with a worked-out theory of relations and this allowed him to develop a large part of Cantor's work in the new system. This he pursued in his first substantial contribution to logic (Russell 1901b, 1902b), which constitutes a bridge between the theory of relations developed by Peirce and Schröder and Peano's formalization of mathematics. At this stage Russell thinks of relations intensionally, that is, he does not identify them with sets of pairs. The notion of relation is taken as primitive. Then the notion of the domain and co-domain of a relation, among others, are introduced. Finally, the axioms of his theory of relations state, among other things, closure properties with respect to the converse, the complement, the relative product, the union, and the intersection (of relations or classes thereof). He also defines the notion of function in terms of that of relation (however, in POM they are both taken as primitive). In this work, Russell treats natural numbers as definable, which stands in stark contrast to his previous view of number as an indefinable primitive. This led him to the famous definition of "the cardinal number of a class u" as "the class of classes similar to u." Russell arrived at it independently of Frege, whose definition was similar, but he was apparently influenced by Peano, who discussed such a definition in 1901 without, however, endorsing it. In any case, Peano's influence is noticeable in Russell's abandonment of the Boolean leanings of his previous logic in favor of Peano's mathematical logic. Russell now accepted, except for a few changes, Peano's symbolism. One of Peano's advances had been a clear distinction between sentences such as "Socrates is mortal" and "All men are mortal," which were previously conflated as being of the same structure. Despite the similar surface structure, the first one indicates a membership relation between Socrates and the class of mortals, whereas the second indicates an inclusion between classes. In Peano's symbolism we have $s \, \varepsilon \, \phi(x)$ for the first and $\phi(x) \supset_x \psi(x)$ for the second. With this distinction Peano was able to define the relation of subsumption between two classes by means of implication. In a letter to Jourdain in 1910, Russell writes:

> Until I got hold of Peano, it had never struck me that Symbolic Logic would be any use for the Principles of mathematics, because I knew the Boolean stuff and found it useless. It was Peano's ε, together with the discovery that relations could be fitted into his system, that led me to adopt Symbolic Logic. (Grattan-Guinness 1977, 133)

What Peano had opened for Russell was the possibility of considering the mathematical concepts as definable in terms of logical concepts. In particular, an analysis in terms of membership and implication is instrumental in accounting for the generality of mathematical propositions. Russell's logicism finds its first formulation in a popular article written in 1901 where he claims that all the indefinables and indemonstrables in pure mathematics stem from general logic: "All pure mathematics—Arithmetic, Analysis, and Geometry—is built up of the primitive ideas of logic, and its propositions are deduced from the general axioms of logic" (1901a, 367).

This is the project that informed the *Principles of Mathematics* (1903). The construction of mathematics out of logic is carried out by first developing arithmetic through the definition of the cardinal number of a class as the class of classes similar to it. Then the development of analysis is carried out by defining real numbers as sets of rationals satisfying appropriate conditions. (For a detailed reconstruction see, among others, Vuillemin 1968, Rodriguez-Consuegra 1991, Landini 1998, Grattan-Guinness 2000.) The main difficulty in reconstructing Russell's logic at this stage consists in the presence of logical notions mixed with linguistic and ontological categories (denotation, definition). Moreover, Russell does not present his logic by means of a formal language.

After Russell finished preparing POM, he also began studying Frege with care (around June 1902). Under his influence, Russell began to notice the limitations in Peano's treatment of symbolic logic, such as the lack of different symbols for class union and the disjunction of propositions, or material implication and class inclusion. Moreover, he changed his symbolism for universal and existential quantification to $(x)f(x)$ and $(\mathsf{E}x)f(x)$. He adopted from Frege the symbol $\vdash$ for the assertion of a proposition. His letter to Frege of June 16, 1902, contained the famous paradox, which had devastating consequences for Frege's system:

> Let w be the predicate: to be a predicate that cannot be predicated of itself. Can w be predicated of itself? From each answer its opposite follows. Therefore we must conclude that w is not a predicate. Likewise there is no class (as a totality) of those classes which, each taken as a totality, do not belong to themselves. From this I conclude that under certain circumstances a definable collection does not form a totality. (Russell 1902a, 125)

The first paradox does not involve the notion of class but only that of predicate. Let Imp(w) stand for "w cannot be predicated of itself," that is, $\sim w(w)$. Now we ask: Is Imp(Imp) true or ∼Imp(Imp)? From either one of the possibilities the opposite follows. However, what is known as Russell's paradox is the second one offered in the letter to Frege. In his work *Grundgesetzte der Arithmetik* (Frege 1893, 1903), Frege had developed a logicist project that aimed at reconstructing arithmetic and analysis out of general logical laws. One of the basic assumptions made by Frege (Basic Law V) implies that every propositional function has an extension, where extensions are a kind of object. In modern terms we could say that Frege's Basic Law V implies that for any property $F(x)$ there exists a set $y = \{ x : F(x) \}$. Russell's paradox consists in noticing that for the specific $F(x)$ given by $x \notin x$, Frege's principle leads to asserting the existence of the set $y = \{ x : x \notin x \}$. Now if one asks whether $y \in y$ or $y \notin y$ from either one of the assumptions one derives the opposite conclusion. The consequences of Russell's paradox for Frege's logicism and Frege's attempts to cope with it are well known, and we will not recount them here (see Garciadiego 1992). Frege's proposed emendation to his Basic Law V, while consistent, turns out to be inconsistent as soon as one postulates that there are at least two objects (Quine 1955a).[18]

Extensive research on the development that led to Russell's paradox has shown that Russell already obtained the essentials of his paradox in the first half of 1901 (Garciadiego 1992; Moore 1994) while working on Cantor's set theory. Indeed, Cantor himself already noticed that treating the cardinal numbers (resp., ordinal numbers) as a completed totality would lead to contradictions. This led him to distinguish, in letters to Dedekind, between "consistent multiplicities," that is, classes that can be considered as completed totalities, from "inconsistent multiplicities," that is, classes that cannot, on pain of contradiction, be considered as completed totalities. Unaware of Cantor's distinction between consistent and inconsistent multiplicities Russell in 1901 convinced himself that Cantor had "been guilty of a very subtle fallacy" (1901a, 375). His reasoning was that the number of all things is the greatest of all cardinal numbers. However, Cantor proved that for every cardinal number there is a cardinal number strictly bigger than it. Within a few months this conundrum led to Russell's paradox. In POM we find, in addition to the two paradoxes we have discussed, also a discussion of what is now known as Burali-Forti's paradox (Moore and Garciadiego 1981).

In POM Russell offered a tentative solution to the paradoxes: the theory of types. The theory of types contained in POM is a version of what is now called the simple theory of types, whereas the one offered in Russell (1908) (and *Principia Mathematica*, Whitehead and Russell 1910, 1913) is called the ramified theory of types (on the origin of these terms, see Grattan-Guinness 2000, 496). Russell's exposition of the theory of types (in 1903 as well as later) is far from perspicuous, and we will simply give the gist of it. The basic idea is that every propositional function $\phi(x)$ has a range of significance, that is, a range of values of x for which it can be meaningfully said to be true or false:

> Every propositional function $\phi(x)$—so it is contended—has, in addition to its range of truth, a range of significance, i.e., a range within which x must lie if $\phi(x)$ is to be a proposition at all, whether true or false. This is the first point in the theory of types; the second point is that ranges of significance form *types*, i.e., if x belongs to the range of significance of $\phi(x)$, then there is a class of objects, the *type* of x, all of which must also belong to the range of significance of $\phi(x)$, however ϕ may be varied. (Russell 1903, 523)

The lowest type, type 0, is the type of all individuals (objects which are not "ranges"). Then we construct the class of all classes of individuals, namely, type 1. Type 2 is the class of all classes of classes of type 1, and so on. This gives an infinite hierarchy of types for which Russell specifies that "in $x \varepsilon u$ the u must always be of a type higher by one than x" (517). In this way $x \varepsilon x$ and its negation are meaningless and thus it is not possible for Russell's paradox to arise, as there are no ranges of significance, that is, types, for meaningless propositions. The other paradoxes considered by Russell are also blocked by the postulated criteria of meaningfulness. The presentation of the theory in POM is vastly complicated by the need to take into account relations and by a number of assumptions which go against the grain of the theory, for instance, that "$x \varepsilon x$ is sometimes significant" (525).

Russell, however, abandoned this version of the theory of types and returned to the theory of types only after trying a number of different theories. His abandonment of this theory is explained by the fact that the theory does not assign types to propositions and thus, as Russell pointed out to Frege (letter of September 29, 1902), this allows for the generation of a paradox through a diagonal argument applied to classes of propositions. His search for a solution to the paradoxes played a central role in his debate with Poincaré concerning impredicative definitions, to which we now turn.

2.2. Russell and Poincaré on Predicativity

In the wake of Russell's paradoxes, many more paradoxes were brought to light,[19] the most famous being Berry's paradox concerning the least ordinal number not definable in a finite number of words, Richard's paradox (see following discussion), and the König–Zermelo contradiction. The latter concerned a contradiction between König's "proof" that the continuum cannot be well ordered and Zermelo's (1904) proof that every set can be well ordered.

Many more were added, and one finds a long list of paradoxes in the opening pages of Russell (1908). What the paradoxes had brought to light was that not every propositional function defines a class. Russell's paradox, for instance, shows that there is a propositional function, or "norm," $\phi(x)$ for which we cannot assume the existence of $\{ x : \phi(x) \}$. When trying to spell out which propositional functions define classes and which do not, Russell proposed in 1906 the distinction between predicative and nonpredicative norms:

> We have thus reached the conclusion that some norms (if not all) are not entities which can be considered independently of their arguments, and that some norms (if not all) do not define classes. Norms (containing one variable) which do not define classes I propose to call *non-predicative*; those which do define classes I shall call *predicative*. (Russell 1906b, 141)

At the time Russell was considering various theories as possible solutions to the paradoxes and in the 1906 article he mentions three of them: the "no-classes" theory, the "zig-zag" theory, and the "limitation of size" theory. Accordingly, the Russellian distinction between predicative and nonpredicative norms gives rise to extensionally different characterizations depending on the theory under consideration. Russell mentions "simplicity" as the criterion for predicativity in the "zig-zag" theory, and "limitation of size" in the "limitation of size" theory. In the case of the "no-classes" theory, no propositional function is predicative as classes are eliminated through contextual definitions. However, it is only with Poincaré's reply to Russell that we encounter the notion of predicativity that was at the center of their later debate.[20] Poincaré's discussion also takes its start from the paradoxes but rejects Russell's suggestion as to what should count as a predicative propositional function, on account of the vagueness of Russell's proposal. Poincaré suggested that nonpredicative classes are those that contain a vicious circle. Poincaré did not provide a general account, but he clarified the proposal through a discussion of Richard's paradox (Richard 1905). Richard's paradox takes its start by a consideration of the set E of all numbers that can be defined by using expressions of finite length over a finite vocabulary. By a diagonal process one then defines (by appealing explicitly to E) a new number N which is not in the list. As the definition of N is given by a finite expression using exactly the same alphabet used to generate E, it follows that N is in E. But by construction N is not in E. Thus N is and is not in E. Poincaré's way out was to claim that in defining N one is not allowed to appeal to E, as N would be defined in terms of the totality to which it belongs. Thus, according to Poincaré, reference to infinite totalities is the source of the nonpredicativity:

> It is the belief in the existence of actual infinity that has given birth to these non-predicative definitions. I must explain myself. In these definitions we find the word *all*, as we saw in the examples quoted above. The word *all* has a very precise meaning when it

> is a question of a finite number of objects; but for it still to have a precise meaning when the number of the objects is infinite, it is necessary that there should exist an actual infinity. Otherwise all these objects cannot be conceived as existing prior to their definition, and then, if the definition of a notion N depends on *all* the objects A, it may be tainted with the vicious circle, if among the objects A there is one that cannot be defined without bringing in the notion N itself. (Poincaré 1906, 194)

Poincaré was appealing to two different criteria in his diagnosis. On the one hand he considered a definition to be nonpredicative if the definiendum in some way involves the object being defined. The second criterion asserts the illegitimacy of quantifying over infinite sets.[21]

Russell, in "Les Paradoxes de la Logique" (1906a), agreed with Poincaré's diagnosis that a vicious circle was involved in the paradoxes, but he found Poincaré's solution to lack the appropriate generality:

> I recognize, however, that the clue to the paradoxes is to be found in the vicious-circle suggestion; I recognize further this element of truth in M. Poincaré's objection to totality, that whatever in any way concerns *all* or *any* or *some* (undetermined) of the members of a class must not be itself one of the members of a class. In M. Peano's language, the principle I want to advocate may be stated: "Whatever involves an apparent variable must not be among the possible values of that variable." (Russell 1973, 198)

Russell's objection to Poincaré was essentially that Poincaré's proposal was not supported by a general theory and thus seemed ad hoc. Moreover, he pointed out that in many paradoxes infinite totalities play no role and thus he concluded that "the contradictions have no essential reference to infinity." Russell's position brought to light the coexistence of different criteria in Poincaré's notion of predicativity. However, what exactly the vicious circle principle amounted to remained vague also in Russell's work, which displayed several nonequivalent versions of the principle. We resume discussion of predicative mathematics in the section on set theory, and we move now to a discussion of the last element we need to discuss the ramified theory of types—the theory of denoting.

2.3. On Denoting

One of the key elements in the formalization of mathematics given in *Principia* is the contextual definition of some of the concepts appearing in mathematics. In other words, not every single mathematical concept is individually defined. Rather, there are concepts that receive a definition only in the context of a proposition in which they appear. The philosophical and technical tools for dealing with contextual definitions was given by the theory of denoting

(Russell 1905; see de Rouilhan 1996; Hylton 1990). The theory of denoting allowed Russell to account for denoting phrases without having to assume that denoting phrases necessarily refer to an object. A denoting phrase is given by a list of examples. The examples include "a man, some man, any man, every man, all men, the present King of England, the present King of France." Whether a phrase is denoting depends solely on its form. However, whether a denoting phrase successfully denotes something does not depend merely on its form. Indeed, although "the present King of England" and "the present King of France" have the same form, only the first one denotes an object (at the time Russell is writing). Expressions of the form "the so-and-so," a very important subclass of denoting expressions, are called definite descriptions. Russell's theory consisted in parsing a definite description such as "the present King of France is bald" as "there exists a unique x such that x is King of France and x is bald." In this way "the so-and-so" is meaningful only in the context of a sentence and does not have meaning independently: "According to the view which I advocate, a denoting phrase is essentially part of a sentence, and does not, like most single words, have any significance on its own account" (Russell 1905, 1973, 113). It is hard to overestimate the importance of this analysis for the foundations of mathematics, as denoting phrases, and definite descriptions in particular, are ubiquitous in mathematical practice. In *Principia*, Russell and Whitehead will talk of "incomplete symbols" which do not have an independent definition but only a "definition in use," which determines their meaning only in relation to the context in which they appear. We are now ready to discuss the basic structure of the ramified theory of types.

2.4. Russell's Ramified Type Theory

Poincaré's criticism of impredicative definitions forced Russell and Whitehead to reconsider some of the work they had previously carried out. In particular, Poincaré had criticized the proof of mathematical induction (due to Russell) presented in Whitehead (1902). Poincaré found the definition of an inductive number as the intersection of all recurrent classes (i.e., a class containing zero and closed under successor) to be impredicative. Russell agreed with Poincaré's claim that a vicious circle is present in impredicative definitions and, as we mentioned, presented several theories as possible solutions for the problems raised by the paradoxes (Russell 1906b). Among the theories developed in this period, the substitutional theory (an implementation of the no-classes theory) has been recently subjected to detailed scrutiny (see de Rouilhan 1996, Landini 1998). However, these theories were eventually abandoned and it was the theory of types, as presented in (1908) and (1910), that became Russell's final choice for a solution to the paradoxes. Let us follow the exposition of Russell (1908) to convey the basic ideas of ramified type theory. Russell begins with a long list of paradoxes: Epimenides ("the liar paradox"), Russell's paradox for classes, Russell's paradox for relations, Berry's paradox on "the least integer not nameable in less than nineteen syllables," the paradox of "the least undefinable ordinal," Richard's paradox, and Burali-Forti's contradiction. Russell detects a common feature to all these paradoxes, which consists in the occurrence of a certain "self-reference or reflexiveness": "Thus all our contradictions have in common the assumption of a totality such that, if it were legitimate, it would at once be enlarged by new members defined in terms of itself" (Russell 1908, 155). Thus, the rule adopted by Russell for avoiding the paradoxes, known as the vicious circle principle, reads: "whatever involves *all* of a collection must not be one of a collection." Russell gives several formulations of the principle. A different formulation reads: "If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total" (Russell 1908, 155).[22]

Notice that the vicious circle principle implies that "no totality can contain members defined in terms of itself." This excludes impredicative definitions. However, Russell insists that the principle is purely negative and that a satisfactory solution to the paradoxes must be the result of a positive development of logic. This development of logic is the ramified theory of types. The second remark concerns the issue of when collections can be considered as having a total. By claiming that a collection has no total, Russell means that statements about *all* its members are nonsense. This leads Russell to a lengthy analysis of the difference between "any" and "all." For him, the condition of possibility for saying something about all objects of a collection rests on the members of that collection as being of the same type. The partition of the universe into types rests on the intuition that to make a collection, the objects collected must be logically homogeneous. The distinction between "all" and "any" is expressed, roughly, by the use of a universally bound variable—which ranges over a type—versus a free variable, whose range is not bounded by a type.

In this way we arrive at the core of the ramified theory of types. Unfortunately, the exposition of the theory, both in 1908 and in *Principia*, suffers from the lack of a clear presentation.[23] We will not give a detailed technical exposition here, but only try to convey the gist of the theory with reference to the effect of the theory on the structuring of the universe into types. The distinction into types, however, can also be applied to propositions and propositional functions.

A type is defined by Russell as "the range of significance of a propositional function, that is, as the collection of arguments for which the said function has values" (Russell 1908, 163). We begin with the lowest type, which is simply the class of individuals. In 1908 the individuals are characterized negatively as being devoid of logical complexity, and hence as different from propositions and propositional functions. This is important to exclude the possibility that quantification over individuals might already involve a vicious circle. Type 1 will contain all the (definable) classes of individuals; type 2 all the (definable) classes of classes of individuals; and so on. What we have described is a form of the simple theory of types. This theory already takes care of some of the paradoxes. For instance, if x is an object of type n and y an object of type $n+1$ it makes sense to write $x \in y$, but it makes no sense to write $x \in x$. Thus, in terms of class existence we already exclude the formation of problematic classes at the syntactic level by declaring that expressions of the form $x \in y$ are significant only if x is of type n, for some n, and y is of type $n+1$. This significantly restricts the classes that can be formed.

However, the simple theory of types is not enough to guarantee that the vicious circle principle is satisfied. The complication arises due to the following possibility. One might define a class of a certain type, say, n, by quantifying, in the propositional function defining the class, over collections of objects which might be of higher type than the one being defined. It is thus essential to keep track of the way classes are defined and not only, so to speak, of their ontological complexity.[24] This leads to a generalized notion of **type** (boldface, to distinguish it from type as in the simple theory) for the ramified theory. Rather than giving the formal apparatus for capturing the theory, we will exemplify the main intuition by considering a few examples.

Type 0: the totality of individuals.

Type 1.0: the totality of classes of individuals that can be defined using only quantifiers ranging over individuals (**type** 0).

Type 2.0: the totality of classes of classes of individuals that can be defined by using only quantifiers ranging over objects of **type** 1.0 and **type** 0.

Type 2.1.0: the totality of classes of classes of individuals of **type** 1.0 that can be defined using only quantifiers ranging over elements in **type** 1.0 and in **type** 0

And so on. Let us say that **type** 0 corresponds to order 0, **type** 1.0 to order 1, and that **type** 2.1.0 and **type** 2.0 are of order 2.

This system of **types** satisfies the vicious circle principle, as defining an object by quantifying over a previously given totality will automatically give a class of higher **type**. But this also implies that the development of mathematics in the ramified theory becomes unnatural. In particular, real numbers will appear at different stages of definition. For instance, given a class of real numbers bounded above, the least upper bound principle will, in general, generate a real number of higher **type** (as the definition of the least upper bound requires a quantification over classes containing the given class of reals). To provide a workable foundation for analysis, Russell is then forced to postulate the so-called axiom of reducibility. For its statement we need the notion of a *predicative propositional function* (notice that this notion of predicative is not to be confused with that which is at stake in impredicative definitions). A propositional function $\phi(x)$ is predicative if its order is one higher than that of its argument. To use the foregoing examples, **type** 1.0 and **type** 2.1.0 are predicative, but **type** 2.0 is not. The axiom of reducibility says that each propositional function is extensionally equivalent to a predicative function. Since predicative functions occupy a well-specified place in the hierarchy of **types**, the axiom has the consequence of rendering many of the **types** redundant, at least extensionally. Thus, to go back to our example, the axiom implies that all classes of **type** 2.0 are all extensionally equivalent to classes in **type** 1.0. The net effect of the axiom for the foundations of the real numbers is that it reestablishes the possibility of treating the reals as being all at the same level. In particular, the least upper bound of a class of reals will also be given, extensionally, at the same level as the class used in generating it. However, it has been often observed (most notably in Ramsey 1925), that the axiom of reducibility defeats the purpose of having a ramified hierarchy in the first place. Indeed, with the axiom of reducibility, the ramified theory is equivalent to a form of simple type theory.

2.5. The Logic of *Principia*

Russell and Whitehead's project consisted in showing that all of mathematics could be developed through appropriate definitions in the system of logic defined in *Principia*. One must distinguish here between the development of arithmetic, analysis, and set theory on the one hand and the development of geometry on the other hand. Indeed, for the former theories the axioms of the theory are supposed to come out to be logical theorems of the system of logic, thereby showing that arithmetic, analysis, and set theory are basically developments of pure logic. However, the logicist reconstruction of these branches of mathematics could only be carried out by assuming the axioms of choice ("the multiplicative axiom"), infinity, and reducibility among the available "logical" principles. This is one of the major reasons for the worries about the prospects of logicism in the twenties and thirties (see Grattan-Guinness 2000).

The situation for geometry, whose development was planned for the fourth volume of *Principia* (never published), is different. The approach there would have been a conditional one. The development of geometry in the system of logic given in *Principia* would have shown that the theorems of geometry can be obtained in the system of *Principia* under the assumption of the axioms of geometry. As these axioms say something about certain specific types of relations holding for the geometrical spaces in question, the development of geometry would result in conditional theorems of the logic of *Principia* with the form "if A then p," where A expresses the set of geometrical axioms in question and p is a theorem of geometry.

In both cases, the inferential patterns must be regulated by a specific set of inferential rules. The development of mathematical logic presented in part I of *Principia* (85–326) divides the treatment into three sections. Section A deals with the theory of deduction and develops the propositional calculus. Section B treats the theory of apparent variables (i.e., quantificational logic for types) and sections C, D, and E the logic of classes and relations. While the treatment is supposed to present the whole of logic, its organization already permits one to isolate interesting fragments of the logic presented. In particular, the axiomatization of propositional logic presented in section A of part I is the basis of much later logical work. Russell and Whitehead take the notion of negation and disjunction as basic. They define material implication, $A \supset B$, as $\sim A \vee B$. The axioms for the calculus of propositions are:

1. Anything implied by a true premiss is true.
2. $\vdash :p \vee p. \supset .p \vee q$
3. $\vdash :q. \supset .p \vee q$
4. $\vdash :p \vee q. \supset .q \vee p$
5. $\vdash :p \vee (q \vee r). \supset .q \vee (p \vee r)$
6. $\vdash :.q \supset r. \supset :p \vee q. \supset .p \vee r$

The sign "$\vdash$" is the sign of assertibility (taken from Frege), and the dotted notation (due to Peano) is used instead of the now common parentheses. The only rule of inference is modus ponens; later Bernays pointed out the need to make explicit the rule of substitution, used but not explicitly stated in *Principia*. The quantificational part cannot be formalized as easily due to the need to specify in detail the type theoretic structure. This also requires checking that the propositional axioms presented remain valid when the propositions contain apparent variables (see Landini 1998 for a careful treatment).

Among the primitive propositions of quantificational logic is the following:

$$\vdash :\phi x. \supset .(\exists z).\phi z \tag{9.1}$$

About it, Russell and Whitehead say that "practically, the above primitive proposition gives the only method of proving 'existence-theorems': to prove such theorems, it is necessary (and sufficient) to find some instance in which an object possesses the property in question" (1910, 131). This is, however, wrong and it will be a source of confusion in later debates (see Mancosu 2002).

2.6. Further Developments

The present itinerary on Russell does not aim at providing a full overview of either Russell's development in the period in question nor of the later discussion on the nature of logicism. The incredible complexity of Russell's system and the wealth of still unpublished material make the first aim impossible to achieve here. As evidenced by the citations throughout this itinerary (limited to the major recent books), in the past decade there has been an explosion of scholarly work on Russell's contributions to logic and mathematical philosophy. Moreover, the history of logicism as a program in the foundations of mathematics in the 1920s would require a book on its own.[25] We thus conclude with a general reflection on the importance of *Principia* for the development of mathematical logic proper.

It is hard to overestimate the importance of *Principia* as the first worked-out example of how to reconstruct in detail from a limited number of basic principles the main body of mathematics (even though *Principia*, despite its length, does not even manage to treat the calculus in full detail). However, it became evident that a number of problematic principles—such as infinity, choice, and reducibility—were needed to carry out the reconstruction of mathematics within logic. These existential principles were not obviously logical, and in the case of reducibility seemed rather ad hoc. The further development of logicism in the twenties can be seen as an attempt to work out a solution to such problems. One possible solution was to simply reject the axiom of reducibility and accept that not all of classical mathematics could be obtained in the ramified theory of types. This was the strategy pursued by Chwistek in a number of articles from the early twenties. A second solution was offered by Ramsey's radical rethinking of the logicist project. Ramsey (1925) distinguished between mathematical and semantical antinomies. The former have to do with concepts of mathematics, which are purely extensional whereas the latter involve intensional notions, like definability, which do not belong to mathematics. By refusing to consider the semantical antinomies of relevance to mathematics, Ramsey was able to propose a simple theory of types that could account for classical mathematics and which he claimed took care of all the mathematical antinomies. This, however, came at the cost of excluding intensional notions from the realm of logic.

However, it can be said that despite their interest for the history of logicism, these developments did not, properly speaking, affect the development of mathematical logic for the period we are considering. What was the influence of *Principia* for developments in mathematical logic in the 1910s?

First of all, we have a number of investigations related to the propositional part of *Principia*. Among the results to be mentioned are Sheffer's (re)discovery (1913) of the possibility of defining all Boolean propositional connectives starting from the notion of incompatibility (Sheffer's stroke). Using Sheffer's stroke, Nicod (1916–1919) was able to provide an axiomatization of the propositional calculus with only one axiom. This work was generalized in the early twenties in Göttingen by extending it to the quantificational part of the calculus. This development also marks the beginning of combinatory logic. A systematic analysis of the propositional part of *Principia* was also carried out in Bernays's *Habilitationsschrift* (1918). Much of this work required a metamathematical approach to logic, which was absent from *Principia* (on all this, see itineraries V and VIII).[26] *Principia* was also influential in the development of systems of logic that were strongly opposed to some of the major assumptions therein contained. In the 1910s the most important work in this direction was Lewis's development of systems of strict implication (Lewis 1918).

However, the major influence of *Principia* might simply be that of having established higher-order logic as the paradigm of logic for the next two decades. While it is true that first-order logic emerges as a (more or less) natural fragment of *Principia* (see itinerary IV) most logicians well into the thirties (Carnap, Gödel, Tarski, Hilbert–Ackermann) still considered higher-order logic the appropriate logic for formalizing mathematical theories (see Ferreiros 2001 for extensive treatment).

3. Itinerary III. Zermelo's Axiomatization of Set Theory and Related Foundational Issues

The history of set theory during the first three decades of the twentieth century has been extensively researched. One area of investigation is the history of set theory as a mathematical discipline and its influence on other areas of mathematics. A second important topic is the relationship between logic and set theory. Finally, much attention has been devoted to the axiomatizations of set theory, and even to the pluralities of set theories (naive set theory, Zermelo, von Neumann, intuitionistic set theory, etc.). Here we focus on Zermelo's axiomatization.

3.1. The Debate on the Axiom of Choice

At the beginning of the century, set theory had already established itself both as an independent mathematical theory as well as in its applications to other branches of mathematics, in particular analysis.[27] In his address to the mathematical congress in Paris, Hilbert singled out the continuum problem as one of the major problems for twentieth-century mathematics. One of the problems that had occupied Cantor, which he was never able to solve, was that of whether every set is an aleph, or equivalently, whether every set can be well ordered. Julius König (1904) presented a proof at the third International Congress of Mathematicians in Heidelberg claiming that the continuum cannot be well ordered. A key step of the proof made use of a result by Felix Bernstein claiming that $\aleph_\alpha^{\aleph_\beta} = 2^{\aleph_\beta}\aleph_\alpha$. But after scrutinizing Bernstein's result in the wake of König's talk, Hausdorff (1904) showed that it holds only when α is a successor ordinal. Soon thereafter, Zermelo showed that every set can be well ordered (Zermelo 1904).[28] Let us recall that an ordered set F is well ordered if and only if every nonempty subset of it has a least element (under the ordering). Zermelo's proof appealed to

> the assumption that coverings γ actually do exist, hence upon the principle that even for an infinite totality of sets there are always mappings that associate with every set one of its elements, or, expressed formally, that the product of an infinite totality of sets, each containing at least one element, itself differs from zero. This logical principle cannot, to be sure, be reduced to a still simpler one, but is applied without hesitation everywhere in mathematical deduction. (Zermelo 1904, 141)[29]

Let M be the arbitrary set for which a well ordering needs to be established. A covering γ for M in Zermelo's proof is what we would call a choice function, which for an arbitrary subset M' of a set M yields an element $\gamma(M')$ of M, called the distinguished element of M'. It is under the assumption of existence of such a covering that Zermelo establishes the existence of special sets called

γ-sets. A γ-set is a set M_γ included in M which is well ordered and such that if $a \in M_\gamma$ and if $A = \{ x : x \in M_\gamma$ and $x < a$ in the well ordering of $M_\gamma \}$, then a is the distinguished element of $M - A$ according to the covering γ. Zermelo then shows that the union of all γ-sets, L_γ, is a γ-set and that $L_\gamma = M$. Thus M can be well ordered.

Zermelo's proof immediately gave rise to a major philosophical and mathematical discussion.[30] The main exchange was published by the *Bulletin de la Société Mathématique de France* in 1905 and consisted of five letters exchanged among Baire, Borel, Lebesgue, and Hadamard (1905). Baire, Borel, and Lebesgue shared certain constructivist tendencies, which led them to object to Zermelo's use of the principle of choice, although in their actual mathematical practice they often made use (implicitly or explicitly) of Cantorian assumptions, including the principle of choice. For instance, Lebesgue's proof of the countable additivity of the measurable subsets of the real line relies on the principle of choice for countable collections of sets. Hadamard took a more liberal stand.

The debate began with an article by Borel, which appeared in *Mathematische Annalen* (Borel 1905). Borel claimed that Zermelo's proof had only shown the equivalence between the well-ordering problem for an arbitrary set M and the problem of choosing an arbitrary element from each subset of M. However, Borel did not accept this as a solution to the first problem, for the postulation of a choice function required by Zermelo was, if anything, even more problematic than the problem one began with. He found the application of the principle to uncountably many sets particularly problematic, but allowed for the possibility that the principle might be acceptable when we are dealing with countable collections of sets. Hadamard's reply to Borel's article defended Zermelo's principle. In the process of defending Zermelo's application of the principle, Hadamard also drew a few important distinctions. For instance, he distinguished between reasonings in which each choice depends on the previous ones (dependent choice) from Zermelo's principle, which postulated simultaneous independent choices. Moreover, he objected to Borel that he saw no essential difference between postulating the principle for a countable or an uncountable collection of sets. Finally, he also pointed at the fact that one had to distinguish between whether the choice could be made "effectively" or simply postulated to exist. He emphasized the essential difference between showing that an object (say, a function) exists, without however specifying the object, and actually providing a unique specification of the object. Hadamard claimed that whether one raises the first or the second problem essentially changes the nature of the mathematical question being investigated. The most radical position was taken by Baire, who defended a strong finitism and refused to accept one of the basic principles underlying Zermelo's proof. Indeed, he claimed that if a set A is given it does not follow that the set of its subsets can also be considered as given. And thus, he rejected that part of Zermelo's argument that allowed him to pick an element from every subset of the given set. Baire claimed that Zermelo's principle was consistent but that it simply lacked mathematical meaning. Lebesgue's point of view also emphasized the issue of definability of mathematical objects. He asked: "Can one prove the existence of a mathematical object without defining it?" He also defended a constructivist attitude and claimed that the only true claim of existence in mathematics must be obtained by defining the object uniquely. In the last of the five letters, Hadamard rejected the constructivist positions of Baire, Borel, and Lebesgue and claimed that mathematical existence does not have to rely on unique definability. He clearly set out the two different conceptions of mathematics that were at the source of the debate. On one conception, the constructivist one, mathematical objects are said to exist if they can be defined or constructed. On the other conception, mathematical existence is not dependent on our abilities to either construct or define the object in question. While allowing the reasonableness of the constructivist position, Hadamard considered it to rely on psychological and subjective considerations that were foreign to the true nature of mathematics.

The debate focused attention not only on the major underlying philosophical issues but also on the important distinctions that one could draw between different forms of the principle of choice. The positions of Baire, Borel, and Lebesgue on definability remained vague but influenced later work by Weyl, Skolem, and others.

Zermelo's proof was widely discussed and criticized. In the article "A new proof of the possibility of a well-ordering" (1908b), Zermelo gave a new proof of the well-ordering theorem, by relying on a generalization of Dedekind's chains, and gave a full reply to the criticism that had been raised against his previous proof (by, among others, Borel, Peano, Poincaré, König, Jourdain, Bernstein, and Schoenflies). We focus on Poincaré's objections.

Poincaré's criticism of Zermelo's proof occurred in his discussion (1906) of logicism and set theory. In particular, he had objected to the formation of impredicative sets which occur in the proof. Recall that in the final part of the first proof Zermelo defined the set L_γ as the union of all γ-sets, that is,

$$L_\gamma = \{\, x : \text{for some } \gamma\text{-set } Y,\ x \in Y \,\}.$$

According to Poincaré, this definition is objectionable because to determine whether an element x belongs to L_γ, one needs to go through all the γ-sets. But among the γ-sets is L_γ itself, and thus a vicious circle is involved in the procedure. Zermelo replied to Poincaré claiming that his critique would "threaten the existence of all of mathematics" (Zermelo 1908b, 198). Indeed, impredicative definitions and procedures occur not only in set theory but in the most established branches of mathematics, such as analysis:

> Now, on the one hand, proofs that have this logical form are by no means confined to set theory; exactly the same kind can be found in analysis wherever the maximum or the minimum of a previously defined "completed" set of numbers Z is used for further inferences. This happens, for example, in the well-known Cauchy proof of the

> fundamental theorem of algebra, and up to now it has not occurred to anyone to regard this as something illogical. (Zermelo 1908b, 190–191)

Poincaré claimed that there was an essential difference between Cauchy's proof (in which the impredicativity is eliminable) and Zermelo's proof. This debate forced Poincaré to be more explicit on his notion of predicativity (see Heinzmann 1985) and contributed to Zermelo's spelling out of the axiomatic structure of set theory. After presenting the axioms of Zermelo's set theory we will return to the issue of impredicativity.

3.2. Zermelo's Axiomatization of Set Theory

Another set of objections that were raised against Zermelo's proof suggested the possibility that Zermelo's assumption might end up generating the set of all ordinals W and therefore fall prey to Burali-Forti's antinomy.[31] Zermelo claimed that a suitable restriction of the notion of set was enough to avoid the antinomies and that in 1904 he had restricted himself "to principles and devices that have not yet by themselves given rise to any antinomy" (Zermelo 1908b, 192). These principles were the subject of another article that contains the first axiomatization of set theory (Zermelo 1908c). Zermelo begins by claiming that no solution to the problem of the paradoxes has yielded a simple and convincing system. Rather than starting with a general notion of set, he proposes to distill the axioms of set theory out of an analysis of the current state of the subject. The treatment has to preserve all that is of mathematical value in the theory and impose a restriction on the notion of set so that no antinomies are generated. Zermelo's solution consists in an axiom system containing seven axioms. The main intuition behind his approach to set theory is one of "limitation of size," that is, sets which are "too large" will not be generated by the axioms. This is ensured by the separation axiom, which in essence restricts the possibility of obtaining new sets only by isolating (definable) parts of already given sets. Following Hilbert's axiomatization of geometry, Zermelo begins by postulating the existence of a domain $\mathfrak{B}$ of individuals, among which are the sets, on which some basic relations are defined. The two basic relations are equality ($=$) and membership ($\in$). For sets A and B, A is said to be a subset of B if and only if every element of A is an element of B. The key definition concerns the notion of definite property:

> A question or assertion $\mathfrak{E}$ is said to be *definite* if the fundamental relations of the domain, by means of the axioms and the universally valid laws of logic, determine without arbitrariness whether it holds or not. Likewise a "propositional function" $\mathfrak{E}(x)$, in which the variable term x ranges over all individuals of a class $\mathfrak{K}$, is said to be definite if it is definite for *each single* individual x of the class $\mathfrak{K}$. (Zermelo 1908c, 201)

This definition plays a central role in the axiom of separation (see the following) which forms the cornerstone of Zermelo's axiomatic construction. However, the notion of a propositional function being "definite" remained unclarified and Zermelo did not specify what "the universally valid laws of logic" are. This lack of clarity was immediately seen as a blemish of the axiomatization; it was given a satisfactory solution only later by, among others, Weyl and Skolem. Let us now list the axioms in Zermelo's original formulation.

Axiom I (Axiom of extensionality). If every element of a set M is also an element of N and vice versa, if, therefore, both $M \mathrel{\underline{\in}} N$ and $N \mathrel{\underline{\in}} M$, then always $M = N$; or, more briefly: Every set is determined by its elements. ...

Axiom II (Axiom of elementary sets). There exists a (fictitious) set, the null set, 0, that contains no element at all. If a is any object of the domain, there exists a set $\{a\}$ containing a and only a as element; if a and b are two objects of the domain, there always exists a set $\{a, b\}$ containing as elements a and b but no object x distinct from both. ...

Axiom III (Axiom of separation). Whenever the propositional function $\mathfrak{E}(x)$ is definite for all elements of a set M, M possesses a subset $M_{\mathfrak{E}}$ containing as elements precisely those elements x of M for which $\mathfrak{E}(x)$ is true. ...

Axiom IV (Axiom of the power set). To every set T there corresponds a set $\mathfrak{U}T$, the *power set* of T, that contains as elements precisely all subsets of T.

Axiom V (Axiom of the union). To every set T there corresponds a set $\mathfrak{S}T$, the *union* of T, that contains as elements precisely all elements of the elements of T. ...

Axiom VI (Axiom of choice). If T is a set whose elements are all sets that are different from 0 and mutually disjoint, its union $\mathfrak{S}T$ includes at least one subset S_1 having one and only one element in common with each element of T. ...

Axiom VII (Axiom of infinity). There exists in the domain at least one set Z that contains the null set as an element and is so constituted that to each of its element a there corresponds a further element of the form $\{a\}$, in other words, that with each of its elements a it also contains the corresponding set $\{a\}$ as an element. (Zermelo 1908c, 201–204)

Let us clarify how Zermelo's axiomatization manages to exclude the generation of the paradoxical sets and at the same time allows the development of classical mathematics, including the parts based on impredicative definitions. Previous developments of set theory operated with a comprehension principle that allowed, given any property $P(x)$, the formation of the set of objects satisfying $P(x)$, that is, $\{\, x : P(x) \,\}$. This unrestricted use of comprehension leads to the possibility of forming Russell's paradoxical "set" of all sets that do not contain themselves as elements, or the "set" of all ordinals W. However, the separation principle essentially restricts the formation of sets by requiring that sets be obtained, through some propositional function $P(x)$, as subsets of previously given sets. Thus, to go back to Russell's set, it is not possible to construct $\{\, x : \sim(x \in x) \,\}$ but only, for a previously given set A, a set $B = \{\, x \in A : \sim(x \in x) \,\}$. Unlike the former, this set is innocuous and does not give rise to an antinomy. In the same way, we cannot form the set of all ordinals but only, for any given set A, the set of ordinals in A. The paradoxes having to do with notions such as denotation and definability, such as Berry's or König's paradoxes, are excluded because the notions involved are not "definite" in the sense required for Axiom III. Zermelo's approach here foreshadows the distinction, later drawn by Ramsey (1925), between mathematical and semantical paradoxes, albeit in a somewhat obscure way. In his essay, Zermelo pointed out that the entire theory of sets created by Cantor and Dedekind could be developed from his axioms, and he himself carried out the development of quite a good amount of cardinal arithmetic.

To connect our discussion to the debate on impredicative definitions let us look more closely at the principles of Zermelo's system that allow the formation of impredicative definitions. We shall consider one classic example, namely, the definition of natural numbers according to Dedekind's theory of chains.

In *Was sind und was sollen die Zahlen?* (1888), Dedekind had given a characterization of the natural numbers starting from the notion of a chain. First he argued, in a notoriously fallacious way, that there are simply infinite systems (or sets), that is, sets that can be mapped one-one into a proper subset of themselves. Then he showed that each simply infinite system S contains an (isomorphic) copy of a K-chain, that is, a set that contains 1 and is closed under successor. Finally, the set of natural numbers is defined as the intersection of all K-chains contained in a simply infinite system. This is the smallest K-chain contained in S. From the logical point of view, the definition of the natural numbers by means of an intersection of sets corresponds to a universal quantification over the power set of the infinite system S. More formally, $N = \{\, X : X \mathrel{=\!\!\!\in} S$ and X is a chain in $S \,\}$. Equivalently, $n \in N$ iff n is a member of all chains in S.

In Zermelo's axiomatization of set theory, the definition of N is justified by appealing to three axioms. First of all, the existence of an infinite simple system S is given through the axiom of infinity. By means of the power set axiom we are also given the set of subsets of S. Finally, we appeal to the separation axiom to construct the intersection of all chains in S.

It thus appears that the formalization of set theory provided by Zermelo had met the goals he set for himself. On the one hand, the notion of set was restricted in such a way that no paradoxical sets could arise. On the other hand, no parts of classical mathematics seemed to be excluded by its formalization. Zermelo's axiomatization proved to be an astounding success. However, there were problems left. Subsequent discussion showed the importance of the issue of definability, and further results in set theory showed that Zermelo's axioms did not quite characterize a single set-theoretic universe. This will be treated in the next section.

3.3. The Discussion on the Notion of "Definit"

One important contribution to the clarification of Zermelo's notion of "definit" came already in Weyl's "Über die Definitionen der mathematischen Grundbegriffe" (1910). After reflecting on the process of "Logisierung der Mathematik," Weyl declares in this paper that from the logical point of view, set theory is the proper foundation of the mathematical sciences. Thus, he adds, if one wants to give general definitional principles that hold for all of mathematics, it is necessary to account for the definitional principles of set theory. First, he begins his definitional analysis with geometry. Relying on Pieri's work on the foundations of geometry, he starts with two relations, $x = y$ and $E(x, y, z)$. $E(x, y, z)$ means that y and z are equidistant from x. Then he adds that all definitions in Pieri's geometry can be obtained by closing the basic relationships under five principles:

1. Permutation of variables: if $\mathfrak{A}(x, y, z)$ is a ternary relation, so is $\mathfrak{A}(x, z, y)$.
2. Negation: if $\mathfrak{A}$ is a relation, then not-$\mathfrak{A}$ is also a relation.
3. Addition: if $\mathfrak{A}(x, y, z)$ is a ternary relation, then $\mathfrak{A}{+}(x, y, z, w)$ is a relation, which holds of x, y, z, w, iff $\mathfrak{A}(x, y, z)$ holds.
4. Subtraction: if $\mathfrak{A}(x, y, z)$ is a relation, then so is $\mathfrak{B}(x, y)$, which holds iff there exists a z such that $\mathfrak{A}(x, y, z)$
5. Coordination: if $\mathfrak{A}(x, y, z)$ and $\mathfrak{B}(x, y, z)$ are ternary relations, so is $\mathfrak{C}(x, y, z)$, which holds if and only if both $\mathfrak{A}(x, y, z)$ and $\mathfrak{B}(x, y, z)$ hold.

For Weyl, these definitional principles are sufficient to capture all the concepts of elementary geometry. In the later part of the article, Weyl poses the question: Can all the concepts of set theory be obtained from $x = y$ and $x \in y$ by closing under the definitional principles 1–5? Here his reply is negative. He claims that the fact that in set theory we have objects that can be characterized uniquely, such as the empty set, presents a situation very different from the geometrical one, where all the points are equivalent. He adds that the definitional principles 1–5 would have to be altered to take care of this situation. However the definitional principles still play an important role in connection to Zermelo's concept of "definit." After pointing out the vagueness of Zermelo's formulation of the comprehension principle he proposes an improvement: "A *definite relation* is one that can be defined from the basic relationships $=$ and $\in$ by finitely many applications of our definitional principles modified in an appropriate fashion" (Weyl 1910, 304). The comprehension principle is then stated not for arbitrary propositional functions, as in Zermelo, but in the restricted form for binary relationships: "If M is an arbitrary set, a an arbitrary object, and $\mathfrak{A}$ is a definite binary relationship, then the elements

x of M which stand in the relationship $\mathfrak{A}$ to the object a constitute a set" (Weyl 1910, 304). In a note to the text, Weyl also expresses his conviction that without a precise formulation of the definitional principles the solution of the continuum problem would not be possible. Weyl's attempt at making precise the notion of definite property is important because, despite a few remaining obscure points, it clearly points the way to a notion of definability based on closure under Boolean connectives and existential quantification over the individuals of the domain (definition principle 4). In *Das Kontinuum* (1918), the analysis of the mathematical concept formation is presented as an account of the principles of combination of judgments with minor differences from the account given in 1910. However, the explicit rejection of the possibility of quantifying over (what he then calls) ideal elements, that is, sets of elements of the domain, which characterizes Weyl's predicative approach in 1918, brings his approach quite close to an explicit characterization of the comprehension principle in terms of first-order definability.[32]

Two very important contributions to the problem of "definiteness" were given by Fraenkel (1922b) and Skolem (1922). The more influential turned out to be Skolem's account. Here is the relevant passage from Skolem's work:

> A very deficient point in Zermelo is the notion "definite proposition." Probably no one will find Zermelo's explanations of it satisfactory. So far as I know, no one has attempted to give a strict formulation of this notion; this is very strange, since it can be done quite easily and, moreover, in a very natural way that suggests itself. (Skolem 1922, 292)

Skolem then listed "the five basic operations of mathematical logic": conjunction, disjunction, negation, universal quantification, existential quantification. His proposal is that "by a definite proposition we now mean a finite expression constructed from elementary propositions of the form $a \in b$ or $a = b$ by means of the five operations mentioned" (292–293). The similarity to Weyl's account is striking. Although Skolem does not mention Weyl (1910), he was familiar with it, as he had reviewed it for the *Jahrbuch für die Fortschritte der Mathematik* (Skolem 1912).

One final point should be mentioned in connection to these debates on the notion of "definit." Weyl, already in 1910, had pointed out that the appeal to a finite number of applications of the definitional principles showed that the notion of natural number was essential to the formulation of set theory, which however was supposed to provide a foundation for all mathematical concepts (including that of natural number). In *Das Kontinuum*, he definitely takes the stand that the concept of natural number is basic, and that set theory cannot give a foundation for it (Weyl 1918, 24). Zermelo took the opposite stand. Analyzing Fraenkel's account of "definit" (1929), he rejected it on account of the fact that an explicit appeal to the notion of finitely many applications of the axiom was involved. But the notion of finite number should be given a foundation by set theory, which therefore cannot presuppose it in its formulation (see also Skolem 1929a).

Thus, two major problems emerged in the discussion concerning a refinement of the notion of "definit." The first concerned the question of whether set theory could be considered a foundation of mathematics. Both Skolem and Weyl (who had abandoned his earlier position) thought that this could not be the case. The second problem had to do with the choice of the formal language. Why restrict oneself to first-order logic as Skolem and Weyl were proposing? Why not use a stronger language? The problem was of course of central significance due to the relativization of set-theoretical notions that Skolem had pointed out in his 1922 paper (see itinerary IV). This topic would be worthy of further discussion, but we limit ourselves here to two observations. First of all, Skolem used the relativity of set-theoretic notions as an argument against considering set theory as a foundation of mathematics. Second, Zermelo proposed in "Über Stufen der Quantifikation und die Logik des Unendlichen" (1931) an infinitary logic with the aim of countering Skolem's position (which Zermelo called, disparagingly, "Skolemism").[33] As Ferreiros (1999, 363) argues, it was only after Gödel's incompleteness results that the idea of using first-order logic as the "natural" logical scaffolding for axiomatic set theory became standard.

3.4. Metatheoretical Studies of Zermelo's Axiomatization

In treating set theory as an axiomatic system, Zermelo had opened the way for a study of the metatheoretical properties of the system itself, such as independence, consistency, and categoricity of the axioms. It should be said from the outset that no real progress was made on the issue of consistency. A proof of the consistency of set theory was one of the major goals of Hilbert's program, but it was not achieved. Of course, much attention was devoted to the axiom of choice. The Polish set-theorist Sierpinski (1918) listed a long set of propositions which seemed to require the axiom of choice essentially, or which were equivalent to the axiom of choice. But was the axiom of choice itself indispensable, or could it be derived from the remaining axioms of Zermelo's system?[34] While this problem was only solved by the combined work of Gödel (1940) and Cohen (1966), an interesting result on independence was obtained by Fraenkel (1922b). Fraenkel was able to show that the axiom of choice is independent of the other axioms of Zermelo's set theory, if we assume the existence of infinitely many urelements, that is, basic elements of the domain $\mathfrak{B}$ which possess no elements themselves. Unfortunately, the assumption of a denumerable set of urelements is essential to the proof, and thus the result does not apply immediately to Zermelo's system. Moreover, there were reasons to consider the assumption of urelements as foreign to set theory. Fraenkel himself (1922c) had criticized the possibility of having urelements as part of the domain $\mathfrak{B}$, posited at the outset by Zermelo, as irrelevant for the goal of giving a foundations of mathematics. The possibility of having interpretations of set theory with urelements, and others without, already suggested the inability of the axioms to characterize a unique model. Skolem (1922) (and independently also Fraenkel in the same year) also discusses interpretations of Zermelo's axioms in which there are infinite descending chains $\cdots \in M_2 \in M_1 \in M$, which he called a descending $\in$-sequence, a fact that had already been pointed out by Mirimanoff (1917).[35] A related shortcoming, which affects both the completeness and the categoricity of Zermelo's theory, is related to the inability of the theory to ensure that certain sets, which are used unproblematically in the practice of set theorists, actually exist. Skolem gives the following example. Consider the set M. By the power set axiom we can form $\mathfrak{U}(M)$, then $\mathfrak{U}(\mathfrak{U}(M))$, and so on for any finite iteration of the power set axiom. However, no axiom in Zermelo's set theory allows us to infer the existence of $\{M, \mathfrak{U}(M), \mathfrak{U}(\mathfrak{U}(M)), \dots\}$. Skolem gives an interpretation that satisfies all the axioms of set theory, which contains M and all finite iterations of the power set of M, but in which $\{M, \mathfrak{U}(M), \mathfrak{U}(\mathfrak{U}(M)), \dots\}$ does not exist. Both shortcomings, infinite descending chains and lack of closure at "limit" stages, pointed out important problems in Zermelo's axiomatization. The existence of infinite descending chains ran against the intuitive conception of set theory as built up in a "cumulative" way, and the lack of closure for infinite sets showed that genuine parts of the theory of ordinal and cardinal numbers could not be obtained in Zermelo's system. The latter problem was addressed by Skolem through the formulation of what came to be known as the replacement axiom:

> Let U be a definite proposition that holds for certain pairs (a, b) in the domain B; assume further, that for every a there exists at most one b such that U is true. Then, as a ranges over the elements of a set M_a, b ranges over all elements of a set M_b. (Skolem 1922, 297)

In other words, starting from a set a and a "definite" functional relationship $f(x)$ on the domain, the range of $f(x)$ is also a set. The name and an independent formulation, albeit very informal, of the axiom of replacement is due to Fraenkel (1922c). For this reason Zermelo (1930, 29) calls the theory Zermelo–Fraenkel set theory. However, Fraenkel had doubts that the axiom was really needed for "general set theory." The real importance of the axiom became clear with the development of the theory of ordinals given by von Neumann, who showed that the replacement axiom was essential to the foundation of the theory.[36] Von Neumann (1923) gave a theory of ordinals in which ordinals are specific well-ordered sets, as opposed to classes of equivalent well-orderings. This opened the way for a development of ordinal arithmetic independently of the theory of ordered sets. The definition he obtained is now standard and it was captured by von Neumann in the claim that "every ordinal is the set of the ordinals that precede it." The formalization of set theory he offered in 1925 is essentially different from that of Zermelo. Von Neumann takes the notion of function as basic (the notion of set can be recovered from that of function) and allows classes in addition to sets. This system of von Neumann was later modified and extended by Bernays and Gödel, to result in what is known as NBG set theory.[37] The central intuition is a "limitation of size" principle, according to which there are collections of objects which are too big (we now call them classes), namely, those that are equivalent to the class of all things. The difference between classes and sets is essentially that the latter but not the former can be elements of other sets or classes. A very important part of von Neumann (1925) consists in the axiomatic investigation of "models" of set theory. We will come back to this issue in itinerary VIII. Here it should be pointed out that von Neumann's technique foreshadowed the studies of inner models of set theory.

It is only with von Neumann that a new axiom intended to eliminate the existence of descending $\in$-sequences (and finite cycles) was formulated (1925, 1928) (although Mirimanoff had foreshadowed this development by means of his postulate of "ordinariness" meant to eliminate "extraordinary" sets, that is, infinite descending $\in$-sequences). This was the axiom of well foundedness (von Neumann 1928, 498), which postulates that every (nonempty) set is such that it contains an element with which it has no element in common. The axiom appears in Zermelo (1930) as the *Axiom der Fundierung*:

> Axiom of Foundation: Every (descending) chain of elements, each member of which is an element of the previous one, terminates with a finite index in an urelement. Or, equivalently: Every subdomain T (of a ZF-model) contains at least one element t_0 that has no element t in T. (Zermelo 1930, 31)

Thus by 1930 we have all the axioms that characterize what we nowadays call ZFC, namely, Zermelo–Fraenkel set theory with choice. However, the formulation given by Zermelo (1930) is not first-order, as it relies on second-order quantification in the statement of the axioms of separation and replacement. Even the second formulation of the axiom of foundation contains an implicit quantification over models of ZF.[38]

During the thirties, there were several competing systems for the foundations of mathematics, such as, in addition to Zermelo's extended system, simple type theory and NBG. It was only in the second half of the 1930s that the first-order formulation of ZFC became standard (see Ferreiros 1999, 2001).

4. Itinerary IV. The Theory of Relatives and Löwenheim's Theorem

4.1. Theory of Relatives and Model Theory

Probably the most important achievements of the algebraic tradition in logic are the axiomatization of the algebra of classes, the theory of relatives, and the proof of the first results of a clearly metalogical character. The origin of the calculus of classes is found in the works of Boole. De Morgan was the first logician to recognize the importance of relations to logic, but he did not develop a theory of relations. Peirce established the fundamental laws of the calculus of classes and created the theory of relatives.[39] Schröder proposed the first complete axiomatization of the calculus of classes and expanded considerably the calculus and the theory of relatives. This theory was the frame that made possible the proof by Löwenheim of the first metalogical theorem. "Über Möglichkeiten im Relativkalkül" (1915), the paper in which Löwenheim published these results, is now recognized as one of the cornerstones in the history of logic (or even in the history of mathematics) due to the fact that it marks the beginning of what we call *model theory*.[40]

The main theorems of Löwenheim's paper are (stated in modern terminology): (1) Not every first-order sentence of the theory of relatives is logically equivalent to a quantifier-free formula of the calculus of relatives (proved by Korselt in a letter to Löwenheim); (2) if a first-order sentence has a model, then it has a countable model; (3) there are satisfiable second-order sentences which have no countable model; (4) the unary predicate calculus is decidable; and (5) first-order logic can be reduced to binary first-order logic.

Nowadays, we use the term "Löwenheim–Skolem theorem" to refer to theorems asserting that if a set of first-order sentences has a model of some infinite cardinality, it also has models of some other infinite cardinalities. The mathematical interest of these theorems is well known. They imply, for example, that no infinite structure can be characterized up to isomorphism in a first-order language. Theorem 2 of Löwenheim's paper was the first one of this group to be proved and, in fact, the first in the history of logic that established a nontrivial relation between first-order formulas and their models.

Löwenheim's theorem poses at least two problems to the historian of logic. The first is to explain why the theory of relatives made it possible to state and prove a theorem which was unthinkable in the syntactic tradition of Frege and Russell. The second problem is more specific. Even today, Löwenheim's proof raises many uncertainties. On the one hand, the very result that is attributed to Löwenheim today is not the one that Skolem—a logician raised in the algebraic tradition—appears to have attributed to him. On the other hand, present-day commentators agree that the proof has gaps, but it is not completely clear which they are. We deal with these questions in the following pages.[41]

Schröder was interested in the study of the algebras of relatives. As Peirce and he himself conceived it, an algebra of relatives consists of a domain of relatives (the set of all relatives included in a given universe), the inclusion relation between relatives (denoted by $\Subset$), six operations (union, intersection, complementation, relative product, relative sum, and inversion) and four distinguished elements called *modules* (the total relation, the identity relation, the diversity relation, and the empty class). Schröder's objective was to study these structures with the help of a calculus. He could have tried to axiomatize the calculus of relatives, but, following Peirce, he preferred to develop it within the theory of relatives. The difference between the theory and the calculus of relatives is roughly this. The calculus permits the quantification over relatives, but deals only with relatives and operations between them. The theory of relatives, on the other hand, is an extension which also allows the quantification over individuals. The advantage of the theory over the calculus is that the operations between relatives can be defined in terms of individuals and these definitions provide a simpler and more intuitive way of proving certain theorems of the calculus.

Neither Peirce nor Schröder thought that the theory of relatives was stronger than the calculus. Schröder in particular was convinced that all logical and mathematical problems could be addressed within the calculus of relatives (Schröder 1898, 53). So he focused on developing the calculus and viewed the theory as a tool that facilitated his task. Schröder did not address problems of a metalogical nature, in that he did not consider the relation between the formulas of a formal language and their models. Arguments or considerations of a semantic type are not completely absent from *Vorlesungen über die Algebra der Logik* (henceforth *Vorlesungen*), but they occur only in the proofs of certain equations, and so we cannot view them as properly metalogical.

Schröder posed numerous problems regarding the calculus of relations, but very few later logicians showed any interest in them, and the study of the algebras of relatives was largely neglected until Tarski. In his first paper on the subject, Tarski (1941) claimed that hardly any progress had been made in the previous 45 years and expressed his surprise that this line of research should have had so few followers.[42]

Schröder was not interested in metalogical questions, but the theory of relatives as he conceived it made it possible to take them into consideration. As a preliminary appraisal, we can say that in the theory of relatives two interpretations coexist: an algebraic interpretation and a propositional interpretation. This means that the same expressions can be seen both as expressions of an algebraic theory and as formulas of logic (i.e., as well-formed expressions of a formal language which we may use to symbolize the statements of a theory to reflect its logical structure). We do not mean by this that the whole theory admits of two interpretations, because not all the expressions can be read in both ways, but the point is that some expressions do.

One way of viewing the theory of relatives that gives a fairly acceptable idea of the situation is as a theory of relations together with a partly algebraic presentation of the logic required to develop it.[43] The theory constitutes a whole, but it is important to distinguish the part that deals with the tools needed to construct and evaluate the expressions that denote a truth value (i.e., the fragment that concerns logic) from the one that deals specifically with relatives. So, to prove his theorem, Löwenheim had to think of logic as a differentiated fragment of the theory of relatives and delimit the formal language at least to the extent required to state and prove the theorem.

With the exception of the distinction between object language and metalanguage (an absence that needs emphasizing as it causes many problems in the proof by Löwenheim of his theorem), the basic components of model theory are found in one way or another in the theory of relatives. On the one hand, the part of the theory dealing with logic contains more or less implicitly the syntactic component of a formal language with quantification over relatives: a set of logical symbols with its corresponding propositional interpretation and a syntax borrowed from algebra. On the other, the algebraic interpretation supplies a semantics for this language in the sense that it is enough to evaluate the expressions of this language. In this situation, all that remains to be done to obtain the first results of model theory is, first, become aware that the theory does include a formal language and single it out; second, focus on this language and, in particular, on its first-order fragment; and third, investigate the relationship between the formulas of this language and the domains in which they hold. As far as we know, Löwenheim was the first in the history of logic to concentrate on first-order logic and investigate some of its nontrivial metalogical properties.

4.2. The Logic of Relatives

To understand Löwenheim's proof and the relationship between his paper and the theory of relatives, we need first to present the logic of relatives (i.e., the fragment of the theory that concerns logic).[44] In our exposition, we distinguish syntax from semantics, although such a distinction is particularly alien to the logic of relatives. Consequently, the exposition should not be used to draw conclusions about the level of precision found in Schröder or in Löwenheim.

Strictly speaking, relatives denote relations on the (first-order) domain and they are the only nonlogical symbols of the logic of relatives. However, as a matter of fact, in the writings of the algebraists the word *relative* refers both to a symbol of the language and to the object denoted by it. The only relatives usually taken into account are binary, on the assumption that all relatives can be reduced to binary.[45]

What we would call today *logical symbols* are the following: (a) indices; (b) module symbols: $1'$ and $0'$; (c) operation symbols: $+$, $\cdot$ and $^{-}$; (d) quantifiers: Σ and Π; (e) equality symbol: $=$; and (f) propositional constants: 1 and 0.

Indices play the role of individual variables. As indices the letters h, i, j, k, and l are the most frequently used.

In the theory of relatives, the term *module* is used to refer to any of the four relatives 1, 0, $1'$, and $0'$. The module 1 is the class of all ordered pairs of elements of the (first-order) domain; 0 is the empty class; $1'$ is the identity relation on the domain; and $0'$ is the diversity relation on the domain. In the logic of relatives, $1'$ and $0'$ are used as relational constants and 1 and 0 are not viewed as modules but as propositional constants, denoting the truth values.

There are six operations on the set of relatives: identical sum (union, denoted by $+$), identical product (intersection, denoted by $\cdot$), complement ($^{-}$), relative sum, relative product, and inversion. None of these operations belongs to the logic of relatives. The symbols corresponding to the first three operations are used ambiguously to refer also to the three well-known Boolean operations defined on the set $\{0, 1\}$. This is the meaning they have in the logic of relatives.

If i and j are elements of the domain and a is a relative or a module, then a_{ij} is a *relative coefficient.* For example, the relative coefficients of z in the domain $\{2, 3\}$ are z_{22}, z_{23}, z_{32}, and z_{33}. Relative coefficients can only take two values: the truth values (1 and 0). That is, if a_{ij} is a relative coefficient, then

$$a_{ij} = 1 \quad \text{or} \quad a_{ij} = 0.$$

Relative coefficients admit of a propositional interpretation: a_{ij} expresses that the individual i is in the relation a with the individual j. This interpretation allows us to regard relative coefficients as atomic formulas of a first-order language, but in the logic of relatives they are considered as terms.

If A and B are expressions denoting a truth value, so are $(A + B)$, $(A \cdot B)$, and $\overline{A}$; for example, $(a_{ij} + b_{ij})$, $(a_{ij} \cdot b_{ij})$, and $(\overline{a_{ij}})$ are meaningful expressions of this sort. Terms denoting a truth value admit a propositional reading when the symbols $+$, $\cdot$, and $\overline{}$ occurring in them are viewed as connectives.

The symbols Σ and Π have different uses in the theory of relatives and they cannot be propositionally interpreted as quantifiers in all cases. We restrict ourselves to their use as quantifiers. If u is a variable ranging over elements (or over relatives) and A_u is an expression denoting a truth value in which u occurs, then

$$\sum_u A_u \quad \text{and} \quad \prod_u A_u$$

are, respectively, the sum and the product of all A_u, where u ranges over the domain (or over the set of relatives). From the algebraic point of view, these expressions are terms of the theory, because they denote a truth value. They also admit a propositional reading, Σ can also be interpreted as the existential quantifier and Π as the universal one. For example, $\Sigma_i \Pi_j z_{ij}$ can also be read as "there exists i such that for every j, i is in the relation z with j."[46]

The canonical formulas of the theory of relatives are the equations, that is, the expressions of the form $A = B$, where both A and B are terms denoting either a relative or a truth value. As a special case, $A = 0$ and $A = 1$ are equations.[47] The logic of relatives only deals with terms that have a propositional interpretation, that is, with terms denoting a truth value. A first-order term is a term of this kind whose quantifiers (if any) range over elements (not over relatives). In his presentation of the logic of relatives (1915), Löwenheim uses the word *Zählausdruck* (first-order expression) to refer to these terms and the word *Zählgleichung* (first-order equation) to refer to the equations whose terms are first-order expressions.[48] To move closer to the current terminology, in what follows we use the word "formula" for what Löwenheim calls *Zählausdruck.*

The set over which the individual variables range is the first-order domain (*Denkbereich der ersten Ordnung*) and is denoted by 1^1. The only condition that this domain must fulfill is to be nonempty. Schröder insists that it must have more than one element, but Löwenheim ignores this restriction. Relative variables range over the set of relations on 1^1. The second-order domain

(*Denkbereich der zweiten Ordnung*), 1^2, is the set of all ordered pairs whose coordinates belong to 1^1. In this exposition we are using the word *domain* as shorthand for "first-order domain."

The current distinction between the individual variables of the object language and the metalinguistic variables ranging over the elements of the domain does not exist in the logic of relatives. From the moment it is assumed that an equation is interpreted in a domain, the indices play simultaneously the role of variables of the formal language and that of variables of the metalanguage. The canonical names of the elements of the domain are then used as individual constants having a fixed interpretation. Thus, the semantic arguments that we find in the logic of relatives are better reproduced when we think of them as arguments carried out in the expanded language that results from adding the canonical names of the elements to the basic language.

Interpreting an equation means fixing a domain and assigning a relation on the domain to each relative occurring in it. We can say that an *interpretation in a domain* D of an equation (without free variables) is a function that assigns a relation on D to each relative occurring in the equation. The interpretation of a relative z can also be fixed by assigning a truth value to each coefficient of z in D, because, in the theory of relatives, for every $a, b \in D$, $\langle a, b\rangle \in z$ if and only if $z_{ab} = 1$. Thus, an interpretation of an equation in a domain D can also be defined as an assignment of truth values to the coefficients in D of the relatives (other than $1'$ and $0'$) occurring in the equation.

The most immediate response to an equation is to inquire about the systems of values that satisfy it. This inquiry has a clear meaning in the context of the logic of relatives and it does not require any particular clarification to understand it. The equations of the logic of relatives are composed of terms which in a domain D take a unique value (1 or 0) for each assignment of values to the coefficients in D of the relatives occurring in them. An equation is satisfied by an interpretation $\mathcal{I}$ in a domain if both members of the equation take the same value under $\mathcal{I}$. There is no essential difference between asking if there is a solution (an interpretation) that satisfies the equation $A = 1$ and asking if the formula A is satisfiable in the modern sense.[49] In this way, in the logic of relatives semantic questions arise naturally, propitiated by the algebraic context. There is no precise definition of any semantic concept, but the meaning of these concepts is clear enough for the proof of theorems such as Löwenheim's.

4.3. Löwenheim's Theorem

The simplest versions of the Löwenheim–Skolem theorem can be stated as follows: For every first-order sentence A,

a. if A is satisfiable, then it is satisfiable in some countable domain;

b. if an interpretation $\mathfrak{I}$ in D satisfies A, there exists a countable subdomain of D such that the restriction of $\mathfrak{I}$ to the subdomain satisfies A.

Version b (the subdomain version) is stronger than version a (the weak version) and has important applications in model theory. Some form of the axiom of choice is necessary to prove the subdomain version, but not to prove the weak one.

All modern commentators of Löwenheim's proof agree that he proved the weak version, and that it was Skolem (1920) who first proved the subdomain version and further generalized it to infinite sets of formulas. By contrast, Skolem (1938, 455), a logician trained in the algebraic tradition, attributed to Löwenheim the proof of the subdomain version, and in our opinion, this attribution must be taken seriously. The fact that Löwenheim's proof allows two readings so at variance with each other shows patently his argument is far from clear.

As far as the correctness of the proof is concerned, no logician of Löwenheim's time asserts that the proof is incorrect or that it has major gaps. The only inconvenience mentioned by Skolem is that the use of fleeing indices complicates the proof unnecessarily.[50] Herbrand thought that Löwenheim's argument lacks the rigor required by metamathematics but considered it "sufficient in mathematics" (Herbrand 1930, 176). The most widely held position today is that the proof has some important gaps, although commentators differ as to precisely how important they are. Without actually stating that the proof is incorrect, van Heijenoort maintains that Löwenheim does not account for one of the most important steps. Dreben and van Heijenoort (1986, 51) accept that Löwenheim proved the weak version, but state that their reading of the proof is a charitable one. For Vaught (1974, 156), the proof has major gaps, but he does not specify what they are. Wang (1970, 27 and 29) considers that Löwenheim's argument is "less sophisticated" than Skolem's in 1922, but does not say that it has any important gaps. Moore's point of view is idiosyncratic (see Moore 1980, 101; 1988, 121–122). In his opinion, the reason Löwenheim's argument appears "odd and unnatural" to the scholars just mentioned is that they consider it inside standard first-order logic instead of considering it in the frame of infinitary logic.

This diversity of points of view makes manifest the difficulty of understanding Löwenheim's argument and at the same time the necessity to provide a new reading of it.

Theorem 2 of Löwenheim's paper is: "If the domain is at least denumerably infinite, it is no longer the case that a first-order fleeing equation is satisfied for arbitrary values of the relative coefficients" (Löwenheim 1915, 235).

A *fleeing equation* is an equation that is not logically valid but is valid in every finite domain. Löwenheim's example of a fleeing equation is:

$$\sum_{l} \prod_{i,j,h} (\overline{z}_{hi} + \overline{z}_{hj} + 1'_{ij})\overline{z}_{li} \sum_{k} z_{ki} = 0.$$

For the proof of the theorem, he assumes without any loss of generality that every equation is in the form $A = 0$. This allows him to go from equations to formulas, bearing in mind that "$A = 0$ is valid" is equivalent to "A is not satisfiable." Thus, Löwenheim's argument can also be read as a proof of the following.

Theorem If a first-order sentence (a *Zählausdruck*) is satisfiable but not satisfiable in any finite domain, then it is satisfiable in a denumerable domain.

Löwenheim's proof can be split into two lemmas. We state them for formulas (not for equations) and comment on their proof separately.

Lemma 1 Every sentence of a first-order language is logically equivalent to a sentence of the form $\Sigma\Pi F$, where Σ stands for a possibly empty string of existential quantifiers, Π stands for a possibly empty string of universal quantifiers, and F is a quantifier-free formula.

The central step in the proof of this lemma involves moving the existential quantifiers in front of the universal quantifiers, preserving logical equivalence. Löwenheim takes this step by applying the equality

$$\prod_i \sum_k A_{ik} = \underline{\sum_{k_i}} \prod_i A_{ik_i}, \tag{1}$$

which is a notational variant of a transformation introduced by Schröder (1895, 513–516). According to Löwenheim, $\underline{\Sigma}_{k_i}$ is an n-fold quantifier, where n is the cardinality of the domain (n may be transfinite).[51] For example, if the domain is the set of natural numbers, then

$$\underline{\sum_{k_i}} \prod_i A_{ik_i} \tag{2}$$

can be developed in this way:

$$\sum_{k_1,k_2,k_3,\ldots} A_{1k_1} A_{2k_2} A_{3k_3} \ldots$$

Löwenheim warns, however, that this development of (2) contravenes the stipulations on language, even if the domain is finite.

Löwenheim calls terms of the form k_i *fleeing indices* (*Fluchtindizes*) and says that these indices are characterized by the fact that their subindices are universally quantified variables, but in fact, he also gives that name to the indices generated by a fleeing index when its universally quantified variables take values on a domain (k_1, k_2, $k_3, \ldots$ in the example).

Schröder's procedure for changing the order of quantifiers is generally considered to be the origin of the concept of the Skolem function, and

$$\forall x \exists y A(x, y) \leftrightarrow \exists f \forall x A(x, fx)$$

as the current way of writing (1).[52] Even if we subscribed to this assertion, we should notice that neither Schröder nor Löwenheim associated the procedure for changing the order of quantifiers with the quantification over functions (as Goldfarb notes in 1979). Skolem did not make this association either. In addition, the interpretation of (2) in terms of *Skolem functions* does not clarify why Schröder and Löwenheim reasoned as they did, nor does it explain some of Skolem's assertions as this one: "But his [Löwenheim's] reasonings can be simplified by using the 'Belegungsfunktionen' (i.e., functions of individuals whose values are individuals)" (Skolem 1938, 455–456). Finally, it is debatable whether fleeing indices are functional terms or not.

The usual way of interpreting Löwenheim's explanation of the meaning of (2) can be summarized as follows: (2) is a schema of formulas that produces different formulas depending on the cardinality of the domain under consideration; when the domain is infinite the result of the development is a formula of infinite length; in each case, (2) should be replaced by its development in the corresponding domain.[53] Against this interpretation the above-mentioned warning could be cited and also the fact that, strictly speaking, no step in Löwenheim's proof consists of the replacement of a formula by its development.

The main characteristic of fleeing indices is their ability to generate a different term for each element of the domain. If a is an element of the domain and k_i is a fleeing index, then k_a is an index. The terms generated by a fleeing index behave like any "normal" index (i.e., like any individual variable). Thus, Löwenheim can assert that k_a, unlike k_i, stands for an element of the domain.

In our view, Löwenheim's recourse to the development of quantifiers in a domain is a rather rough and ready way of expressing the semantics of formulas with fleeing indices. The purpose of the development of (2) is to facilitate the understanding of this kind of formulas. Today's technical and expressive devices allow us to express the meaning of (2) without recourse to developments. If for the sake of simplicity let us suppose that (2) has no free variables, then

(3) $\underline{\Sigma}_{k_i} \Pi_i A_{ik_i}$ is satisfied by an interpretation $\mathfrak{I}$ in a domain D if and only if there is an indexed family $\langle k_a \mid a \in D \rangle$ of elements of D such that for all $a \in D$: A_{ik_i} is satisfied by $\mathfrak{I}$ in D when i takes the value a and k_i the value k_a.

This interpretation of (2) is what Löwenheim attempts to express and is all we need to account for the arguments in which (2) intervenes. Löwenheim (unlike Schröder) does not see (2) as a schema of formulas. The developments are informal explanations (informal, because they contravene the stipulations on language) whose purpose is to facilitate the understanding of quantification over fleeing indices. Löwenheim has no choice but to give examples, because the limitations of his conceptual apparatus (specifically, the lack of a clear distinction between syntax and semantics) prevents him from giving the meaning of (2) in a way analogous to (3). Many of Schröder's and Löwenheim's arguments and remarks are better understood when they are read in the light of (3). In particular, some of these remarks show that they did not relate quantification over fleeing indices with quantification over functions, because they did not relate the notion of indexed family with that of function.

In the proof of Lemma 1, Löwenheim aims to present a procedure for obtaining a formula of the form $\Sigma\,\Pi\,F$ logically equivalent to a given formula A. One of the most striking features of Löwenheim's procedure is that the order in which he proceeds is the opposite of the one we would follow today. First he moves the existential quantifiers of A in front of the universal ones, and then obtains the prenex form. This way of arriving at a formula of the form $\Sigma\,\Pi\,F$ introduces numerous, totally unnecessary complications. One of the most unfortunate consequences of the order that Löwenheim follows is that the prenex form cannot be obtained in a standard first-order language, because the formula that results from changing the order of the quantifiers will contain quantified fleeing indices. Thus, to obtain the prenex form we need equivalences that tell us how to deal with these expressions and how to resolve the syntactic difficulties that they present. Löwenheim ignores these problems.

The proof of the lemma presents some problems, but its first part, the one in which existential quantifiers are moved in front of universal ones, is an essentially correct proof by recursion. Löwenheim is not aware of the recursion involved, but his proof shows that he intuits the recursive structure of a formal language.

Lemma 2 If $\Sigma\,\Pi\,F$ is satisfiable but not satisfiable in any finite domain, then it is satisfiable in a denumerable domain.

First of all, Löwenheim shows with the aid of examples that for this proof we can ignore the existential quantifiers of $\Sigma\,\Pi\,F$. He notes that a formula of the form $\Pi\,F$ is satisfiable in a domain D if there exists an interpretation of the relatives occurring in F and an assignment of values (elements in D) to the free variables of F and to the indices generated by the fleeing indices when their subindices range over the domain. But this is precisely what it means to assert that $\Sigma\,\Pi\,F$ is satisfiable in D.

The proof proper begins with the recursive definitions of a sequence (C_n, $n \geq 1$) of subsets of $C = \{1, 2, 3, \dots\}$ and of some sequences of formulas as follows.

1. If $\Pi\,F$ is a sentence, $C_0 = \{1\}$. If $\{j_1, \dots, j_m\}$ are the free variables of $\Pi\,F$, then $C_0 = \{1, \dots, m\}$. Let $\Pi\,F'$ be the result of replacing in $\Pi\,F$ the constant n $(1 \leq n \leq m)$ for the variable j_n. Let F_1 be the product of all the formulas that are obtained by dropping the quantifiers of $\Pi\,F'$ and replacing the variables that were quantified by elements of C_0. For example, if $\Pi\,F = \Pi_i\,F(i, j_1, j_2, k_i)$ then, $C_0 = \{1, 2\}$ and

$$F_1 = F(1, 1, 2, k_1) \cdot F(2, 1, 2, k_2).$$

If F_1 has p fleeing indices, we enumerate them in some order from $m+1$ to $m+p$. P_1 is the result of replacing in F_1 the individual constant n for the fleeing index t_n $(m+1 \leq n \leq m+p)$ and C_1 is the set of individual constants of P_1, that is, $C_1 = \{1, 2, \dots, m, \dots, m+p\}$. If $\Pi\,F$ and, therefore, F_1 has no fleeing indices, then $P_1 = F_1$ and $C_1 = C_0$. If in our example, the fleeing indices are enumerated from 2 onward in the order in which they occur in F_1, then

$$P_1 = F(1,1,2,3) \cdot F(2,1,2,4).$$

At this point Löwenheim makes the following claim.

Claim 2.1 If P_1 is not satisfiable, then ΠF is not satisfiable.

To determine whether P_1 is satisfiable or not, Löwenheim takes identity into account and considers all possible systems of equalities and inequalities between the constants that occur in P_1.[54] He implicitly assumes that we choose a representative of each equivalence class of each equivalence relation. Then, for each system of equalities between the constants of P_1, we obtain the formula resulting from

i. replacing each constant of P_1 by the representative of its class; and

ii. evaluating the coefficients of $1'$ and $0'$. This means that in place of $1'_{ab}$, we will write 1 or 0, depending on whether $a = b$ or $a \neq b$, and analogously for the case of $0'_{ab}$. Thus, each system of equalities determines the values of the relative coefficients of $1'$ and $0'$ and this allows us to eliminate these coefficients.

Because C_1 is finite, we obtain by this method a finite number of formulas:

$$P_1^1, P_1^2, \ldots, P_1^q.$$

Following Skolem's terminology (1922, 296), we use the expression *formulas of level* 1 to refer to these formulas.

Löwenheim goes on by stating the following.

Claim 2.2 If no formula of level 1 is satisfiable, then ΠF is not satisfiable.

He could now have applied the hypothesis of the theorem to conclude that there are satisfiable formulas at level 1, but instead of doing so, he argues as follows: if no formula of level 1 is satisfiable, we are done; if some formula is satisfiable, we proceed to the next step of the construction.

2. Let F_2 be the product of all the formulas that are obtained by dropping the quantifiers of $\Pi F'$ and replacing the variables that were quantified by elements of C_1. Evidently, the fleeing indices of F_1 are also fleeing indices of F_2. Suppose that F_2 has q fleeing indices that do not occur in F_1. Enumerate these new fleeing indices in some order starting at $m+p+1$. Now, P_2 is the result of replacing in F_2 each individual constant n for the corresponding fleeing index t_n $(m+1 \leq n \leq m+p+q)$, and C_2 is the set of individual constants of P_2, that is, $C_2 = \{1, 2, \ldots, m+p+q\}$. If ΠF and, therefore, F_1 has no fleeing indices, then $P_2 = P_1$ and $C_2 = C_1$. If in our example, the fleeing indices are enumerated from 4 onward in the order in which they occur in F_1, then

$$P_2 = F(1,1,2,3) \cdot F(2,1,2,4) \cdot F(3,1,2,5) \cdot F(4,1,2,6).$$

As before, we take into account all possible systems of equalities between the elements of C_2, and for each of these systems, we obtain the formula resulting from replacing each constant by the representative of its class and from evaluating the coefficients of $1'$ and $0'$. Let the formulas obtained by this method (the formulas of level 2) be:

$$P_2^1, P_2^2, \ldots, P_2^r.$$

If no formula of level 2 is satisfiable, we are done; if any of them is satisfiable, we repeat the process to construct P_3, C_3, and the formulas of level 3. By repeatedly applying this method, we can construct for each $n \geq 1$, the formula P_n, the subset C_n and the associated formulas of level n.

We emphasize a number of points that will be important in the final part of the proof.

a. The number of formulas at each level is finite, since for each n, C_n is finite.

b. Let us say that a formula A is an *extension* of a formula B, if A is of the form $B \cdot B'$. Löwenheim assumes that for every n, F_{n+1} is an extension of F_n. Thus, if $n < m$, P_m is an extension of P_n, and each formula Q of level m is an extension of one and only one formula of level n. The relation of extension on the set of all formulas occurring at some level (the formulas P_n^r obtained from P_1, $P_2, \ldots$) is a partial order on the set of all formulas. This kind of partial order is what we today call a *tree*.

c. Because what we said about the formulas of level 1 goes for any $n > 1$ as well, the following generalization of Claim 2.2 can be considered as proven.

Claim 2.3 If there exists n such that no formula of level n is satisfiable, then ΠF is not satisfiable.

We now present the last part of Löwenheim's argument. We deliberately leave a number of points unexplained—points which, in our opinion, Löwenheim does not clarify. In the subsequent discussion we argue for our interpretation and explain all the details.

By the hypothesis of the theorem, there is an interpretation in an infinite domain D that satisfies $\Sigma\Pi F$ and, therefore, ΠF. As a consequence, at each level there must be at least one true formula under this interpretation and, therefore, the tree of formulas constructed by following Löwenheim's procedure is infinite. Among the true formulas of the first level which, we recall, is finite, there must be at least one which has infinitely many true extensions (i.e., one that has true extensions at each of the following levels). Let Q_1 be one of these formulas. At the second level, which is also finite, there are true formulas which are extensions of Q_1 and also have infinitely many true extensions. Let us suppose that Q_2 is one of these formulas. In the same way, at the third level there must be true formulas which are extensions of Q_2 (and, therefore, of Q_1) and have infinitely many true extensions. Let Q_3 be one of these formulas. In this way, there is a sequence of formulas

Q_1, Q_2, $Q_3, \ldots$ such that for each $n > 0$: Q_{n+1} is a true extension of Q_n. Consequently,

$$Q_1 \cdot Q_2 \cdot Q_3 \cdots = 1. \tag{4}$$

The values taken by the various kind of indices whose substitution gives rise to the sequence Q_1, Q_2, $Q_3, \ldots$ determine a subdomain of D on which ΠF has the same truth value as $Q_1 \cdot Q_2 \cdot Q_3 \cdots$. Because this subdomain cannot be finite, because ΠF is not satisfiable in any finite domain, we conclude that $\Pi F = 1$ in a denumerable domain. This ends the proof of the theorem.

Basically, this part of Löwenheim's argument is the proof of a specific case of what we know today as the infinity lemma proved later with all generality by Denes König (1926, 1927). The proof of this lemma requires the use of some form of the axiom of choice, but when the tree is countable (as in this case) any enumeration of its nodes allows us to choose one from each level without appealing to the axiom of choice. Since Löwenheim does not choose the formulas on the basis of any ordering, we can assume that he is implicitly using some form of the axiom of choice.

Modern commentators have seen in the construction of the tree an attempt to construct an interpretation of ΠF in a denumerable domain. Van Heijenoort (1967a, 231) reads the final step in this way: "for every n, Q_n is satisfiable; therefore, $Q_1 \cdot Q_2 \cdot Q_3 \cdots$ is satisfiable." This step is correct but, as the compactness theorem had not been proven in 1915 and Löwenheim does not account for it, van Heijenoort concludes that the proof is incomplete. Wang considers that Löwenheim is not thinking of formulas but of interpretations. According to his reading, the tree that Löwenheim constructs should be seen as if any level n were formed by all the interpretations in D (restricted to the language of P_n) that satisfy P_n. The number of interpretations at each level is also finite, although it is not the same as the number of formulas that Löwenheim considers. Thus, when Löwenheim fixes an infinite branch of the tree, it should be understood that he is fixing a sequence of partial interpretations such that each one is an extension of the one at the previous level. The union of all these partial interpretations is an interpretation in a denumerable domain that satisfies P_n for every $n \in N$, and therefore ΠF.

The main difference between these readings of Löwenheim's argument and the foregoing version is that instead of constructing the sequence Q_1, Q_2, $Q_3, \ldots$ with satisfiable formulas or interpretations, we do so with formulas that are true under the interpretation that, by hypothesis, satisfies ΠF in D. Obviously, this means that we subscribe to the view that Löwenheim meant to prove the subdomain version of the theorem.

The aim of Löwenheim's proof is to present a method for determining a domain. The determination is made when all the possible systems of equalities are introduced. In a way, it is as if the satisfiable formulas of a level n represented all the possible ways of determining the values of the constants occurring in P_n. Thus, when Löwenheim explains how to construct the different levels of the tree, what he means to be explaining is how to determine a domain on the basis of an interpreted formula; consequently, when the construction is completed, he states that he has constructed it.

In Löwenheim's view the problem of determining the system of equalities between numerals is the same (or essentially the same) as that of fixing the values taken by the summation indices of ΠF (the free variables, and the indices generated by the fleeing indices). Each system of equalities between the numerals of P_n is biunivocally associated to a formula of level n. The formulas of any level n *represent*, from Löwenheim's perspective, all the possible ways of determining the values taken by the numerals that occur in P_n and, in the last resort, the values taken by the indices replaced by the numerals (i.e., the free variables in ΠF and the indices generated when their fleeing indices range over the set of numerals occurring in P_{n-1}). Thus, any assignment of values to these indices is represented by a formula of level n. Now, if ΠF is satisfiable, at each level there must be at least one satisfiable formula. In the same way, if ΠF is true in a domain D, at each level there must be at least one true formula (in other words, for each n there exists an assignment of elements of D to the numerals of P_n that satisfies P_n, assuming that the relative coefficients are interpreted according to the interpretation that, by hypothesis, satisfies $\Sigma\Pi F$ in D). The infinite branches of the tree represent the various ways of assigning values to the summation indices of ΠF in a denumerable domain. The product of all the formulas of any infinite branch can be seen as a possible development of ΠF in a denumerable domain. This assertion is slightly inexact, but we think this is how Löwenheim sees it, and for this reason he claims without any additional clarification that for the values of the summation indices that give rise to the sequence $Q_1, Q_2, Q_3, \ldots$, the formula ΠF takes the same truth value as the product $Q_1 \cdot Q_2 \cdot Q_3 \cdots$. Thus, showing that the tree has an infinite branch of true formulas (in the sense just described) amounts, from this perspective, to constructing a subdomain of D in which ΠF is true, and this is what Löwenheim set out to do.

One of the reasons for seeing in Löwenheim's argument an attempt to construct an interpretation in a denumerable domain is probably that when it is seen as a proof of the subdomain version of the theorem, the construction of the tree appears to be an unnecessary complication. He could, it seems, have offered a simpler proof which would not have required that construction and which would have allowed him to reach essentially the same conclusion. Löwenheim reasons in the way he does because he lacks the conceptual distinctions required to pose the problem accurately. The meaning of ΠF and the relation between this formula and $\Sigma\Pi F$ cannot be fully grasped without the concept of assignment or at least without sharply distinguishing between the terms of the language and the elements they denote. From Löwenheim's point of view, the assumption that ΠF is satisfied by an interpretation in D does not imply that the values taken by the summation indices are fixed. All he manages to intuit is that the problem of showing that $\Sigma\Pi F$ is satisfiable is equivalent to the problem of showing that ΠF is satisfiable. He then proceeds essentially as he would with $\Sigma\Pi F$, but without the inconvenience of having to eliminate the existential quantifiers each time that a formula of the sequence P_1, $P_2, \ldots$ is constructed: he assumes that the nonlogical relatives (i.e., relatives other than $1'$ and $0'$) of ΠF have a fixed meaning in a domain D and proposes fixing the values of summation indices in a denumerable subdomain of D. This means that in practice Löwenheim is arguing as he would do if the prefix had the form $\Pi\,\Sigma$.

Löwenheim's strategy is then as follows: First he presents a procedure of a general nature to construct a tree of a certain type, and then (without any warning, and without differentiating between the two ideas) he applies the hypothesis of the theorem to the construction. The reason for the style that he adopts in the construction of the tree probably lies in his desire to make it clear that the technique he is presenting is applicable to any formula in normal form and not only to one that meets the conditions of the hypothesis. If the starting formula is not satisfiable, we will conclude the construction in a finite number of steps because we will reach a level at which none of the formulas is satisfiable; if the starting formula is satisfiable in a domain D, then, according to Löwenheim, this construction will allow us to determine a finite or denumerable subdomain of D in which it is satisfiable.

We must distinguish between what Löwenheim actually constructs and what he thinks is constructing. On the one hand, the tree (which he constructs) naturally admits a syntactic reading and can be viewed as a method of analyzing quantified formulas. This proof method was later used by Skolem, Herbrand, Gödel, and more recently by Quine (though he related it with Skolem and not with Löwenheim) (Quine 1955b; 1972, 185ff.). On the other hand, it is obvious that, contrary to Löwenheim's belief, the process of constructing the sequence Q_1, Q_2, $Q_3, \ldots$ does not represent the process of constructing a subdomain, because neither these formulas nor their associated systems of equalities can play the role of partial assignments of values to the summation indices. If we wanted to reflect what Löwenheim is trying to express, we should construct a tree with partial assignments rather than with formulas and modify his argument accordingly. Thus, Löwenheim's proof is not completely correct, but any assessment of it must take into account that he lacked the resources that would allow him to express his ideas better.

4.4. Skolem's First Versions of Löwenheim's Theorem

Although Skolem did not explicitly state the subdomain version until 1929, this was the version that he proved in 1920. At the beginning of this paper (1920, 254), Skolem asserts explicitly that his aim is to present a simpler proof of Löwenheim's theorem that avoids the use of fleeing indices. He then introduces what today we know as *Skolem normal form for satisfiability* (a prenex formula with the universal quantifiers preceding the existential ones), and then shows the subdomain version of the theorem for formulas in that form. This change of normal form is significant, because Löwenheim reasons as if the starting formula were in the form $\Pi\,\Sigma\,F$ (as remarked) and, therefore, the recourse to

$\Pi\,\Sigma$ formulas seems to be the natural way of dispensing with fleeing indices. Skolem's construction of a countable subdomain is, in essence, the usual one. Let us suppose that $\Pi_{x_1} \ldots \Pi_{x_n} \Sigma_{y_1} \ldots \Sigma_{y_m} U_{x_1 \ldots x_n y_1 \ldots y_m}$ (his notation) is the $\Pi\,\Sigma$ formula that is satisfied by an interpretation $\mathcal{I}$ in a domain D. By virtue of the axiom of choice, there is a function h that assigns to each n-tuple $(a_1, \ldots, a_n)$ of elements in D the m-tuple $(b_1, \ldots, b_m)$ of elements in D such that $U_{a_1 \ldots a_n b_1 \ldots b_m}$ is satisfied by $\mathcal{I}$ in D. Let a be any element in D. The countable subdomain D' is the union $\bigcup_n D_n$, where $D_0 = \{a\}$ and for each n, D_{n+1} is the union of D_n and the set of elements in the m-tuples $h(a_1, \ldots, a_n)$ for $a_1, \ldots, a_n \in D_n$.

In 1922, Skolem proved the weak version of the theorem, which allowed him to avoid the use of the axiom of choice. The schema of Skolem's argument is as follows: (1) he begins by transforming the starting formula A into one in normal form for satisfiability which is satisfiable if and only if A is; (2) he then constructs a sequence of formulas which, in essence, is Löwenheim's P_1, $P_2, \ldots$, and, for each n, he defines a linear ordering on the finite set of (partial) interpretations that satisfy P_n in the set of numerals of P_n; and (3) after observing that the extension relation defined in the set of all partial interpretations is an infinite tree whose levels are finite, Skolem fixes an infinite branch of this tree; this branch determines an interpretation that satisfies A in set of natural numbers (assuming that A is formula without identity).

Skolem's (1922) proof seems similar to Löwenheim's in certain aspects, but the degree of similarity depends on our reading of the latter. If Löwenheim was attempting to construct a subdomain, the two proofs are very different: Each one uses a distinct notion of normal form, fleeing indices do not intervene in Skolem's proof, and, more important, the trees constructed in each case involve different objects (in Löwenheim's proof the nodes represent partial assignment of values to the summation indices, while in Skolem's the nodes are partial interpretations). These are probably the differences that Skolem saw between his proof and Löwenheim's. The fact is that in 1922 he did not relate one proof to the other. This detail corroborates the assumption that Skolem did not see in Löwenheim's argument a proof of the weak version of the theorem.

In 1964 Gödel wrote to van Heijenoort:

> As for Skolem, what he could justly claim, but apparently does not claim, is that, in his 1922 paper, he implicitly proved: "Either A is provable or $\neg A$ is satisfiable" ("provable" taken in an informal sense). However, since he did not clearly formulate this result (nor, apparently, had he made it clear to himself), it seems to have remained completely unknown, as follows from the fact that Hilbert and Ackermann (1928) do not mention it in connection with their completeness problem. (Dreben and van Heijenoort 1986, 52)

Gödel made a similar assertion in a letter to Wang in 1967 (Wang 1974, 8). Gödel means that Skolem's (1922) argument can be viewed as (or can easily be transformed into) a proof of a version of the completeness theorem (see itinerary VIII). This is so because the laws and transformations used to obtain the normal form of a formula A, together with the rules employed in the construction of the sequence $P_1, P_2, \ldots$ associated with A and the rules used to decide whether a formula without quantifiers is satisfiable can be viewed as an informal refutation procedure. From this point of view, to say that P_n $(n \geq 1)$ is not satisfiable is equivalent to saying that the informal procedure refutes it. Now, we can define what it means to be provable as follows:

1. A formula A is refutable if and only if there exists n such that the informal procedure refutes P_n;
2. A formula A is provable if and only if $\neg A$ is refutable.

An essential part of Skolem's argument is the proof of the following result.

Lemma 3 If for every n, P_n is satisfiable, then A is satisfiable.

With the aid of the foregoing definitions, this lemma can be restated as follows.

Lemma 4 If A is not satisfiable, then A is refutable.

This lemma (which is equivalent to Gödel's formulation: Either A is provable or $\neg A$ is satisfiable) asserts the completeness of the informal refutation procedure.[55]

Since the laws and rules used by Löwenheim in his proof can also be transformed into an informal refutation procedure (applicable even to formulas with equality), it is interesting to ask whether he proves Lemma 3 (for ΠF formulas). The answer to this question depends on our reading of his proof. If we think, as van Heijenoort and Wang do, that Löwenheim proved the weak version, then we are interpreting the last part of his argument as an (incomplete or unsatisfactory) proof of Lemma 3. Thus, if we maintain that Löwenheim proved the weak version, we have to accept that what Gödel asserts in the quotation applies also to Löwenheim as well. In our view, Löwenheim did not try to construct an interpretation, but a subdomain. He did not set out to prove Lemma 3, and as a consequence, Gödel's assertion is not applicable to him.

5. Itinerary V. Logic in the Hilbert School

5.1. Early Lectures on Logic

David Hilbert's interests in the foundations of mathematics began with his work on the foundations of geometry in the 1880s and 1890s (Hilbert 1899, 2004). Although he was then primarily concerned with geometry, he was interested more broadly in the principles underlying the axiomatic method, and in Dedekind's work (1888). A number of factors worked together to persuade Hilbert around 1900 that a fundamental investigation of logic and its relationship to the foundation of mathematics was needed. These were his correspondence with Frege (1899–1900) on the nature of axioms and the realization that his formulation of geometry was incomplete without an axiom of completeness. They were manifest in his call for an independent consistency proof of arithmetic in his 1900 address, and in his belief that every meaningful mathematical problem had a solution ("no ignorabimus").

Although the importance of logic was clear to Hilbert in the early years of the 1900s, he himself did not publish on logic. His work and influence then consisted mainly in a lecture course he taught in 1905 and a number of administrative decisions he made at Göttingen. The latter are described in detail in Peckhaus (1990, 1994, 1995), and include his involvement with the appointment of Edmund Husserl and Ernst Zermelo at Göttingen.

Hilbert's first in-depth discussion of logic occurred in his course "Logical Principles of Mathematical Thought" in the summer term of 1905. The lectures centered on set theory (axiomatized in natural language, just like his axiomatic treatment of geometry), but in chapter V, Hilbert also discussed a basic calculus of propositional logic. The presentation is influenced mainly by Schröder's algebraic approach.

> Axiom I If $X \equiv Y$ then one can always replace X by Y and Y by X.
>
> Axiom II From 2 propositions X, Y a new one results ("additively")
>
> $$Z \equiv X + Y$$
>
> Axiom III From 2 propositions X, Y a new one results in a different way ("multiplicatively")
>
> $$Z \equiv X \cdot Y$$
>
> The following identities hold for these "operations":
>
> IV. $X + Y \equiv Y + X$
>
> V. $X + (Y + Z) \equiv (X + Y) + Z$
>
> VI. $X \cdot Y \equiv Y \cdot X$
>
> VII. $X \cdot (Y \cdot Z) \equiv (X \cdot Y) \cdot Z$
>
> VIII. $X \cdot (Y + Z) \equiv X \cdot Y + X \cdot Z$
>
> There are 2 definite propositions 0, 1, and for each proposition X a different proposition $\overline{X}$ is defined, so that the following identities hold:
>
> IX. $X + \overline{X} \equiv 1$
>
> X. $X \cdot \overline{X} \equiv 0$
>
> XI. $1 + 1 \equiv 1$
>
> XII. $1 \cdot X \equiv X$
>
> (Hilbert 1905a, 225–228)

Hilbert's intuitive explanations make clear that X, Y, and Z stand for propositions, $+$ for conjunction, $\cdot$ for disjunction, $\overline{\cdot}$ for negation, 1 for falsity, and 0 for truth. In the absence of a first-order semantics, neither statement nor proof of a semantic completeness claim could be given. Hilbert does, however, point out that not every unprovable formula renders the system inconsistent when added as an axiom, that is, the full function calculus is not (what we now call) Post-complete.

5.2. The Completeness of Propositional Logic

Hilbert's work on the foundations of logic begins in earnest with a lecture course on the principles of mathematics he taught in the winter semester 1917/18 (1918b). These form the basis of Hilbert and Ackermann (1928) (see Section 5.5 and Sieg 1999), and contain a wealth of material on propositional and first-order logic, as well as Russell's type theory. We focus here on the development of the propositional calculus in these lectures. Syntax and axioms are modeled after the propositional fragment of *Principia Mathematica* (Whitehead and Russell 1910). The language consists of propositional variables (*Aussage-Zeichen*) X, Y, $Z, \ldots$, as well as signs for particular propositions, and the connectives $\overline{\cdot}$ (negation) and $\times$ (disjunction). The conditional, conjunction, and equivalence are introduced as abbreviations. Expressions are defined by recursion:

1. Every propositional variable is an expression.
2. If α is an expression, so is $\overline{\alpha}$.
3. If α and β are expressions, so are $\alpha \times \beta$, $\alpha \to \beta$, $\alpha + \beta$ and $\alpha = \beta$.

Hilbert introduces a number of conventions, for example, that $X \times Y$ may be abbreviated to XY, and the usual conventions for precedence of the connectives. Finally, the logical axioms are introduced. Group I of the axioms of the function calculus gives the formal axioms for the propositional fragment (unabbreviated forms are given on the right, recall that XY is "X or Y").

1. $XX \to X$ $\quad \overline{XX}X$
2. $X \to XY$ $\quad \overline{X}(XY)$
3. $XY \to YX$ $\quad \overline{XY}(YX)$
4. $X(YZ) \to (XY)Z$ $\quad \overline{X(YZ)}((XY)Z)$
5. $(X \to Y) \to (ZX \to ZY)$ $\quad \overline{\overline{X}Y}(\overline{\overline{Z}X}(ZY))$

The formal axioms are postulated as correct formulas (*richtige Formel*), and we have the following two rules of derivation ("contentual axioms"):

a. Substitution: From a correct formula another one is obtained by replacing all occurrences of a propositional variable with an expression.

b. If α and $\alpha \to \beta$ are correct formulas, then β is also correct.

Although the calculus is very close to the one given in *Principia Mathematica*, there are some important differences. Russell uses (2′) $X \to YX$ and (4′) $X(YZ) \to Y(XZ)$ instead of (2) and (4). *Principia* also does not have an explicit substitution rule.[56] The division between syntax and semantics, however, is not quite complete. The calculus is not regarded as concerned with uninterpreted formulas; it is not separated from its interpretation. (This is also true of the first-order part, see Sieg 1999, B3.) Also, the notion of a "correct formula" which occurs in the presentation of the calculus is intended not as a concept defined, as it were, by the calculus (as we would nowadays define the term "provable formula" for instance) but rather should be read as a semantic stipulation: The axioms are true, and from true formulas we arrive at more true formulas using the rules of inference.[57] Read this way, the statement of modus ponens is not that much clearer than the one given in *Principia:* "Everything implied by a true proposition is true" (*1.1).

Hilbert goes on to give a number of derivations and proves additional rules. These serve as stepping stones for more complicated derivations. He proves a normal form theorem to establish decidability and completeness. In the new propositional calculus, however, Hilbert has to establish that arbitrary subformulas can be replaced by equivalent formulas, that is, that the rule of replacement is a dependent rule. He does so by establishing the admissibility of rule (c): If $\varphi(\alpha)$, $\alpha \to \beta$, and $\beta \to \alpha$ are provable, then so is $\varphi(\beta)$. With that, the admissibility of using commutativity, associativity, distributivity, and duality inside formulas is quickly established, and Hilbert obtains the normal form theorem just as he did for the first propositional calculus in the 1905 lectures. Normal forms again play an important role in proofs of decidability and now also completeness.

5.3. Consistency and Completeness

"This system of axioms would have to be called inconsistent if it were to derive two formulas from it which stand in the relation of negation to one another" (Hilbert 1918b, 150). Hilbert proves that the system of axioms is not inconsistent in this sense using an arithmetical interpretation. The propositional variables are interpreted as ranging over the numbers 0 and 1, $\times$ is just multiplication, and $\overline{X}$ is $1 - X$. One sees that the five axioms represent functions which are constant equal to 0, and that the two rules preserve that property. Now if α is derivable, $\overline{\alpha}$ represents a function constant equal to 1, and thus is underivable.

Hilbert then poses the question of completeness in the syntactic sense for the propositional calculus in the following way:

> Let us now turn to the question of *completeness*. We want to call the system of axioms under consideration complete if we always obtain an inconsistent system of axioms by adding a formula which is so far not derivable to the system of basic formulas. (Hilbert 1918b, 152)

This is the first time that completeness is formulated as a precise mathematical question to be answered for a system of axioms. Before this, Hilbert (1905a, 13) had formulated completeness as the question of whether the axioms suffice to prove all "facts" of the theory in question. The notion of completeness is of course related to the *axiom of completeness*. This axiom was missing from the first edition of *Grundlagen der Geometrie*, but was added in subsequent editions. Hilbert also added such an axiom to his axiomatization of the reals in (1900b); it states that it is not possible to extend the system of real numbers by adding new entities so that the other axioms are still satisfied. Following the formulation of the completeness axiom in Hilbert (1905a), we read:

> This last axiom is of a general kind and has to be added to every axiom system whatsoever in some form. It is of special importance in this case, as we shall see. Following this axiom, the system of numbers has to be so that whenever new elements are added contradictions arise, regardless of the stipulations made about them. If there are things which can be adjoined to the system without contradiction, then in truth they already belong to the system. (Hilbert 1905a, 17)

The formulation of completeness can be seen to arise directly out of the completeness axioms of Hilbert's earlier axiomatic systems, only this time completeness is a theorem *about* the system instead of an axiom *in* the system. The completeness axiom stated that the domain cannot be extended without producing contradictions; the domain of objects is the system of real numbers in one case, the system of provable propositional formulas in the other.[58]

The completeness proof in the 1917/18 lectures itself is an ingenious application of the normal form theorem: Every formula is interderivable with a conjunctive normal form. As has been proven earlier in the lectures, a conjunction is provable if and only if each of its conjuncts is provable. A disjunction of propositional variables and negations of propositional variables is provable only if it represents a function which is constant equal to 0, as the consistency proof shows. A disjunction of this kind is equal to 0 if and only if it contains a variable and its negation, and conversely, every such disjunction is provable. So a formula is provable if and only if every conjunct in its normal form contains a variable and its negation. Now suppose that α is an underivable formula. Its conjunctive normal form β is also underivable, so it must contain a conjunct γ where every variable occurs only negated or unnegated but not both. If α were added as a new axiom, then β and γ would also be derivable. By substituting X for every unnegated variable and $\overline{X}$ for every negated variable in γ, we would obtain X as a derivable formula (after some simplification), and the system would be inconsistent.[59]

In a footnote, the result is used to establish the converse of the characterization of provable formulas used for the consistency proof: Every formula representing a function which is constant equal to 0 is provable. For, supposing there were such a formula that was not provable, then adding this formula to the axioms would not make the system inconsistent, by the same argument as in the consistency proof. This would contradict syntactic completeness (Hilbert 1918b, 153).

We have seen that the lecture notes to *Principles of Mathematics* 1917–18 contain consistency and completeness proofs (relative to a syntactic completeness concept) for the propositional calculus of *Principia Mathematica*. They also implicitly contain the familiar truth value semantics and a proof of semantic soundness and completeness. In his *Habilitationsschrift* (Bernays 1918), Bernays fills in the last gaps between these remarks and a completely modern presentation of propositional logic.

Bernays introduces the propositional calculus in a purely formal manner. The concept of a formula is defined and the axioms and rules of derivation are laid out almost exactly as done in the lecture notes. §2 of Bernays (1918) is titled "Logical interpretation of the calculus. Consistency and completeness." Here Bernays first gives the interpretation of the propositional calculus, which is the motivation for the calculi in Hilbert's earlier lectures (Hilbert 1905a, 1918b). The reversal of the presentation—first calculus, then its interpretation—makes it clear that Bernays is fully aware of a distinction between syntax and semantics, a distinction not made precise in Hilbert's earlier writings. There, the calculi were always introduced with the logical interpretation built in, as it were. Bernays writes:

> The axiom system we set up would not be of particular interest, were it not capable of an important contentual interpretation.
>
> Such an interpretation results in the following way:
>
> The variables are taken as symbols for *propositions* (sentences).
>
> That propositions are either true or false, and not both simultaneously, shall be viewed as their characteristic property.
>
> The symbolic product shall be interpreted as the connection of two propositions by "or," where this connection should not be understood in the sense of a proper disjunction, which excludes the case of both propositions holding jointly, but rather so that "X or Y" holds (i.e., is true) if and only if at least one of the two propositions X, Y holds. (Bernays 1918, 3–4)

Similar truth-functional interpretations of the other connectives are given as well. Bernays then defines what a provable and what a valid formula is, thus making the syntax-semantics distinction explicit:

> The importance of our axiom system for logic rests on the following fact: If by a "provable" formula we mean a formula which can be shown to be correct according to the axioms [footnote in text: It seems to me to be necessary to introduce the concept of a provable formula in addition to that of a correct formula (which

> is not completely delimited) in order to avoid a circle], and by a "valid" formula one that yields a true proposition according to the interpretation given for any arbitrary choice of propositions to substitute for the variables (for arbitrary "values" of the variables), then the following theorem holds:
>
> *Every provable formula is a valid formula and conversely.*
>
> The first half of this claim may be justified as follows: First one verifies that all basic formulas are valid. For this one only needs to consider finitely many cases, for the expressions of the calculus are all of such a kind that in their logical interpretation their truth or falsehood is determined uniquely when it is determined of each of the propositions to be substituted for the variables whether it is true or false. The content of these propositions is immaterial, so one only needs to consider truth and falsity as values of the variables. (Bernays 1918, 6)

We have here all the elements of a modern discussion of propositional logic: A formal system, a semantics in terms of truth values, soundness and completeness relative to that semantics. As Bernays points out, the consistency of the calculus follows from its soundness. The semantic completeness of the calculus is proved in §3, along the lines of the note in Hilbert (1918b) just mentioned. The formulation of syntactic completeness given by Bernays is slightly different from the lectures and independent of the presence of a negation sign: It is impossible to add an unprovable formula to the axioms without thus making all formulas provable.[60] Bernays sketches the proof of syntactic completeness along the lines of Hilbert's lectures, but leaves out the details of the derivations.

Bernays also addresses the question of decidability. In the lecture notes, decidability was not mentioned, even though Hilbert had posed it as one of the fundamental problems in the investigation of the calculus of logic. In his talk in Zürich in 1917, he said that an axiomatization of logic cannot be satisfactory until the question of decidability by a finite number of operations is understood and solved (Hilbert 1918a, 1143). Bernays gives this solution for the propositional calculus by observing that

> this consideration does not only contain the proof for the completeness of our axiom system, but also provides a uniform method by which one can decide after finitely many applications of the axioms whether an expression of the calculus is a provable formula or not. To decide this, one need only determine a normal form of the expression in question and see whether at least one variable occurs negated and unnegated as a factor in each simple product. If this is the case, then the expression considered is a provable formula, otherwise it is not. The calculus therefore can be completely trivialized. (Bernays 1918, 15–16)

Consistency and independence are the requirements that Hilbert laid down for axiom systems of mathematics time and again. Consistency was established—but the "contributions to the axiomatic treatment" of propositional logic could not be complete without a proof that the axioms investigated are independent. In fact, however, the axiom system for the propositional calculus, slightly modified from the postulates in (*1) of *Principia Mathematica*, is not independent. Axiom 4 is provable from the other axioms. Bernays devotes §4 of the *Habilitationsschrift* to give the derivation, and also the interderivability of the original axioms of *Principia* (2′) and (4′) with the modified versions (2) and (4) in presence of the other axioms.

Independence is, of course, more challenging. The method Bernays uses is not new, but it is applied masterfully. Hilbert had already used arithmetical interpretations (1905a) to show that some axioms are independent of the others. The idea was the same as that originally used to show the independence of the parallel postulate in Euclidean geometry: To show that an axiom α is independent, give a model in which all axioms but α are true, the inference rules are sound, but α is false. Schröder was the first to apply that method to logic. §12 of his *Algebra of Logic* (Schröder 1890) gives a proof that one direction of the distributive law is independent of the axioms of logic introduced up to that point (see Thiel 1994). The interpretation he gives is that of the "calculus of algorithms," developed in detail in appendix 4. Bernays combines Schröder's idea with Hilbert's arithmetical interpretation and the idea of the consistency proof for the first propositional calculus in Hilbert (1918b) (interpreting the variables as ranging over a certain finite number of propositions, and defining the connectives by tables). He gives six "systems" to show that each of the five axioms (and a number of other formulas) is independent of the others. The systems are, in effect, finite matrices. He introduces the method as follows:

> In each of the following independence proofs, the calculus will be reduced to a finite system (a finite group in the wider sense of the word [footnote: that is, without assuming the associative law or the unique invertability of composition]), where for each element a composition ("symbolic product") and a "negation" is defined. The reduction is given by letting the variables of the calculus refer to elements of the system as their values. The "correct formulas" are characterized in each case as those formulas which only assume values from a certain subsystem T for arbitrary values of the variables occurring in it. (Bernays 1926, 27–28)

We shall not go into the details of the derivations and independence proofs; see section 8.2.[61] Bernays's method was of some importance in the investigation of alternative logics. For instance, Heyting (1930a) used it to prove the independence of his axiom system for intuitionistic logic, and Gödel (1932b) was influenced by it when he defined a sequence of sentences F_n so that each F_n is independent of intuitionistic propositional calculus together with all F_i, $i > n$ (see section 7.1.7).[62]

5.4. Axioms and Inference Rules

In the final section of his *Habilitationsschrift*, Bernays considers the question of whether some of the axioms of the propositional calculus may be replaced by rules. This seems like a natural question, given the relationship between inference and implication: For instance, axiom 5 suggests the following rule of inference (recall that $\alpha\beta$ is Hilbert's notation for the disjunction of α and β):

$$\frac{\alpha \to \beta}{\gamma\alpha \to \gamma\beta}\,\mathrm{c},$$

which Bernays used earlier as a derived rule. Indeed, axiom 5 is in turn derivable using this rule and the other axioms and rules. Bernays considers a number of possible rules,

$$\frac{\begin{array}{l}\alpha \to \beta\\ \beta \to \gamma\end{array}}{\alpha \to \gamma}\,\mathrm{d} \qquad \frac{\alpha\alpha}{\alpha}\,\mathrm{r}_1 \qquad \frac{\alpha}{\alpha\beta}\,\mathrm{r}_2 \qquad \frac{\alpha\beta}{\beta\alpha}\,\mathrm{r}_3 \qquad \frac{\alpha(\beta\gamma)}{(\alpha\beta)\gamma}\,\mathrm{r}_4$$

$$\frac{\varphi(\alpha\alpha)}{\varphi(\alpha)}\,\mathrm{R}_1 \qquad \frac{\varphi(\alpha\beta)}{\varphi(\beta\alpha)}\,\mathrm{R}_3,$$

and shows that the following sets of axioms and rules are equivalent (and hence, complete for propositional logic):

1. Axioms: 1, 2, 3, 5; rules: a, b.
2. Axioms: 1, 2, 3; rules: a, b, c.
3. Axioms: 2, 3; rules: a, b, c, r_1.
4. Axioms: 2; rules: a, b, c, r_1, R_3.
5. Axioms: $\overline{X}X$; rules: a, b, c, r_1, r_2, r_3, r_4.

Bernays also shows, using the same method as before, that these axiom systems are independent, and also the following independence results:[63]

6. Rule c is independent of axioms: 1, 2, 3; rules: a, b, d (showing that in (2), rule c cannot in turn be replaced by d).
7. Rule r_2 is independent of axioms: 1, 3, 5; rules: a, b, (thus showing that in (1) and (2), axiom 2 cannot be replaced by rule r_2).
8. Rule r_3 is independent of axioms: 1, 2; rules: a, b, c (showing similarly, that in (1) and (2), rule r_3 cannot replace axiom 3).
9. Rule R_3 is independent of axioms: $\overline{X}X$, 3; rules: a, b (showing that R_3 is stronger than r_3, since 3 is provable from R_3 and $\overline{X}X$).
10. Rule R_1 is independent of axioms: $\overline{X}X$, 1; rules: a, b (showing that R_1 is stronger than r_1, since 1 is provable from $\overline{X}X$ and R_1).

11. Axiom 2 is independent of axioms: $\overline{X}X$, 1, 3, 5; rules: a, b.
12. Axiom 2 is independent of axioms: $\overline{X}X$; rules: a, b, c, r_1, R_3 (showing that in (5), $\overline{X}X$ together with r_2 is weaker than axiom 2).

The detailed study exhibits, in particular, a sensitivity to the special status of rules like R_3, where subformulas have to be substituted. The discussion foreshadows developments of formal language theory in the 1960s. Bernays also mentions that a rule (corresponding to the contrapositive of axiom 2), allowing inference of $\varphi(\alpha)$ from $\varphi(\alpha\beta)$ would be incorrect (and hence, "there is no such generalization of r_2").

Bernays's discussion of axioms and rules, together with his discussion of expressibility in the "Supplementary remarks to §2–3," shows his acute sensitivity for subtle questions regarding logical calculi. His remarks are quite opposed to the then-prevalent tendency (e.g., Sheffer and Nicod) to find systems with fewer and fewer axioms, and foreshadow investigations of relative strength of various axioms and rules of inference, for example, of Lewis's modal systems, or more recently of the various systems of substructural logics.

At the end of the "Supplementary remarks," Bernays isolates the positive fragment of propositional logic (i.e., the provable formulas not containing negation; here $+$ and $\rightarrow$ are considered primitives) and claimed that he had an axiomatization of it. He did not give an axiom system, but stated that it is possible to choose a finite number of provable sentences as axioms so that completeness follows by a method exactly analogous to the proof given in §3. The remark suggests that Bernays was aware that the completeness proof is actually a proof schema, in the following sense. Whenever a system of axioms is given, one only has to verify that all the equivalences necessary to transform a formula into conjunctive normal form are theorems of that system. Then completeness follows just as it does for the axioms of *Principia*.

In his next set of lectures on the "Logical Calculus" given in the winter semester of 1920 (Hilbert 1920a), Hilbert makes use of the fact that these equivalences are the important prerequisite for completeness. The propositional calculus we find there is markedly different from the one in Hilbert (1918b) and Bernays (1918), but the influences are clearly visible. The connectives are all primitive, not defined, this time. The sole axiom is $\overline{X}X$, and the rules of inference are:

$$\frac{X}{XY}\ \mathrm{b2}, \qquad \frac{\begin{matrix}X\\ Y\end{matrix}}{X+Y}\ \mathrm{b3},$$

plus the rule (b4), stating: "Every formula resulting from a correct formula by transformation is correct." "Transformation" is meant as transformation according to the equivalences needed for normal forms: commutativity, associativity, de Morgan's laws, $\overline{\overline{X}}$ and X, and the definitions of $\rightarrow$ and $=$ (biconditional). These transformations work in both directions, and also on subformulas of formulas (as did R_1 and R_3).[64] One equivalence corresponding to modus ponens must be added, it is: $(X+\overline{X})Y$ is intersubstitutable with Y.

Anyone familiar with the work done on propositional logic elsewhere might be puzzled by this seemingly unwieldy axiom system. It would seem that the system in Hilbert (1920a) is a step backward from the elegance and simplicity of the *Principia* axioms. Adjustments, if they are to be made at all, it would seem, should go in the direction of even more simplicity, reducing the number of primitives (as Sheffer did) and the number of axioms (as in the work of Nicod and later Łukasiewicz). Hilbert was motivated by different concerns. He was interested not only in the simplicity of his axioms but in their efficiency. Decidability, in particular, supersedes considerations of independence and elegance. The presentation in Hilbert (1920a) is designed to provide a decision procedure which is not only efficient but also more intuitive to use for a mathematician trained in algebraic methods. Bernays's study of inference rules made clear, on the other hand, that such an approach can in principle be reduced to the axiomatics of *Principia.* The subsequent work on the decision problem is also not strictly axiomatic, but uses transformation rules and normal forms. The rationale is formulated by Behmann:

> The form of presentation will not be axiomatic, rather, the needs of practical calculation shall be in the foreground. The aim is thus not to reduce everything to a number (as small as possible) of logically independent formulas and rules; on the contrary, I will give as many rules with as wide an application as possible, as I consider appropriate to the practical need. The logical dependence of rules will not concern us, insofar as they are merely of independent practical importance. ... Of course, this is not to say that an axiomatic development is of no value, nor does the approach taken here preempt such a development. I just found it advisable not to burden an investigation whose aim is in large part the exhibition of new results with such requirements, as can later be met easily by a systematic treatment of the entire field. (Behmann 1922, 167)

Such a systematic treatment, of course, was necessary if Hilbert's ideas regarding his logic and foundation of mathematics were to find followers. Starting in (1922c) and (1923), Hilbert presents the logical calculus not in the form of *Principia* but by grouping the axioms governing the different connectives. In (1922c), we find the "axioms of logical consequence," in (1923), "axioms of negation." The first occurrence of axioms for conjunction and disjunction seems to be in a class taught jointly by Hilbert and Bernays during winter 1922–23, and in print in Ackermann's dissertation (Ackermann 1924). The project of replacing the artificial axioms of *Principia* with more intuitive axioms grouped by the connectives they govern, and the related idea of considering subsystems such as the positive fragment, is Bernays's. In 1918, he had already noted that one could refrain from taking $+$ and $\rightarrow$ as defined symbols and consider the problem of finding a complete axiom system for the positive fragment. The notes to the lecture course from 1922–23 (Hilbert and Bernays 1923a, 17) indicate that the material in question was presented by Bernays. In 1923, he gives a talk titled "The role of negation in propositional logic," in which he points out the importance of separating axioms for the different connectives, in particular, giving axioms for negation separately. This emphasis of separating negation from the other connectives is of course necessitated by Hilbert's considerations on finitism as well. Full presentations of the axioms of propositional logic are also found in Hilbert (1928a), and in slightly modified form in a course on logic taught by Bernays in 1929–1930. The axiom system we find there is almost exactly the one later included in Hilbert and Bernays (1934).

I. $A \to (B \to A)$
$(A \to (A \to B)) \to (A \to B)$
$(A \to (B \to C)) \to (B \to (A \to C))$
$(B \to C) \to ((A \to B) \to (A \to C))$

II. $A \,\&\, B \to A$
$A \,\&\, B \to B$
$(A \to B) \to ((A \to C) \to (A \to B \,\&\, C))$

III. $A \to A \vee B$
$B \to A \vee B$
$(B \to A) \to ((C \to A) \to (B \vee C \to A))$

IV. $(A \sim B) \to (A \to B)$
$(A \sim B) \to (B \to A)$
$(A \to B) \to ((B \to A) \to (A \sim B))$

V. $(A \to B) \to (\overline{B} \to \overline{A})$
$(A \to \overline{A}) \to \overline{A}$
$A \to \overline{\overline{A}}$
$\overline{\overline{A}} \to A$.[65]

Bernays (1927) claims that the axioms in groups I–IV provide an axiomatization of the positive fragment and raises the question of a decision procedure. This is where he first follows up on his claim in 1918 that such an axiomatization is possible.

5.5. *Grundzüge der theoretischen Logik*

Hilbert and Ackermann's textbook *Grundzüge der theoretischen Logik* (Hilbert and Ackermann 1928) provided an important summary of the work on logic done in Göttingen in the 1920s. Although (as documented by Sieg 1999), the book is in large parts a polished version of Hilbert's 1917–1918 lectures (Hilbert 1918b), it is important especially for the influence it had in terms of making the work available to an audience outside of Göttingen. Both Gödel and Herbrand, for instance, became acquainted with the methods developed by Hilbert and his students through it.

In addition, *Grundzüge* contained a number of minor but significant improvements over Hilbert (1918b). The first is a much simplified presentation of the axioms of the predicate calculus. Whereas Hilbert (1918b) listed six axioms and three inference rules governing the quantifiers, the formulation in Hilbert and Ackermann (1928) consisted simply in:

e. $(x)F(x) \to F(y)$,

f. $F(y) \to (Ex)F(x)$,

with the following form of the rule of generalization. If $\mathfrak{A} \to \mathfrak{B}(x)$ is provable, and x does not occur in $\mathfrak{A}$, then $\mathfrak{A} \to (x)\mathfrak{B}(x)$ is provable. Similarly, if $\mathfrak{B}(x) \to \mathfrak{A}$ is provable, then so is $(Ex)\mathfrak{B}(x) \to \mathfrak{A}$.

Another important part of *Grundzüge* concerns the semantics of the predicate calculus and the decision problem. The only publication addressing the decision problem had been Behmann (1922); Bernays and Schönfinkel (1928) and Ackermann (1928a) appeared the same year as *Grundzüge* (although Bernays and Schönfinkel's result was obtained much earlier). Thus, the book was important in popularizing the decision problem as a fundamental problem of mathematical foundations. In a similar vein, although the completeness of the propositional calculus had been established already in 1918 by Bernays and in 1920 by Post, the Post-completeness and semantic completeness of predicate logic remained an open problem. Ackermann solved the former in the negative; this result is first reported in *Grundzüge*. It motivates the question of semantic completeness, posed on p. 68: "Whether the axiom system is complete at least in the sense that all logical formulas that are correct for every domain of individuals can be derived from it is still an unsolved question." This offhand remark provided the motivation for Gödel's landmark completeness theorem (see section 8.4).

5.6. The Decision Problem

The origin of the decision problem in Hilbert's work is no doubt his conviction, expressed in his 1900 address to the Paris Congress, that every mathematical problem has a solution:

> This conviction of the solvability of every mathematical problem is a powerful incentive to the worker. We hear within us the perpetual call: There is the problem. Seek its solution. You can find it by pure reason, for in mathematics there is no *ignorabimus*. (Hilbert 1900a, 1102)

A few years later, Hilbert first explicitly took the step that this *no ignorabimus* should be reflected in the decidability of the problem of whether a mathematical statement is derivable from the axiom system for the domain in question:

> So it turns out that for every theorem there are only *finitely many possibilities of proof*, and thus we have solved, in the most primitive case at hand, the old problem that it must be possible to achieve any correct result by a *finite proof*. This problem was the original starting point of all my investigations in our field, and the solution to this problem in the most general case[,] the proof that there can be no "ignorabimus" in mathematics, has to remain the ultimate goal.[66]

Hilbert's emphasis on the axiomatic method was thus not only motivated by providing a formal framework in which questions such as independence, consistency, and completeness could be given mathematical treatment, but so could the question of the solvability of all mathematical problems. In "Axiomatic Thought" (1918a, 1113), the problem of "decidability of a mathematical question in a finite number of operations" is listed as one of the fundamental problems for the axiomatic method.

Without a semantics for first-order logic in hand, it is not surprising that the formulation of the problem as well as the partial results obtained only made reference to derivability from an axiom system. For instance, as discussed, Bernays infers the decidability of the propositional calculus in this sense as a consequence of the completeness theorem. The development of semantics for first-order logic in the following years made it possible to reformulate the decision problem as a question of validity (*Allgemeingültigkeit*) or, dually, as one of satisfiability: "The decision problem is solved, if one knows a procedure which allows for any given logical expression to decide whether it is valid or satisfiable, respectively" (Hilbert and Ackermann 1928, 73). Hilbert and Ackermann (1928) call the decision problem the main problem of mathematical logic. No wonder it was pursued with as much vigor as the consistency problem for arithmetic.

5.6.1. The Decision Problem in the Tradition of Algebra of Logic

In the algebra of logic, results on the decision problem were obtained in the course of work on elimination problems. The first major contribution to the decision problem was Löwenheim's (1915) result. His theorem 4, "There are no fleeing equations between singulary relative coefficients, not even when the relative coefficients of $1'$ and $0'$ are included as the only binary ones" (Löwenheim 1915, 243), amounts to the proposition that every monadic first-order formula, if satisfiable, is satisfiable in a finite domain. Recall from itinerary IV that a fleeing equation is one that is not valid but valid in every finite domain. If there are no fleeing equations between singulary relative coefficients (i.e., monadic predicates), then every monadic formula valid in every finite domain is also valid.

It should be noted that both Löwenheim (1915) and Skolem (1919), who gave a simpler proof, state the theorem as a purely algebraic result. Neither draws the conclusion that the result shows that satisfiability of monadic formulas is decidable, indeed, this only follows by inspection of the particular normal forms they give in their proofs. In particular, the proofs do not contain bounds on the size of the finite models that have to be considered when determining if a formula is satisfiable.

Löwenheim (1915) proved a second important result, namely, that validity of an arbitrary first-order formulas is equivalent to a formula with only binary predicate symbols. This means that dyadic predicate logic forms a reduction class, that is, the decision problem for first-order logic can be reduced to that of dyadic logic. Löwenheim, of course, did not draw this latter conclusion, since he was not concerned with decidability in this sense. He does, however, remark that

> since, now, according to our theorem the whole relative calculus can be reduced to the binary relative calculus, it follows that we can decide whether an arbitrary mathematical proposition is true provided that we can decide whether a binary relative equation is identically satisfied or not. (Löwenheim 1915, 246)

A related result is proved in (Skolem 1920, theorem 1). A formula is in (satisfiability) Skolem normal form if it is a prenex formula and all universal quantifiers precede all existential quantifiers, that is, it is of the form

$$(\forall x_1) \dots (\forall x_n)(\exists y_1) \dots (\exists y_m)A(x_1, \dots, x_n, y_1, \dots, y_m).$$

Skolem's result is that for every first-order formula there is a formula in Skolem normal form that is satisfiable if and only if the original formula is. From this, it follows that the formulas in Skolem normal form are a reduction class as well.

5.6.2. Work on the Decision Problem after 1920

The word *Entscheidungsproblem* first appears in a talk given by Behmann to the Mathematical Society in Göttingen on May 10, 1921, titled "Entscheidungsproblem und Algebra der Logik."[67] Here, Behmann is very explicit about the kind of procedure required, characterizing it as a "mere calculational method," as a procedure following the "rules of a game," and stating its aim as an "elimination of thinking in favour of mechanical calculation."

The result Behmann reports on in this talk is that of his *Habilitationsschrift* (Behmann 1922), in which he proves, independently of Löwenheim and Skolem, that monadic second-order logic with equality is decidable. The proof is by a quantifier elimination procedure, that is, a transformation of sentences of monadic-second order logic (with equality) into a disjunctive normal form involving expressions "there are at least n objects" and "there are at most n objects."

The problem was soon taken up by Moses Schönfinkel, who was a student in Göttingen at the time. In December 1922, he gave a talk to the Mathematical Society in which he proved the decidability of validity of formulas of the form $(\exists x)(\forall y)A$, where A is quantifier-free and contains only one binary predicate symbol (Schönfinkel 1922). This result was subsequently extended by Bernays to apply to formulas with arbitrary many predicate symbols (Bernays and Schönfinkel 1928). The published paper also discusses Behmann's (1922) result and gives a bound on the size of finite models for monadic formulas, as well as the cases of prenex formulas with quantifier prefixes of the form $\forall^* A$, $\exists^* A$ and $\forall^*\exists^*$. In particular, it is shown there that a formula $(\forall x_1)\dots(\forall x_n)(\exists y_1)\dots(\exists y_m)A$ is valid iff it is valid in all domains with n individuals. In its dual formulation, the main result is that satisfiability of prenex formulas with prefix $\exists^*\forall^*$ (the Bernays-Schönfinkel class) is decidable. The result was later extended by Ramsey (1930) to include identity; along the way, Ramsey proved his famous combinatorial theorem.

The result dual to Bernays and Schönfinkel's first, namely, the decidability of satisfiability of formulas of the form $(\forall x)(\exists x)A$, was extended by Ackermann (1928a) to formulas with prefix $\exists^*\forall\exists^*$. The same result was proved independently later the same year by Skolem (1928); this paper as well as the follow-up (1935) also prove some related decidability results.

Herbrand (1930, 1931b) draws some important conclusions regarding the decision problem from his *theorème fondamental* (see following discussion) as well, giving new proofs of the decidability of the monadic class, the Bernays-Schönfinkel class, the Ackermann class, and the Herbrand class (prenex formulas where the matrix is a conjunction of atomic formulas and negated atomic formulas).

The last major partial solution of the decision problem before Church's (1936a) and Turing's (1937) proofs of the undecidability of the general problem was the proof of decidability of satisfiability for prenex formulas with prefix of the form $\exists^*\forall\forall\exists^*$. This was carried out independently by Gödel (1932a), Kalmár (1933), and Schütte (1934a, 1934b). Gödel (1933b) also showed that prenex formulas with prefix $\forall\forall\forall\exists^*$ form a reduction class.[68]

5.7. Combinatory Logic and λ-Calculus

In the early 1920s, there was a significant amount of correspondence between Hilbert and his students (in particular, Bernays and Behmann) and Russell on various aspects of *Principia* (see Mancosu 1999a, 2003). One of the things Russell mentioned to Bernays was Sheffer's (1913) reduction of the two primitive connectives $\sim$ and $\vee$ of *Principia* to the Sheffer stroke. In 1920, Schönfinkel extended this reduction to the quantifiers by means of the operator $|^x$, where $\phi(x) \mid^x \psi(x)$ means "for no x is $\phi(x)$ and $\psi(x)$ both true." Then $(x)\phi(x)$ can be defined by $(\phi(x) \mid^y \phi(x)) \mid^x (\phi(x) \mid^y \phi(x))$. This led Schönfinkel to consider further possibilities of reducing the fundamental notions of the logic of *Principia*, namely, those of propositional function and variables themselves.

In a manuscript written in 1920, and later edited by Behmann and published (1924), Schönfinkel gave a general analysis of mathematical functions, and presented a function calculus based on only application and three basic functions (the combinators). First, Schönfinkel explains how one only needs to consider unary functions: A binary function $F(x, y)$, for instance, may be considered instead as a unary function which depends on the argument x, or, equivalently, as a unary function of the argument x which has a unary function as its value. Hence, $F(x, y)$ becomes $(fx)y$; fx now is the unary function which, for argument y has the same value as the binary function $F(x, y)$. Application associates to the left, so that $(fx)y$ can more simply be written fxy.

Just as functions in Schönfinkel's system can have functions as values, they can also be arguments to other functions. Schönfinkel introduces five primitive functions I, C, T, Z, and S by the equations

$$\begin{aligned} Ix &= x \\ (Cx)y &= x \\ (T\phi)xy &= \phi yx \\ Z\phi\chi x &= \phi(\chi x) \\ S\phi\chi x &= (\phi x)(\chi x) \end{aligned}$$

I is the identity; its value is always simply its argument. C is the constancy function: Cx is the function whose value is always x. T allows the interchange of argument places; $T\phi$ is the function which has as its value for xy the value of ϕyx. Z is the composition function: $Z\phi\chi$ is the function which takes its argument, first applies χ, and then applies ϕ to the resulting value. The fusion function S is similar to composition, but here ϕ is to be thought of as a binary function $F(x, y)$: Then $S\phi\chi x$ is the unary function $F(x, \chi x)$.

So far this constitutes a very general theory of functions. In applying this to logic, Schönfinkel obtains an elegant system in which formulas without free variables can be written without connectives, quantifiers, or variables at all. In light of the reduction to unary functions, first of all relations can be eliminated; for example, instead of a binary relation $R(x, y)$ we have a unary function r from arguments x to functions that themselves take individuals as arguments, and whose value is a truth value. Then, instead of $|^x$, Schönfinkel introduces a new combinator, U: $Ufg = fx \,|^x\, gx$—note that in the expression on the left the bound variable x no longer occurs. Together with the other combinators, this allows Schönfinkel to translate any sentence of even higher-order logic into an expression involving only combinators. For instance, $(f)(Eg)(x)\overline{fx \,\&\, gx}$ first becomes, using $|^x$:

$$[(fx \,|^x\, gx) \,|^g\, (fx \,|^x\, gx)] \,|^f\, [(fx \,|^x\, gx) \,|^g\, (fx \,|^x\, gx)].$$

Now replacing $|^x$ and $|^g$ by the combinator U, we get

$$[U(Uf)(Uf)] \,|^f\, [U(Uf)(Uf)].$$

To remove the last $|^f$, the expressions on either side must end with f; however, $U(Uf)(Uf) = S(ZUU)Uf$, and so finally we get $U[S(ZUU)U][S(ZUU)U]$.

Schönfinkel's ideas were further developed in great detail by Haskell Curry, who wrote a dissertation under Hilbert in 1929 (1929, 1930).[69]

Similar ideas led Church (1932) to develop his system of λ-calculus. Like Schönfinkel's and Curry's combinatory logic, the λ-calculus was intended in the first instance to provide an alternative to Russellian type theory and to set theory as a foundation for mathematics. Like combinatory logic, the λ-calculus is a calculus of functions with *application* (st) as the basic operation; like Curry, Church defined a notion of equality between terms using certain conversion relations. If t is a term in the language of the calculus with free variable x, the λ operator is used to form a new term $\lambda x.t$, which denotes a function with argument x. A term of the form $(\lambda x.t)s$ *converts* to the term $t(x/s)$ (t with all free occurrences of x replaced by s). This is one of three basic kinds of conversion; a term on which no conversion can be carried out is in *normal form*.

Unfortunately, as Kleene and Rosser (1935) showed, both Curry's and Church's systems were inconsistent and hence unsuitable in their original formulation to provide a foundation for mathematics. Nevertheless, combinatory logic and λ-calculus proved incredibly useful as theories of functions; in particular, versions of the λ-calculus were developed as systems of computable functions. In fact, Church's (1936b, 1936a) (negative) solution to the decision problem essentially involved the λ-calculus. Church (1933) and Kleene (1935) found a way to define the natural numbers as certain λ-terms $\bar{n}$ in normal form (Kleene numerals). The notion of λ-definability of a number theoretic function is then simply: A function f is λ-definable if there is a term t such that t applied to the Kleene numeral $\bar{n}$ converts to a normal form which is the Kleene numeral of the value of $f(n)$. Church (1936b) showed that λ-definability coincides with (general) recursiveness and that the problem of deciding whether a term converts to a normal form is not general recursive. Church (1936a) uses this result to show that the decision problem is unsolvable.

5.8. Structural Inference: Hertz and Gentzen

Another important development in logic that originated in Hilbert's school was the introduction of sequent calculus and natural deduction by Gentzen. This grew out of the logical work of Paul Hertz. Hertz was a physicist working in Göttingen between 1912 and 1933. From the 1920s onward, he was also working in philosophy and in particular, logic. In a series of papers (Hertz 1922, 1923, 1928, 1929), he developed a theory of structural inference based on expressions of the form $a_1, \ldots, a_n \to b$. Hertz calls such expressions *sentences*; the signs on the left are the antecedents, the sign on the right the succedent. It is understood that in the antecedents each sign occurs only once. The two rules which he considers are what he calls *syllogism:*

$$\frac{\begin{array}{ll} a_1^1, a_2^1, \ldots \to b^1 & \\ a_1^2, a_2^2, \ldots \to b^2 & \\ \quad\vdots & \\ & a^1, a^2, \ldots, b^1, b^2 \to c \end{array}}{a_1^1, a_2^1, \ldots, a_1^2, a_2^2, \ldots, a^1, a^2 \to c}$$

and *direct inference:*

$$\frac{a_1, a_2, \ldots \to b}{a^1, a^2, \ldots, a_1, a_2, \ldots \to b}\,.$$

In the syllogism, the premises on the left are called lower sentences, the premise on the right the upper sentence of the inference.

A set of sentences is called *closed* if it is closed under these two rules of inference. Hertz's investigations concern in the main criteria for when a closed system of sentences has a set of independent axioms—a concern typical for the Hilbert school. Hertz's other concern, and this is his lasting contribution, is that of proof transformations and normal forms. We cannot give the details of all these results, but a statement of one will give the reader an idea: A sentence is called *tautological* if it is of the form $a \to a$. An *Aristotelian normal proof* is one in which each inference has a nontautological upper sentence that is an initial sentence of the proof (i.e., not the conclusion of another inference). For instance, the following is an Aristotelian normal proof:

$$\frac{\dfrac{\dfrac{a \to b \qquad b \to c}{a \to c} \qquad c \to m}{a \to m} \qquad m, b \to d}{a, b \to d}\,.$$

Hertz proves that every proof can be transformed into an Aristotelian normal proof.

Gentzen's first contribution to logic was a continuation of Hertz's work. In (1933b), Gentzen shows a similar normal form theorem, as well as a completeness result relative to a simple semantics which interprets the elements of the sentences as propositional constants. A sentence $a_1, \ldots, a_n \to b$ is interpreted as: either one of the a_i is false or b is true. Gentzen's result is that if a sentence S follows from (is a tautological consequence of) some other sentences $S_1, \ldots,$ S_n, then there is a proof of a certain normal form of S from $S_1, \ldots, S_n$.[70]

The basic framework of sentences and inferences, as well as the interest in normal form theorems, was contained in Gentzen's more important work on the proof theory of classical and intuitionistic logic. Gentzen (1934) extended Hertz's framework from propositional atoms to formulas of predicate logic. Sentences are there called *sequents*, and the succedent is allowed to contain more than one formula (for intuitionistic logic, the restriction to at most one formula on the right stands). Hertz's direct inference is now called "thinning"; there is an analogous rule for thinning the succedent: The antecedent and succedent of a sequent are now considered sequences of formulas (denoted by uppercase Greek letters). Thus, Gentzen adds rules for changing the order of formulas in a sequent and for contracting two of the same formulas to one. Syllogism is restricted to one lower sentence; this is the cut rule:

$$\frac{\Gamma \to \Theta, A \qquad A, \Delta \to \Lambda}{\Gamma, \Delta \to \Theta, \Lambda}\,.$$

To deal with the logical connectives and quantifiers, Gentzen adapts the axiom systems developed by Hilbert and Bernays in the 1920s by turning the axioms governing a connective into rules introducing the connective in the antecedent and succedent of a sequent. For instance, axiom group (III),

$$\begin{aligned} \text{III.}\quad & A \to A \lor B \\ & B \to A \lor B \\ & (B \to A) \to ((C \to A) \to (B \lor C \to A)), \end{aligned}$$

results in the rules

$$\text{OES:}\quad \frac{\Gamma \to \Theta, A}{\Gamma \to \Theta, A \lor B} \quad \frac{\Gamma \to \Theta, B}{\Gamma \to \Theta, A \lor B} \qquad \text{OEA:}\quad \frac{A, \Gamma \to \Theta \qquad B, \Gamma \to \Theta}{A \lor B, \Gamma \to \Theta}.$$

The rules, together with axioms of the form $A \to A$, result in the system **LK** for classical logic, and **LJ** for intuitionistic logic, where **LJ** is like **LK** with the restriction that each sequent can contain at most one formula in the succedent. The soundness and completeness of these systems is proved in the last section of the paper, by showing that they derive the same formulas as ordinary axiomatic presentations of Hilbert (1928a) and Glivenko (1929) (for the intuitionistic case).

Gentzen's main result (1934) is the *Hauptsatz*. It states that any derivation in **LK** (or **LJ**) can be transformed into one that does not use the cut rule; thus it is now also called the cut-elimination theorem. It has some important consequences: It establishes the decidability of intuitionistic propositional logic, and provides new proofs of the consistency of predicate logic as well as the nonderivability of the principle of the excluded middle in intuitionistic propositional calculus. Gentzen also proves an extension of the Hauptsatz, now called the midsequent theorem: Every derivation of a prenex formula in **LK** can be transformed into one that is cut-free and in which all propositional inferences precede all quantifier inferences. An important consequence of this theorem is a form of Herbrand's theorem (see section 6.4).

The second main contribution of Gentzen (1934) is the introduction of calculi of natural deduction. It was intended to capture actual "natural" reasoning more accurately than axiomatic systems do. Such patterns of reasoning are for instance the methods of conditional proof (to prove a conditional, give a proof of the consequent under the assumption that the antecedent is true) and dilemma (if a conclusion C follows from both A and B individually, it follows from $A \lor B$). In natural deduction then, a derivation is a tree of formulas. The uppermost formulas are assumptions, and each formula is either an assumption, or must follow from preceding formulas according to one of the rules:

$$\frac{A \qquad B}{A \,\&\, B} \qquad \frac{A \,\&\, B}{A} \qquad \frac{A \,\&\, B}{B} \qquad \frac{A}{A \lor B} \qquad \frac{B}{A \lor B} \qquad \frac{A \lor B \quad \begin{matrix}[A] \\ C\end{matrix} \quad \begin{matrix}[B] \\ C\end{matrix}}{C}$$

$$\frac{Fa}{\forall x\, Fx} \qquad \frac{\forall x\, Fx}{Fa} \qquad \frac{Fa}{\exists x\, Fx} \qquad \frac{\exists x\, Fx \qquad \begin{matrix}[Fa]\\ C\end{matrix}}{C}$$

$$\frac{\begin{matrix}[A]\\ B\end{matrix}}{A \supset B} \qquad \frac{A \qquad A \supset B}{B} \qquad \frac{\begin{matrix}[A]\\ \bigwedge\end{matrix}}{\neg A} \qquad \frac{A \quad \neg A}{\bigwedge} \qquad \frac{\bigwedge}{D}.$$

In the rules, the notation $[A]$ indicates that the subproof ending in the corresponding premise may contain any number of formulas for the form A as assumptions, and that the conclusion of the inference is then *independent* of these assumptions. A derivation is a proof of A, if A is the last formula of the derivation and is not dependent on any assumptions.

6. Itinerary VI. Proof Theory and Arithmetic

6.1. Hilbert's Program for Consistency Proofs

The basic aim and structure of Hilbert's program in the philosophy of mathematics is well known: To put classical mathematics on a firm foundation and to rescue it from the attempted *Putsch* of intuitionism, two things were to be accomplished. First, formalize classical mathematics in a formal system; second, give a direct, finitistic consistency proof for this formal system. This project is first outlined in Hilbert (1922c) and received its most popular presentation in "On the infinite" (1926). The project has an important philosophical aspect, which we cannot do justice here (see Zach 2006). This philosophical aspect is the finitist standpoint—the methodological position from which the consistency proofs were to be carried out. At its most basic, the finitist standpoint is characterized as the domain of reasoning about sequences of strokes (the finitist numbers), or sequences of signs in general. From the finitist standpoint, only such finite objects which, according to Hilbert, are "intuitively given" are admissible as objects of finitist reflection; specifically, the finitist standpoint cannot operate with or assume the existence of completed infinite totalities such as the set of all numbers. Furthermore, only such methods of construction and inference are allowed that are immediately grounded in the intuitive representation we have of finitist objects. This includes, for example, definition by primitive recursion and induction as the basic method of proof. A consistency proof for a formal system, in particular, has to take roughly the following form: Give a finitist method by which any given proof in the formal system of classical mathematics can be transformed into one which by its very form cannot be a derivation of a contradiction, such as $0 = 1$. Such a finitist consistency proof not only grounds classical mathematics but also can be taken as a reductio of one of the intuitionist's motivations, viz., that classical reasoning may lead to outright contradictions, since the finitist methods themselves are acceptable intuitionistically.

Hilbert envisaged the consistency proof for classical mathematics to be accomplished in stages of consistency proofs for increasingly stronger systems, starting with propositional logic and ending with full set theory. The crucial development that enabled Ackermann and von Neumann to give partial solutions to the consistency problem was the invention of the ε-calculus around 1922.[71] The ε-calculus is an extension of quantifier-free logic and number theory by term forming ε-operators: If $A(a)$ is a formula, then $\varepsilon_a A(a)$ is a term, intuitively, the least a such that $A(a)$ is true. Using such ε-terms, it is then possible to define the quantifiers by $(\exists a)A(a) \equiv A(\varepsilon_a A(a))$ and $(\forall a)A(a) \equiv A(\varepsilon_a \overline{A(a)})$. The axioms governing the ε-operator are the so-called transfinite axioms

$$A(a) \to A(\varepsilon_a(A(a))) \quad \text{and}$$
$$\varepsilon_a A(a) \neq 0 \to \overline{A(\delta\varepsilon_a A(a))}.$$

The first axiom allows the derivation of the usual axioms for $\exists$ and $\forall$; the second derives the induction axiom (δ is the predecessor function). The ε-substitution method used by Ackermann and von Neumann goes back to an idea of Hilbert: In a given proof, replace the ε-terms by actual numbers so that the result is a derivation of the same formula; then apply the consistency proof for quantifier-free systems.

6.2. Consistency Proofs for Weak Fragments of Arithmetic

Around 1900, Hilbert began championing the axiomatic method as a foundational approach, not only to geometry but also to arithmetic. He proposed the axiomatic method in contradistinction to the *genetic method*, by which the reals were constructed out of the naturals (which were taken as primitive) through the usual constructions of the integer, rational, and finally real numbers through constructions such as Dedekind cuts. In Hilbert's opinion, the axiomatic method is to be preferred for "the final presentation and the complete logical grounding of our knowledge [of arithmetic]" (Hilbert 1900b). The first order of business, then, is to provide an axiomatization of the reals, which Hilbert first attempted in "Über den Zahlbegriff" (1900b). To complete the "logical grounding," however, one would also have to prove the consistency (and completeness) of the axiomatization. For geometry, consistency proofs can be given by exhibiting models in the reals; but a consistency proof of arithmetic requires a direct method. Hilbert considered such a direct proof of consistency the most important question that has to be answered for the axiomatization of the reals, and he formulated it as the second of his "Mathematical problems" (Hilbert 1900a). Attempts at such a proof were made in (Hilbert 1905b) and his course on "Logical principles of mathematical thought" (1905a). It became clear that a successful direct consistency proof requires a further development of the underlying logical systems. This development was carried out by Russell and Whitehead, and following a period of intense study of the *Principia* between 1914 and 1917 in Göttingen (see Mancosu 1999a, 2003), Hilbert renewed his call for a direct consistency proof of arithmetic in “Axiomatic thought” (1918a). This was followed by an increased focus on foundations in Göttingen. Until 1920, Hilbert seems to have been sympathetic to Russell’s logicist approach, but soon became dissatisfied by it. In his course “Problems of mathematical logic,” he explains:

> Russell starts with the idea that it suffices to replace the predicate needed for the definition of the union set by one that is extensionally equivalent, and which is not open to the same objections. He is unable, however, to exhibit such a predicate, but sees it as obvious that such a predicate exists. It is in this sense that he postulates the “axiom of reducibility,” which states approximately the following: “For each predicate, which is formed by referring (once or multiple times) to the domain of predicates, there is an extensionally equivalent predicate, which does not make such reference.
>
> With this, however, Russell returns from constructive logic to the axiomatic standpoint. . . .
>
> The aim of reducing set theory, and with it the usual methods of analysis, to logic, has not been achieved today and maybe cannot be achieved at all. (Hilbert 1920b, 32–33)

Precipitated by increasing interest in Brouwer’s intuitionism and Poincaré’s and Weyl’s predicativist approaches to mathematics (Weyl 1918, 1919), and especially Weyl’s (1921) conversion to intuitionism, Hilbert finally formulated his own approach to mathematical foundations. This approach combined his previous aim of providing a consistency proof that does not proceed by exhibiting a model, or reducing consistency to the consistency of a different theory, with a philosophical position delineating the acceptable methods for a direct consistency proof. In the same course on “Problems of mathematical logic,” he presented a simple axiom system for the naturals, consisting of the axioms

$$
\begin{aligned}
1 &= 1 \\
(a = b) &\to (a + 1 = b + 1) \\
(a + 1 = b + 1) &\to (a = b) \\
(a = b) &\to ((a = c) \to (b = c)) \\
a + 1 &\neq 1.
\end{aligned}
$$

An equation between terms containing only 1’s and +’s is called *correct* if it is either $1 = 1$, results from the axioms by substitution, or is the end formula of a proof from the axioms using modus ponens. The system was later extended by induction, but for the purpose of describing the kind of consistency proof he has in mind, Hilbert observed that the axiom system would be inconsistent in the sense of deriving a formula and its negation iff it were possible to derive a substitution instance of $a+1 = 1$. In this case, then, a direct consistency proof requires a demonstration that no such formula can be the end formula of a formal proof.

> Thus we are led to make the proofs themselves the object of our investigation; we are urged toward a *proof theory*, which operates with the proofs themselves as objects.
>
> For the way of thinking of ordinary number theory the numbers are then objectively exhibitable, and the proofs about the numbers already belong to the area of thought. In our study, the proof itself is something which can be exhibited, and by thinking about the proof we arrive at the solution of our problem.
>
> Just as the physicist examines his apparatus, the astronomer his position, just as the philosopher engages in critique of reason, so the mathematician needs his proof theory, to secure each mathematical theorem by proof critique.[72]

This is the first occurrence of the term "proof theory" in Hilbert's writings.[73] This approach to consistency proofs is combined with a philosophical position in Hilbert's address in Hamburg in July 1921 (1922c), which emphasizes the distinction between the "abstract operation with general concept-scopes [which] has proved to be inadequate and uncertain," and contentual arithmetic which operates on signs. In a famous passage, Hilbert makes clear that the immediacy and security of mathematical "contentual" thought about signs is a precondition of logical thought in general, and hence is the only basis on which a direct consistency proof for formalized mathematics must be carried out:

> As a precondition for the application of logical inferences and for the activation of logical operations, something must already be given in representation: certain extra-logical discrete objects, which exist intuitively as immediate experience before all thought. If logical inference is to be certain, then these objects must be capable of being completely surveyed in all their parts, and their presentation, their difference, their succession (like the objects themselves) must exist for us immediately, intuitively, as something that cannot be reduced to something else. . . . The solid philosophical attitude that I think is required for the grounding of pure mathematics—as well as for all scientific thought, understanding, and communication—is this: *In the beginning was the sign.* (Hilbert 1922c, 1121–1122)

Just as a contentual mathematics of number signs enjoys the epistemological priority claimed by Hilbert, so does contentual reasoning about combinations of signs in general. Hence, contentual reasoning about formulas and formal proofs, in particular, contentual demonstrations that certain formal proofs are impossible, are the aim of proof theory and metamathematics. This philosophical position, together with the ideas about how such contentual reasoning about derivations can be applied to prove consistency of axiomatic systems—ideas outlined in the 1920 course and going back to 1905—make up Hilbert's program for the foundation of mathematics.

In the following two years, Hilbert and Bernays elaborate the research project in a series of courses and talks (Hilbert 1922a, 1923; Hilbert and Bernays 1923b; Bernays 1922). The courses from 1921–1922 and 1922–1923 are most important. It is there that Hilbert introduces the ε-calculus in 1921–1922 to deal with quantifiers and the approach using the ε-substitution method as a proof of consistency for systems containing quantification and induction. The system used in 1922–23 is given by the following axioms (Hilbert and Bernays 1923b, 17, 19):

1. $A \to B \to A$
2. $(A \to A \to B) \to A \to B$
3. $(A \to B \to C) \to (B \to A \to C)$
4. $(B \to C) \to (A \to B) \to A \to C$
5. $A \,\&\, B \to A$
6. $A \,\&\, B \to B$
7. $A \to B \to A \,\&\, B$
8. $A \to A \vee B$
9. $B \to A \vee B$
10. $(A \to C) \to (B \to C) \to A \vee B \to C$
11. $A \to \overline{A} \to B$
12. $(A \to B) \to (\overline{A} \to B) \to B$
13. $a = a$
14. $a = b \to A(a) \to A(b)$
15. $a + 1 \neq 0$
16. $\delta(a + 1) = a$

Here, " $+\,1$" is a unary function symbol. In Hilbert's systems, Latin letters are variables; in particular, a, b, $c, \ldots$ are individual variables and A, B, $C, \ldots$ are formula variables. The rules of inference are modus ponens and substitution for individual and formula variables.

The idea of the consistency proof is this: Suppose a proof of a contradiction is available. (We may assume that the end formula of this proof is $0 \neq 0$.)

1. *Resolution into proof threads.* First, we observe that by duplicating part of the proof and leaving out steps, we can transform the derivation to one where each formula (except the end formula) is used exactly once as the premise of an inference. Hence, the proof is in tree form.

2. *Elimination of variables.* We transform the proof so that it contains no free variables. This is accomplished by proceeding backward from the end formula: The end formula contains no free variables. If a formula is the conclusion of a substitution rule, the inference is removed. If a formula is the conclusion of modus ponens it is of the form

$$\frac{\mathfrak{A} \qquad \mathfrak{A} \to \mathfrak{B}}{\mathfrak{B}'},$$

where $\mathfrak{B}'$ results from $\mathfrak{B}$ by substituting terms (*functionals*, in Hilbert's terminology) for free variables. If these variables also occur in $\mathfrak{A}$, we substitute the same terms for them. Variables in $\mathfrak{A}$ that do not occur in $\mathfrak{B}$ are replaced with 0. This yields a formula $\mathfrak{A}'$ not containing variables. The inference is replaced by

$$\frac{\mathfrak{A}' \qquad \mathfrak{A}' \to \mathfrak{B}'}{\mathfrak{B}'}.$$

3. *Reduction of functionals.* The remaining derivation contains a number of terms which now have to be reduced to numerical terms (i.e., standard numerals of the form $(\ldots(0+1)+\cdots)+1)$. In this case, this is done easily by rewriting innermost subterms of the form $\delta(0)$ by 0 and $\delta(\mathfrak{n}+1)$ by $\mathfrak{n}$. In later stages, the set of terms is extended by function symbols introduced by recursion, and the reduction of functionals there proceeds by calculating the function for given numerical arguments according to the recursive definition.

To establish the consistency of the axiom system, Hilbert suggests, we have to find a decidable property of formulas (*konkret feststellbare Eigenschaft*) so that every formula in a derivation which has been transformed using the foregoing steps has the property, and the formula $0 \neq 0$ lacks it. The property Hilbert proposes to use is *correctness.* This, however, is not to be understood as truth in a model: The formulas still occurring in the derivation after the transformation are all Boolean combinations of equations between numerals. An equation between numerals $\mathfrak{n} = \mathfrak{m}$ is *correct* if $\mathfrak{n}$ and $\mathfrak{m}$ are equal, and the negation of an equality is correct if $\mathfrak{n}$ and $\mathfrak{m}$ are not equal.

> If we call a formula which does not contain variables or functionals other than numerals an "*explicit* [*i.e., numerical*] *formula*," then we can express the result obtained thus: Every provable explicit formula is end formula of a proof all the formulas of which are explicit formulas.
>
> This would have to hold in particular of the formula $0 \neq 0$, if it were provable. The required proof of consistency is thus completed if we show that there can be no proof of the formula which consists of only explicit formulas.
>
> To see that this is impossible it suffices to find a concretely determinable [*konkret feststellbar*] property, which first of all holds of all explicit formulas which result from an axiom by substitution, which furthermore transfers from premises to end formula in an inference, which however does not apply to the formula $0 \neq 0$. (Hilbert 1922b, part 2, 27–28)

This basic model for a consistency proof is then extended to include terms containing function symbols defined by primitive recursion and terms containing the ε-operator. Hilbert's *Ansatz* for eliminating ε-terms from formal derivations is first outlined in the 1921–1922 lectures and in more detail in the 1922–1923 course.[74]

Suppose a proof involves only one ε-term $\varepsilon_a \mathfrak{A}(a)$ and corresponding *critical formulas*

$$\mathfrak{A}(\mathfrak{k}_i) \to \mathfrak{A}(\varepsilon_a \mathfrak{A}(a)),$$

that is, substitution instances of the transfinite axiom

$$A(a) \to A(\varepsilon_a A(a)).$$

We replace $\varepsilon_a \mathfrak{A}(a)$ everywhere with 0, and transform the proof as before by rewriting it in tree form ("dissolution into proof threads"), eliminating free variables, and evaluating numerical terms involving primitive recursive functions. Then the critical formulas take the form

$$\mathfrak{A}(\mathfrak{z}_i) \rightarrow \mathfrak{A}(0),$$

where $\mathfrak{z}_i$ is the numerical term to which $\mathfrak{k}_i$ reduces. A critical formula can now only be false if $\mathfrak{A}(\mathfrak{z}_i)$ is true and $\mathfrak{A}(0)$ is false. If that is the case, repeat the procedure, now substituting $\mathfrak{z}_i$ for $\varepsilon_a \mathfrak{A}(a)$. This yields a proof in which all initial formulas are correct and no ε terms occur.

If critical formulas of the second kind, that is, substitution instances of the induction axiom,

$$\varepsilon_a A(a) \neq 0 \rightarrow \overline{A(\delta \varepsilon_a A(a))},$$

also appear in the proof, the witness $\mathfrak{z}$ has to be replaced with the least $\mathfrak{z}'$ so that $\mathfrak{A}(\mathfrak{z}')$ is true.

The challenge is to extend this procedure to (a) cover more than one ε-term in the proof, (b) take care of nested ε-terms, and last (c) extend it to second-order ε's and terms involving them, that is, $\varepsilon_f \mathfrak{A}_a(f(a))$, which are used in formulations of second-order arithmetic. This was attempted in Ackermann's (1924) dissertation.

6.3. Ackermann and von Neumann on Epsilon Substitution

Ackermann's dissertation (1924) is a milestone in the development of proof theory. The work contains the first unified presentation of a system of second-order arithmetic based on the ε-calculus, a complete and correct consistency proof of the ε-less fragment (an extension of what is now known as primitive recursive arithmetic, PRA), and an attempt to extend Hilbert's ε-substitution method to the full system.

The consistency proof for the ε-free fragment extends a sketch of a consistency proof for primitive recursive arithmetic contained in Hilbert and Bernays's 1922–1923 lectures. For primitive recursive arithmetic, the basic axiom system is extended by definitional equations for function symbols which define the corresponding functions recursively, for example,

$$\psi(0, \vec{c}) = \mathfrak{a}(\vec{c})$$
$$\psi(a+1, \vec{c}) = \mathfrak{b}(a, \psi(a, \vec{c}), \vec{c}).$$

To prove consistency for such a system, the "reduction of functionals" step has to be extended to deal with terms containing the function symbols defined by evaluating innermost terms with leading function symbol ψ according to the primitive recursion specified by the defining equations. It should be noted right away that such a consistency proof requires the possibility of evaluating an arbitrary primitive recursive function, and as such exceeds primitive recursive methods. This means that Hilbert, already in 1922, accepted nonprimitive recursive methods as falling under the methodological, "finitary" standpoint of proof theory. Ackermann's dissertation extends this consistency proof by also dealing with what might be called second-order primitive recursion. A second-order primitive recursive definition is of the form

$$\phi_{\vec{b}_i}(0, \vec{f}(\vec{b}_i), \vec{c}) = \mathfrak{a}_{\vec{b}_i}(\vec{f}(\vec{b}_i), \vec{c})$$
$$\phi_{\vec{b}_i}(a+1, \vec{f}(\vec{b}_i), \vec{c}) = \mathfrak{b}_{\vec{b}_i}(a, \phi_{\vec{d}_i}(a, \vec{f}(\vec{d}_i), \vec{c}), \vec{f}(\vec{b}_i)).$$

The subscript notation indicates λ-abstraction; in modern notation the schema would more conspicuously be written as

$$\phi(0, \lambda\vec{b}_i.\vec{f}(\vec{b}_i), \vec{c}) = \mathfrak{a}(\lambda\vec{b}_i.\vec{f}(\vec{b}_i), \vec{c})$$
$$\phi(a+1, \lambda\vec{b}_i.\vec{f}(\vec{b}_i), \vec{c}) = \mathfrak{b}(a, \phi(a, \lambda\vec{d}_i.\vec{f}(\vec{d}_i), \vec{c}), \lambda\vec{b}_i.\vec{f}(\vec{b}_i)).$$

Second-order primitive recursion allows the definition of the Ackermann function, which was shown by Ackermann (1928b) to be itself not primitive recursive.

The first consistency proof given by Ackermann is for this system of second-order primitive recursive arithmetic. While for PRA, the reduction of functionals only requires the relatively simple evaluation of primitive recursive terms, the situation is more complicated for second-order primitive recursion. Ackermann locates the difficulty in the following: Suppose you have a functional $\phi_b(2, \mathfrak{b}(b))$, where ϕ is defined by

$$\phi_b(0, f(b)) = f(1) + f(2)$$
$$\phi_b(a+1, f(b)) = \phi_b(a, f(b)) + f(a) \cdot f(a+1).$$

Here, $\mathfrak{b}(b)$ is a term that denotes a function, and so there is no way to replace the variable b with a numeral before evaluating the entire term. In effect, the variable b is bound (in modern notation, the term might be more suggestively written $\phi(2, \lambda b.\mathfrak{b}(b))$). To reduce this term, we apply the recursion equations for ϕ twice and end up with a term like

$$\mathfrak{b}(1) + \mathfrak{b}(2) + \mathfrak{b}(0) \cdot \mathfrak{b}(1) + \mathfrak{b}(1) \cdot \mathfrak{b}(2).$$

The remaining $\mathfrak{b}$'s might in turn contain ϕ, for example, $\mathfrak{b}(b)$ might be $\phi_c(b, \delta(c))$, in which case the above expression would be

$$\phi_c(1, \delta(c)) + \phi_c(2, \delta(c)) + \phi_c(0, \delta(c)) \cdot \phi_c(1, \delta(c)) + \phi_c(1, \delta(c)) \cdot \phi_c(2, \delta(c)).$$

By contrast, reducing a term $\psi(\mathfrak{z})$ where ψ is defined by first-order primitive recursion results in a term which does not contain ψ, but only the function symbols occurring on the right-hand side of the defining equations for ψ.

To overcome this difficulty, Ackermann defines a system of indexes of terms containing second-order primitive recursive terms and an ordering on these indexes. Ackermann's indexes are, essentially, ordinal notations for ordinals $< \omega^{\omega^\omega}$, and the ordering he defines corresponds to the ordering on the ordinals. He then defines a procedure to evaluate such terms by successively applying the defining equations; each step in this procedure results in a new term whose index is less than the index of the preceding term. Because the ordering of the indexes is well founded, this constitutes a proof that the procedure always terminates, and hence that the process of reduction of functionals in the consistency proof comes to an end, resulting in a proof with only correct equalities and inequalities between numerical terms (not containing function symbols).[75] This proof very explicitly proceeds by transfinite induction up to ω^{ω^ω}, and foreshadows Gentzen's (1936) use of transfinite induction up to ε_0. Ackermann was completely aware of the involvement of transfinite induction in this case, but did not see in it a violation of the finitist standpoint:

> The disassembling of functionals by reduction does not occur in the sense that a finite ordinal is decreased each time an outermost function symbol is eliminated. Rather, to each functional corresponds as it were a transfinite ordinal number as its rank, and the theorem that a constant functional is reduced to a numeral after carrying out finitely many operations corresponds to the other [theorem], that if one descends from a transfinite ordinal number to ever smaller ordinal numbers, one has to reach zero after a finite number of steps. Now there is naturally no mention of transfinite sets or ordinal numbers in our metamathematical investigations. It is however interesting, that the mentioned theorem about transfinite ordinals can be formulated so that there is nothing transfinite about it any more. (Ackermann 1924, 13–14)

The full system for which Ackermann attempted to give a consistency proof in the second part of the dissertation consists of the system of second-order primitive recursive arithmetic together with the transfinite axioms:

1. $A(a) \to A(\varepsilon_a A(a))$ $\qquad$ $A_a(f(a)) \to A_a((\varepsilon_f A_b(f(b))(a)))$
2. $A(\varepsilon_a A(a)) \to \pi_a A(a) = 0$ $\qquad$ $A_a(\varepsilon_f A_b(f(b))(a)) \to \pi_f A_a(f(a)) = 0$
3. $\overline{A(\varepsilon_a A(a))} \to \pi_a A(a) = 1$ $\qquad$ $\overline{A_a(\varepsilon_f A_b(f(b))(a))} \to \pi_f A_a(f(a)) = 1$
4. $\varepsilon_a A(a) \neq 0 \to \overline{A(\delta(\varepsilon_a A(a)))}$.

The intuitive interpretation of ε and π, based on these axioms is this: $\varepsilon_a \mathfrak{A}(a)$ is a witness for $\mathfrak{A}(a)$ if one exists, and $\pi_a \mathfrak{A}(a) = 1$ if $\mathfrak{A}(a)$ is false for all a, and $= 0$ otherwise. The π functions are not necessary for the development of mathematics in the axiom system. They do, however, serve a function in the consistency proof, viz., to keep track of whether a value of 0 for $\varepsilon_a \mathfrak{A}(a)$ is a "default value" (i.e., a trial substitution for which $\mathfrak{A}(a)$ may or may not be true) or an actual witness (a value for which $\mathfrak{A}(a)$ has been found to be true).

To give a consistency proof for this system, Ackermann first has to extend the ε-substitution method to deal with proofs in which terms containing more than one ε-operator (and corresponding critical formulas) occur, and then argue (finitistically), that the procedure so defined always terminates in a substitution of numerals for ε-terms which transform the critical formulas into correct formulas of the form $\mathfrak{A}(\mathfrak{t}) \to \mathfrak{A}(\mathfrak{s})$ (where $\mathfrak{A}$, $\mathfrak{t}$, and $\mathfrak{s}$ do not contain ε-operators or primitive recursive function symbols). To solve the first task, Ackermann has to deal with the various possibilities in which ε-operators can occur in the scope of other ε's. For instance, an instance of the transfinite axiom might be

$$\mathfrak{A}(\mathfrak{t}, \varepsilon_y \mathfrak{B}(y)) \to \mathfrak{A}(\varepsilon_x \mathfrak{A}(x, \varepsilon_y \mathfrak{B}(y)), \varepsilon_y \mathfrak{B}(y)).$$

To find a substitution for $\varepsilon_x \mathfrak{A}(x, \varepsilon_y \mathfrak{B}(y))$ here, it is necessary to first have a substitution for $\varepsilon_y \mathfrak{B}(y)$. This case is rather benign, since the value for $\varepsilon_y \mathfrak{B}(y)$ can be determined independently of that for $\varepsilon_x \mathfrak{A}(x, \varepsilon y \mathfrak{B}(y))$. If $\varepsilon_y \mathfrak{B}(y)$ occurs in the term $\mathfrak{t}$ on the left-hand side, the situation is more complicated. We might have, for example, a critical formula of the form

$$\mathfrak{A}(\varepsilon_y \mathfrak{B}(y, \varepsilon_x A(x))) \to \mathfrak{A}(\varepsilon_x \mathfrak{A}(x)).$$

With an initial substitution of 0 for $\varepsilon_x \mathfrak{A}(x)$, we can determine a value for $\varepsilon_y \mathfrak{B}(y, \varepsilon_x \mathfrak{A}(x))$, that is, for $\varepsilon_y \mathfrak{B}(y, 0)$. With this value for $\varepsilon_y \mathfrak{B}(y)$, we then find a value for $\varepsilon_x \mathfrak{A}(x)$. This, however, now might change the "correct" substitution for $\varepsilon_x \mathfrak{A}(x)$, say to n, and hence the initial determination of the value of the term on the left-hand side changes: We now need a value for $\varepsilon_y \mathfrak{B}(y, \mathfrak{n})$.

The procedure proposed by Ackermann is too involved to be discussed here (see Zach 2003 for details). In short, what is required is an ordering of terms based on the level of nesting and of cross-binding of ε's, and a procedure based on this ordering which successively approximates a "solving substitution," that is, an assignment of numerals to ε-terms which results in all correct critical formulas. In this successive approximation, the values found for some ε-terms may be discarded if the substitutions for enclosed ε-terms change. A correct consistency proof would then require a proof that this procedure does in fact always terminate with a solving substitution. Unfortunately, Ackermann's argument in this regard is opaque.

The system to which Ackermann applied the ε-substitution method, as indicated, is a system of second-order arithmetic. Ackermann (and Bernays) soon realized that the proposed consistency proof had problems. Already in the published version, a footnote on p. 9 restricts the system in the following way: Only such terms are allowed in substitutions for formula and function variables in which individual variables do not occur in the scope of a second-order ε. Von Neumann clarified the restriction and its effect: In Ackermann's system, the second-order ε-axiom $A(f) \to \varepsilon_f A(f)$ does duty for the comprehension principle. In this system, the comprehension principle is $(\exists f)(\forall x)(f(x) = t)$, where t is a term possibly containing ε-terms. Under Ackermann's restriction, only such instances of the comprehension principle are permitted in which x is not in the scope of a second-order ε-operator; essentially this guarantees the existence of only such f's which can be defined by arithmetical formulas. Von Neumann (1927) also remarked that Ackermann's restriction makes the system predicative; it is roughly of the strength of the system ACA_0.

This alone restricts the consistency proof to a system much weaker than analysis; however, other problems and lacunae were known to Ackermann, one being that the proof does not cover ε-extensionality,

$$(\forall f)(A(f) \leftrightarrow B(f)) \rightarrow \varepsilon_f A(f) = \varepsilon_f B(f),$$

which serves as the ε-analog of the axiom of choice. Ackermann continued to work on the proof, amending and correcting the ε-substitution procedure even for first-order ε-terms. These corrections used ideas of von Neumann (1927), which was already completed in 1925. Von Neumann (1927) used a different terminology than Ackermann, and the precise connection between Ackermann's and von Neumann's proofs is not clear. Von Neumann's system does not include the induction axiom explicitly, because induction can be proved once a suitable second-order apparatus is available. Hence, the consistency proof for the first-order fragment of his theory does not deal with induction, whereas Ackermann's system has an induction axiom in the form of the second ε-axiom, and his substitution procedure takes into account critical formulas of this second kind. Another significant feature of von Neumann's proof is the precision with which it is executed: Von Neumann gives numerical bounds for the number of steps required until a solving substitution is found.[76]

Ackermann gave a revised ε-substitution proof, using von Neumann's ideas, and communicated it to Bernays in 1927. Both Ackermann and Bernays believed that the new proof would go through for full first-order arithmetic. Hilbert reported on this result in his lectures in Hamburg 1928 (1928a) (see also Bernays 1928b) and Bologna (Hilbert 1928b, 1929). Only with Gödel's (1930b, 1931) incompleteness results did it become clear that the consistency proofs did not even go through for first-order arithmetic. Bernays later gave an analysis of Ackermann's second proof (Hilbert and Bernays 1939) and showed that the bounds obtained hold for induction restricted to quantifier-free formulas, but not for induction axioms of higher complexity. Ackermann eventually, using ideas from Gentzen, gave an ε-substitution proof for full first-order arithmetic (1940).

6.4. Herbrand's Theorem

Herbrand's (1930) thesis "Investigations in proof theory" marks another milestone in the development of first-order proof theory. Herbrand's main influences in this work were Russell and Whitehead's *Principia*, from which he took the notation and some of the presentations of his logical axioms, the work of the Hilbert school, which provided the motivations and aims for proof theoretic research; and Löwenheim's (1915) and Skolem's (1920) work on normal forms. The thesis contains a number of important results, among them a proof of the deduction theorem and a proof of quantifier elimination for induction-free successor arithmetic (no addition or multiplication). The most significant contribution, of course, is Herbrand's theorem.

Herbrand's theorem shares a fundamental feature with Hilbert's approaches to proof theory and consistency proofs: Consistency for systems including quantifiers (ε-terms) is established by giving a procedure that removes quantifiers from a proof, reducing proofs containing such "ideal elements" to quantifier-free (essentially, propositional) proofs. Herbrand's theorem provides a general necessary and sufficient condition for when a formula of the predicate calculus is provable by reducing such provability to the provability of an associated "expansion" in the propositional calculus. The way such an expansion is obtained is closely related to obtaining a Skolem normal form of the formula. The Löwenheim–Skolem theorem reduces the validity of a formula in general to its validity in a canonical countable model. Skolem's and Löwenheim's methods, however, were semantic and used infinitary methods, both features that make it unsuitable for employment in the framework of Hilbert's finitist program. Herbrand's theorem can thus be seen as giving finitary meaning to the Löwenheim–Skolem theorem.

Let us now give a brief outline of the theorem. We follow Herbrand (1931b), which is in some respects clearer than the original (1930). Suppose A is a formula of first-order logic. For simplicity, we assume A is in prenex normal form; Herbrand gave his argument without making this restriction. So let P be $(\mathsf{Q}_1 x_1)\dots(\mathsf{Q}_n x_n)B(x_1, \dots, x_n)$, where Q_i is either $\forall$ or $\exists$, and B is quantifier-free. Then the Herbrand normal form H of A is obtained by removing all existential quantifiers from the prefix of A, and replacing each universally quantified x_i by a term $f_i(x_{j_1}, \dots, x_{j_n})$, where $x_{j_1}, \dots, x_{j_n}$ are the existentially quantified variables preceding x_i. Herbrand (1931a) calls this the elementary proposition associated with P, and f_i is the index function associated with x_i.

To state the theorem, we have to define what Herbrand calls *canonical domains of order k*. This notion, in essence, is a first-order interpretation with the domain being the term model generated from certain initial elements and function, and the terms all have height $\leq k$. (The height of a term is defined as usual: Constants have height 0, and a term $f_j(t_1, \dots, t_k)$ has height $h + 1$ if h is the maximum of the heights of $t_1, \dots, t_k$.) Herbrand did not use terms explicitly as objects of the domain, but instead considered domains consisting of letters, such that each term (of height $\leq k$) has an element of the domain associated with it as its value and such that if terms $t_1, \dots, t_k$ have values $b_1, \dots, b_k$, and the value of $f_i(b_1, \dots, b_k)$ is c, then the value associated with $f(t_1, \dots, t_k)$ is also c. A domain is *canonical* if it furthermore satisfies the condition that any two distinct terms have distinct values associated with them (i.e., the domain is freely generated from the initial elements and the function symbols). Last, a domain is of order k if each term of height $\leq k$ with constants only from among the initial elements has a value in the domain, but some term of height $k + 1$ does not.

The canonical domain of order k associated with P then is the canonical domain of order k with some nonempty set of initial elements and the functions occurring in the Herbrand normal form H of P. P is true in the canonical domain if some substitution of elements for the free variables in H makes H true in the domain, and false otherwise. Herbrand's statement of the theorem then is:

1. If [for some k] there is no system of logical values [truth value assignment to the atomic formulas] making P false in the associated canonical domain of order k, then P is an identity [provable in first-order logic].
2. If P is an identity, then there is a number k obtainable from the proof of P, such that there is no system of logical values making P false in every associated canonical domain of order equal to or greater than k. (Herbrand 1931b, 229)

By introducing canonical domains of order k, Herbrand has thus reduced provability of P in the predicate calculus to the validity of H in certain finite term models. If $H_1, \ldots, H_{n_k}$ are all the possible substitution instances of H in the canonical domain of order k, then the theorem may be reformulated as: (1) If $\bigvee H_i$ is a tautology, then P is provable in first-order logic; (2) if P is provable in first-order logic, then there is a k obtainable from the proof of P so that $\bigvee H_i$ is a tautology.

Herbrand's original proof contained a number of errors that were found by Peter Andrews and corrected by Dreben, Andrews, and Aanderaa (1963); Gödel had independently found a correction (see Goldfarb 1993; Andrews 2003 gives a detailed account of the discovery of the errors). Gentzen (1934) gave a different proof based on the midsequent theorem, which, however, only applies to prenex formulas and does not provide a bound on the size of the Herbrand disjunction $\bigvee H_i$. Another early complete and correct proof was given by Bernays (Hilbert and Bernays 1939) using the ε-calculus.

Herbrand was able to apply the fundamental theorem to give consistency proofs of various fragments of arithmetic, including the case of arithmetic with quantifier-free induction. The idea is to reduce the consistency of arithmetic with quantifier-free induction to induction-free (primitive recursive) arithmetic. This is done by introducing new primitive recursive functions that "code" the induction axioms used. The proof of Herbrand's theorem then produces finite term models for the remaining axioms, and consistency is established (Herbrand 1931a).

6.5. Kurt Gödel and the Incompleteness Theorems

Hilbert had two main aims in his program in the foundation of mathematics: first, a finitistic consistency proof of all of mathematics, and second, a precise mathematical justification for his belief that all well-posed mathematical problems are solvable, that is, that "in mathematics, there is no *ignorabimus*." This second aim resulted in two specific convictions: that the axioms of mathematics, in particular, of number theory, are complete in the sense that for every formula A, either A or $\sim A$ is provable,[77] and second that the validities of first-order logic are decidable (the decision problem). The hopes of achieving both aims were dashed in 1930, when Gödel proved his incompleteness theorems (1930b, 1931). The summary of his results (Gödel 1930b) addresses the impact of the results quite explicitly:

> I. The system S [of *Principia*] is *not* complete [*entscheidungsdefinit*]; that is, it contains propositions A (and we can in fact exhibit such propositions) for which neither A nor $\overline{A}$ is provable and, in particular, it contains (even for decidable properties F of natural numbers) undecidable problems of the simple structure $(Ex)F(x)$, where x ranges over the natural numbers.
>
> II. Even if we admit all the logical devices of *Principia mathematica* ... in metamathematics, there does *not* exist a *consistency proof* for the system S (still less so if we restrict the means of proof in any way). (Gödel 1930b, 141–143)

Soon thereafter, Church and Turing were able to show, using some of the central ideas in Gödel (1931), that the remaining aim of proving the decidability of predicate logic was likewise doomed to fail (Church 1936a,b; Turing 1937).

Gödel obtained his results in the second half of 1930. After proving the completeness of first-order logic, a problem posed by Hilbert and Ackermann (1928), Gödel set to work on proving the consistency of analysis (recall that according to Hilbert (1929), the consistency of arithmetic was already established). Instead of directly giving a finitistic proof of analysis, Gödel attempted to first reduce the consistency of analysis to that of arithmetic, which led him to consider ways to enumerate the symbols and proofs of analysis in arithmetical terms. It soon became evident to him that truth of number-theoretic statements is not definable in arithmetic, by reasoning analogous to the liar paradox. By the end of summer 1930, he had a proof that the analogous fact about *provability* is formalizable in the system of *Principia*, and hence that there are undecidable propositions in *Principia*. At a conference in Königsberg in September 1930, Gödel mentioned the result to von Neumann, who inquired whether the result could be formalized not only in type theory but already in first-order arithmetic. Gödel subsequently showed that the coding mechanism he had come up with could be carried out with purely arithmetical methods using the Chinese remainder theorem. Thus the first incompleteness theorem, that arithmetic contains undecidable propositions, was established. The second incompleteness theorem, namely, that in particular the statement formalizing consistency of number theory is such an undecidable arithmetical statement, was found shortly thereafter (and also independently by von Neumann).[78]

Let us now give a brief outline of the proof. The system P Gödel considers is a version of simple type theory in addition to Peano arithmetic. To carry out the formalization of predicates about formulas and proofs, Gödel introduces what is now known as "Gödel numbering." To each symbol of the system P

a natural number is associated. A finite sequence of symbols a (e.g., a formula) can then be coded by $\Phi(a) = 2^{n_1} \cdot 3^{n_2} \cdots p_k^{n_k}$, where k is the length of the sequence, p_k is the kth prime, and n_i is the Gödel code of the ith symbol in the sequence. Similarly, a sequence of formulas (i.e., a sequence of sequences of numbers) with codes $n_1, \ldots, n_k$ is coded by $2^{n_1} \cdot 3^{n_2} \cdots p_k^{n_k}$.

To carry out the metamathematical treatment of formulas and proofs within the system, Gödel next defines the class of primitive recursive functions and relations of natural numbers (he simply calls them "recursive") and proves (theorems I–IV) that primitive recursive functions and relations are closed under composition, the logical operations of negation, disjunction, conjunction, bounded minimization, and bounded quantification. Using this characterization, he then shows that a collection of 45 functions can be defined primitive recursively. The functions are those necessary to carry out simple manipulations on formulas and proofs, or represent predicates about formulas and proofs. For instance, (31) is the function $Sb(x^v_y)$, the function the value of which is the code of a formula that results from the formula A (with code x) where every free occurrence of the variable with code v is replaced by the term with code y; (45) is the primitive recursive relation xBy, which holds if x is the code of a proof of a formula with code y. (46), finally is $Bew(x)$, expressing that x is the code of a provable formula with code x. $Bew(x)$ is not primitive recursive, because it results from xBy by unbounded existential generalization: $Bew(x) \equiv (Ey)yBx$. Gödel then proves (theorem V) that every recursive relation is numeralwise representable in P, that is, that if $R(x_1, \ldots, x_n)$ is a formula representing a recursive relation (according to the characterization of recursive relations given in theorems I–IV), then:

1. if $R(n_1, \ldots, n_k)$ is true, then P proves Bew(m), where m is the code of $R(n_1, \ldots, n_k)$, and
2. if $R(n_1, \ldots, n_k)$ is false, then P proves Bew(m), where m is the code of $\sim R(n_1, \ldots, n_k)$.

Then Gödel proves the main theorem.

> Theorem VI For every ω-consistent recursive class κ of FORMULAS there are recursive CLASS SIGNS r such that neither v Gen r nor Neg(v Gen r) belongs to Flg(κ) (where v is the FREE VARIABLE of r. (Gödel 1931, 173)

Here κ is the recursive relation defining a set of codes of formulas to be considered as axioms, r is the code of a recursive formula $A(v)$ (i.e., one containing no unbounded quantifiers) with free variable v, v Gen r is the code of the generalization $(v)A(v)$ of $A(v)$, Neg(v Gen r) the code of its negation $\sim(v)A(v)$, and Flg(κ) is the set of codes of formulas that are provable in P together with κ. We may thus restate theorem IV somewhat more perspicuously thus: If P_κ is an ω-consistent theory resulting by adding a recursive set of axioms κ to P, then there is a formula $A(x)$ such that neither $(x)A(x)$ nor

$\sim(x)A(x)$ is provable in P_κ. The requirement that P_κ is ω-consistent states that for no formula $A(x)$ does P_κ prove both $A(n)$ for all numerals n and $\sim(x)A(x)$; Rosser (1936) later weakened this requirement to the simple consistency of P_κ.

In the following sections, Gödel sharpens the result in several ways. First, he shows that (theorem VII) primitive recursive relations are arithmetical, that is, that the basic functions $+$, and $\times$ of arithmetic suffice to express all primitive recursive functions (this is where the Chinese remainder theorem is used). From this, theorem VIII follows, namely, that not only are there undecidable propositions of the form $(x)A(x)$ with A recursive (in particular, possibly using exponentiation x^y) but even with $A(x)$ arithmetical (i.e., containing only $+$ and $\times$). Finally, in section 4, Gödel states the second incompleteness theorem.

> Theorem XI Let κ be any recursive consistent class of FORMULAS; then the SENTENTIAL FORMULA stating that κ is consistent is not κ-PROVABLE; in particular, the consistency of P is not provable in P, provided P is consistent (in the opposite case, of course, every proposition is provable). (Gödel 1931, 193)

Although theorems VI and XI are formulated for the relatively strong system P, Gödel remarks that the only properties of P which enter into the proof of theorem VI are that the axioms are recursively definable, and that the recursive relations can be defined within the system. This applies, so Gödel, also to systems of set theory as well as to number theoretical systems such as that of von Neumann (1927).

Gödel's result is of great importance to the development of mathematical logic after 1930, but its most immediate impact at the time consisted in the doubts it cast on the feasibility of Hilbert's program. Von Neumann and Bernays immediately realized that the result shows that no consistency proof for a formal system of mathematics can be given by methods which can be formalized within the system—and since finitistic methods presumably were so formalizable in relatively weak number theoretic systems already, no finitistic consistency proofs could be given for such systems. This led Gentzen (1935, 1936), in particular, to rethink the role of consistency proofs and the character of finitistic reasoning; following him, work in proof theory has concentrated on, in a sense, *relative* consistency proof.

> From [Gödel's incompleteness theorems] it follows that the consistency of elementary number theory, for example, cannot be established by means of *part* of the methods of proof used in elementary number theory, nor indeed by *all* of these methods. To what extent, then, is a genuine reinterpretation [*Zurückführung*] still possible?
>
> It remains quite conceivable that the consistency of elementary number theory can in fact be verified by means of techniques which, in part, no longer belong to elementary number theory, but which can nevertheless be considered to be *more reliable* than the doubtful components of elementary number theory itself. (Gentzen 1936, 139)

Gentzen's proof uses transfinite induction on constructive ordinals $< \varepsilon_0$, and argues that these methods in fact are finitary, and hence "more reliable" than the infinitistic methods of elementary number theory.[79]

7. Itinerary VII. Intuitionism and Many-Valued Logics

7.1. Intuitionistic Logic

7.1.1. Brouwer's Philosophy of Mathematics

One of the most important positions in philosophy of mathematics of the 1920s was the intuitionism of Luitzen Egbertus Jan Brouwer (1881–1966).[80] Although our emphasis will be on the logical developments that emerged from Brouwer's intuitionism (as opposed to his philosophy of mathematics or the development of intuitionistic mathematics), it is essential to begin by saying something about his position in philosophy of mathematics. The essay "Intuitionism and Formalism" (1912b) contains many of the theses characteristic of Brouwer's approach. In it Brouwer discusses on what grounds one can base the conviction about the "unassailable exactness" of mathematical laws and distinguishes the position of the intuitionist from that of the formalist. The former, represented mainly by the school of French analysts (Baire, Borel, Lebesgue),[81] would posit the human mind as the source of the exactness; by contrast the formalist, by which Brouwer also means realists such as Cantor, would say that the exactness resides on paper. This rough and ready characterization of the situation, although objectionable, is very typical of Brouwer's style and perhaps contributed to the appeal of his radical proposal. Brouwer traces the origins of the intuitionist position back to Kant.[82] For Kant, time and space were the forms of our intuition, which shaped our perception of the world. He famously defended the idea that geometry and arithmetic are synthetic a priori. Brouwer only retains part of the Kantian intuitionism, in that he rejects the aprioricity of space but preserves that of time. The foundation of the Brouwerian account of mathematics is to be found in fact in the basal intuition of time:

> The neo-intuitionism considers the falling apart of moments of life into qualitatively different parts, to be reunited only while remaining separate by time, as the fundamental phenomenon of the human intellect, passing by abstracting from its emotional content into the fundamental phenomenon of mathematical thinking, the intuition of the bare two-oneness. (Brouwer 1912a, 80)

The rest of mathematics is, according to Brouwer, built out of this basal intuition. Together with the emphasis on the centrality of intuition, Brouwer denigrates the use of language in mathematical activity and reserves to it only an auxiliary role. Talking about the construction of (countable) sets he writes:

> And in the construction of these sets neither the ordinary language nor any symbolic language can have any other role than that of

> serving as a non-mathematical auxiliary, to assist the mathematical memory or to enable different individuals to build up the same set. (Brouwer 1912a, 81)

This is at the root of Brouwer's skeptical attitude toward a foundational role for formal work in logic and mathematics. Thus, the intuitionist position finds itself at odds with formalists, logicists, and Platonists, all guilty, according to Brouwer, of relying on "the presupposition of the existence of a world of mathematical objects, a world independent of the thinking individual, obeying the laws of classical logic and whose objects may possess to each other the 'relation of a set to its elements.' " For this reason Brouwer criticized, among other things, the foundation of set theory provided by Zermelo and eventually produced (starting in 1916–1917) his own intuitionist set theory. While in the realm of the finite there is agreement in the results (although not in the method) between intuitionists and formalists, the real differences emerge in the treatment of the infinite and the continuum. There is an important development in Brouwer's ideas here. Whereas in the 1912 essay he thought of real numbers as given by laws, later on (starting in 1917) he developed a very original conception of the continuum based on choice sequences.[83] This will lead him to the development of an alternative construction of mathematics, intuitionistic mathematics. Brouwer presented his new approach in two papers, titled "Foundation of set theory independent from the logical law of the excluded middle" (1918) and the companion paper "Intuitionist set theory" (1921). As already mentioned, the new approach to mathematics was characterized by the admission of "free choice" sequences, that is, procedures in which the subject is not limited by a law but can also proceed freely in the generation of arbitrary elements of the sequence. These sequences are seen as being generated in time and thus as "growing" or "becoming." This new conception of mathematics with the inclusion of free growth and indeterminacy goes hand in hand with one of the major claims of Brouwer's intuitionism, that is, the denial of the idea that mathematical entities and properties are always completely determined. The latter assumption is embodied, according to Brouwer, in the logical law of the excluded middle:

> The use of the principle of the excluded middle is *not permissible* as part of a mathematical proof. It has only scholastic and heuristic value, so that the theorems which in their proof cannot avoid the use of this principle lack all mathematical content. (Brouwer 1921, 23)

Thus, for the intuitionist the only acceptable mathematical entities and properties are those that are constructed in thought; mathematical objects and properties do not have an independent existence. As a consequence, this leads to an abandonment of the unrestricted validity of the principle of the excluded middle and thus to a restriction of the available means of proof in classical mathematics. However, intuitionistic mathematics is not simply a subset of classical mathematics obtained by eliminating the excluded middle but rather a different development, due to the fact that the admission of "incomplete entities" such as free-choice sequences leads to a new and original theory of the mathematical continuum. One of the new concepts introduced by Brouwer is that of Species. This is the intuitionist equivalent of "property" in the classical setting. The constructive interpretation of property is presented by Brouwer in opposition to the principle of comprehension formulated by Cantor and in a restricted form by Zermelo. While in the classical setting any well-formed formula partitions the universe into the set of objects that satisfy the formula and those that do not, the new interpretation of property, or "Species," is obtained by limiting its domain to the entities whose constructions have already been achieved. However, the Species does not partition the already constructed entities into those that satisfy the Species and those that do not. An entity will belong to the Species if one can successfully carry out a proof that the constructed entity does indeed have the property in question (in Brouwer's terminology, "fitting in"). An entity will not belong if one can successfully carry out a construction that will show that the assumption of its belonging to the Species generates a contradiction. However, it is clear that the alternatives to a demonstration of "fitting in" can be twofold: either the demonstration of the absurdity of a "fitting in" or the absence of a demonstration either of "fitting in" or of its absurdity. The consequences of this strict interpretation of negation are that Brouwer has to produce a reconstruction of mathematics in which the principles of double negation and the principle of the excluded middle do not hold. The intuitionistic reconstruction of mathematics cannot be given here;[84] our focus is on the logical aspects of the situation.

7.1.2. Brouwer on the Excluded Middle

From the beginning of his publishing career, Brouwer gave pride of place to the mental mathematical activity and downplayed the foundational rôle of language and logic in mathematics. The system of logical laws is then seen as a mere linguistic edifice that at best can only accompany the communication of successful mathematical constructions. In 1908, Brouwer expresses doubts as to the validity of the principle of the excluded middle, since he claims that it is not the case that for an arbitrary statement S, we either have a proof of S or we have a proof of the negation of S. Of course, this already presupposes a constructive interpretation of the logical connectives. Issues about the excluded middle became central once Brouwer developed his new conception of mathematics based on the admissibility of "becoming" entities (such as choice sequences) and constructive properties (Species) for which, as we have seen, there is more than one alternative to the successful "fitting" of a constructed object to the Species. After the publication of "The Foundations of set theory independent of the logical principle of the excluded middle," which develops parts of mathematics without appeal to the excluded middle, he wrote a number of essays in which he analyzed the logic of negation implicit in the new reconstruction of mathematics. In "On the significance of the excluded middle in mathematics, especially in function theory" (1923b), Brouwer proposes a positive account of how we illegitimately move from the excluded middle on finite domains to infinite domains:

> Within a specific finite "main system" we can always *test* (i.e., either prove or reduce to absurdity) properties of systems.... On the basis of the testability just mentioned, there hold, for properties conceived within a specific finite main system, the *principle of excluded middle*, that is, the principle that for every system every property is either correct or impossible, and in particular the *principle of the reciprocity of the complementary species*, that is, the principle that for every system the correctness of a property follows from the impossibility of the impossibility of this property. (Brouwer 1923b, 335)

However, the validity on finite domains was arbitrarily extended to mathematics in general:

> An a priori character was so consistently ascribed to the laws of theoretical logic that until recently these laws, including the principle of excluded middle, were applied without reservation even in the mathematics of infinite systems. (Brouwer 1923b, 336)

7.1.3. The Logic of Negation

In "Intuitionistic Splitting of the Fundamental Notions of Mathematics" (1923a), Brouwer for the first time engages in an analysis of the consequences of his viewpoint, in particular, his conception of negation as contradiction, for logic proper. He begins by pointing out that the

> the intuitionist conception of mathematics not only rejects the principle of the excluded middle altogether but also the special case, contained in the principle of reciprocity of complementary species, that is, the principle that for any mathematical system infers the correctness of a property from the absurdity of its absurdity." (1923a, 286)

The rejection of the principle of the excluded middle is then argued by means of an example, which is paradigmatic of what are now called (weak) Brouwerian counterexamples.[85] Let k_1 be the least n such that there is a sequence 0123456789 appearing between the nth place and the $(n+9)$th place of the decimal expansion of π, and let

$$c_n = \begin{cases} (-1/2)^{k_1} & \text{if } n \geq k_1 \\ (-1/2)^n & \text{otherwise.} \end{cases}$$

Then the sequence c_1, c_2, c_3, converges to a real number r. We define a real number g to be rational if one can calculate two rational integers p and q

whose ratio equals g. Then r cannot be rational and at the same time the rationality of r cannot be absurd. This is because if r were rational we could compute the two integers, thereby solving a problem for which no computation is known (i.e., finding k_1). On the other hand, it is not contradictory that it be rational, because in that case k_1 would not exist and thus r would be 0, that is, a rational after all. In fact, the problem giving rise to the weak counterexample used by Brouwer has now been solved. But one can use other unsolved problems to generate similar counterexamples.

The counterexample shows that intuitionistically we cannot assert (until the problem is solved) "r is either rational or irrational," something that is of course perfectly legitimate from the classical point of view. However, the argument goes through only if one grants that the property of being rational requires the explicit computation of the integers p and q, which is of course not required in the classical setting. The consequences for the logic of negation are stated by Brouwer in the following principles:

1. Intuitionistically, absurdity-of-absurdity follows from correctness but not vice versa.
2. However, intuitionistically, the absurdity-of-absurdity-of absurdity is equivalent with absurdity.

As a consequence of these principles, any finite sequence of absurdity predicates can be reduced either to an absurdity or to an absurdity-of-absurdity.

It should be pointed out in closing this section that the notion of absurdity obviously involves the notion of a "contradiction" or "the impossibility of fitting in" or an "incompatibility." All these notions presuppose negation or difference, but Brouwer never spells out with clarity how to avoid the potential circularity involved here, although he refers to a primitive intuition of difference (not definable in terms of classical negation) in 1975 (73).

7.1.4. Kolmogorov

Kolmogorov's contribution to the formalization of intuitionistic logic and its properties date from "On the principle of the excluded middle" (1925), which however was not known to many logicians until much later, undoubtedly due to the fact that it was written in Russian. Thus, the debate that we describe in section 7.1.5 on the nature of Brouwer's logic, does not refer to Kolmogorov. In the introduction to his article, Kolmogorov states his aim as follows:

> We shall prove that every conclusion obtained with the help of the principle of the excluded middle is correct provided every judgment that enters in its formulation is replaced by a judgement asserting its double negation. We call the double negation of a judgement its "pseudotruth." Thus, in the metamathematics of pseudotruth it is legitimate to apply the principle of the excluded middle. (Kolmogorov 1925, 416)

Kolmogorov's declared goal in the paper was to show why the illegitimate use of the excluded middle does not lead to contradiction. His results predate similar results by Gentzen (1933a) and Gödel (1933a), which are known as double negation interpretations or negative translations. Kolmogorov's points of departure are Brouwer's critique of classical logic and the formalization of classical logic given by Hilbert (1922c). He introduces two propositional calculi: $\mathfrak{B}$ and $\mathfrak{H}$.

Calculus $\mathfrak{B}$:

1. $A \rightarrow (B \rightarrow A)$,
2. $\{A \rightarrow (A \rightarrow B)\} \rightarrow (A \rightarrow B)$,
3. $\{A \rightarrow (B \rightarrow C)\} \rightarrow \{B \rightarrow (A \rightarrow C)\}$,
4. $(B \rightarrow C) \rightarrow \{(A \rightarrow B) \rightarrow (A \rightarrow C)\}$,
5. $(A \rightarrow B) \rightarrow \{(A \rightarrow \overline{B}) \rightarrow \overline{A}\}$.

Calculus $\mathfrak{H}$ is obtained by adding to $\mathfrak{B}$ the axiom

6. $\overline{\overline{A}} \rightarrow A$.

Rules of inference for both calculi are substitution and modus ponens.

It has been argued that Kolmogorov anticipated Heyting's formalization of intuitionistic propositional calculus (see section 7.1.6). This is almost true. The system $\mathfrak{B}$ (known after Johansson as the minimal calculus) differs from the negation-implication fragment of Heyting's axiomatization only by the absence of axiom

h. $A \supset (\overline{A} \supset B)$.

$\mathfrak{H}$ is equivalent to the formalization of classical propositional calculus given in Hilbert (1922c). We find in Kolmogorov also an attempt at a formalization of the intuitionistic predicate calculus, although he is not completely formal on this point. He regards as intuitive the rule "whenever a formula $\mathfrak{S}$ stands by itself [i.e., is proved], we can write the formula $(a)\mathfrak{S}$" (433; rule **P**) and states the following axioms:

I. $(a)\{A(a) \rightarrow B(a)\} \rightarrow \{(a)A(a) \rightarrow (a)B(a)\}$.

II. $(a)\{A \rightarrow B(a)\} \rightarrow \{A \rightarrow (a)B(a)\}$.

III. $(a)\{A(a) \rightarrow C\} \rightarrow \{(Ea)A(a) \rightarrow C\}$.

IV. $A(a) \rightarrow (Ea)\,A(a)$.

Adding to system $\mathfrak{B}$ the axioms I–IV and rule **P** would result in a complete system for intuitionistic predicate logic (Heyting 1930b) if axiom h and the following axiom,

g. $(a)A(a) \rightarrow A(a)$,

were also added. Kolmogorov considered axiom g to be true (see Wang 1967). He conjectured that $\mathfrak{B}$ is complete with respect to its intended interpretation ("the intuitively obvious" class of propositions), but he cautiously observed that "the question whether this axiom system is a complete axiom system for the intuitionistic general logic of judgments remains open" (422).

Whereas calculus $\mathfrak{B}$ corresponds, according to Kolmogorov, to the "general logic of judgments," calculus $\mathfrak{H}$ corresponds to the "special logic of judgments," since its range of application is narrower (it produces true propositions only when the propositional variables range over a narrower class of propositions). In section III of his paper, Kolmogorov individuates a class of judgments with the property that "the judgment itself follows [intuitively] from its double negation." Finitary judgments are of such type. Let $A^\bullet$, $B^\bullet$, $C^\bullet, \ldots$ denote judgments of the mentioned kind. Then $\overline{\overline{(A^\bullet \to B^\bullet)}} \to (A^\bullet \to B^\bullet)$ and $\overline{\overline{A^\bullet}} \to A^\bullet$ are provable in $\mathfrak{B}$. Moreover, for every negative formula $\overline{A}$, $\mathfrak{B}$ proves $\overline{\overline{\overline{A}}} \to \overline{A}$. It is also shown that substitution for propositional variables, modus ponens, and the axioms of $\mathfrak{H}$ are all valid for this class of propositions. This shows that the system $\mathfrak{H}$ is intuitionistically correct if we restrict it to the class of judgments of the form $A^\bullet$. Thus, the domain for which the calculus $\mathfrak{H}$ is valid is the class of propositions that follow (intuitively) from their double negation, and this includes finitary statements and all negative propositions. This amounts to showing that all of propositional logic is included in intuitionistic propositional logic, if the domain of propositions is restricted to propositions of the form $A^\bullet$. In section IV, Kolmogorov introduces a translation from formulas of classical mathematics to formulas of intuitionistic mathematics:

> We shall construct alongside of ordinary mathematics, a "pseudo-mathematics" that will be such that to every formula of the first there corresponds a formula of the second and, moreover, that every formula of pseudomathematics is a formula of type $A^\bullet$. (Kolmogorov 1925, 418)

The translation is defined as follows: If A is atomic, then $A^* = \overline{\overline{A}}$; $\overline{A}^* = \overline{\overline{\overline{A^*}}}$; and $(A \to B)^* = \overline{\overline{A^* \to B^*}}$. Thus, if $A_1, \ldots, A_k$ are axioms of classical mathematics (comprising the logical axioms), then we have $A_1, \ldots, A_k$ proves A in $\mathfrak{H}$ iff A_1^*, $\ldots$, A_k^* proves A^* in $\mathfrak{B}$. The theorem is proved by showing that applications of substitution and modus ponens remain derivable in $\mathfrak{B}$ under the $*$-translation, using the results about double negations previously established. Moreover, the $*$-translations of the logical axioms are derivable in $\mathfrak{B}$.

Kolmogorov did not extend the result to predicate logic but the extension is straightforward. It should be pointed out that he asserts (IV, §5–6) that every axiom A of classical mathematics is such that A^* is intuitionistically true. But this would imply that all of classical mathematics is intuitionistically consistent, a result which is not established, for analysis and set theory, even to this day. However, as Wang remarks, "it seems not unreasonable to assert that Kolmogorov did foresee that the system of classical number theory is translatable into intuitionistic number theory and therefore is intuitionistically consistent" (Wang 1967, 415). We return to these results after describing the discussion on Brouwer's logic in the West.

7.1.5. The Debate on Intuitionist Logic

In 1926, Wavre published an article contrasting "logique formelle" (classical) and "logique empiriste" (intuitionist). This was, apart from Kolmogorov (1925), the first attempt to discuss systematically the features of "Brouwer's logic." Whereas classical logic is a logic of truth and falsity, "empirical" logic is a logic of truth and absurdity, where true means "effectively demonstrable" and absurd "effectively reducible to a contradiction." Wavre begins by listing similar principles between the two logics:

1. $((A \supset B) \mathbin{\&} (B \supset C)) \supset (A \supset C)$.
2. From A and $A \supset B$, one can infer B.
3. $\neg(A \mathbin{\&} \neg A)$.
4. $(A \supset B) \supset (\neg B \supset \neg A)$.

Among the different principles Wavre mentions the excluded middle and double negation. He then shows that $\neg A$ is equivalent, in empirical logic, to $\neg\neg\neg A$. Moreover he observed that in empirical logic the converse of (4) does not hold, unless B is a negative proposition. Much of Wavre's article only restated observations that were, implicitly or explicitly, contained in Brouwer (1923b). However, it had the merit of opening a debate in the *Revue de Metaphysique et de Morale* on the nature of intuitionistic logic which saw contributions by Wavre, Levy, and Borel. However, this debate did not directly touch on the principles of intuitionistic logic.[86] By contrast, Barzin and Errera (1927) claimed that Brouwerian logic was inconsistent, thereby sparking a long debate on the possibility of an intuitionistic logic, which saw contributions by Church, Levy, Glivenko, Khintchine, and others. Barzin and Errera incorrectly interpreted Brouwer's talk of undecided propositions (i.e., those for which there is neither an effective proof of their validity nor an effective proof of their absurdity) as claiming that there are propositions which are neither true nor false. These propositions are "tierce." Their aim was then to show that the admission of a "tierce" led to formal contradictions. They interpreted these "third" propositions not as a state of objective ignorance but rather as an "objective logical fact." They denoted "p is tierce" by p'. With this notation in place, they stated a principle of "quartum non datur": $p \vee \neg p \vee p'$ and claimed that Brouwer must accept it, if "tierce" is defined as being "neither true nor false." Finally, the equivalent of the principle of noncontradiction, which they claimed Brouwer must admit, is that no proposition can be true and false, or true and tierce, or false and tierce. Under these assumptions they claimed to show that one could prove the collapse of the truth values, that is, that in the calculus one could prove that every proposition that is true is also tierce, and every proposition that is tierce is also false. The proof is, however, inconclusive. First of all, there is a constant confusion between the object level and the metalevel of analysis; moreover, the proof makes use of principles that are classically but not intuitionistically valid.

Of the many replies to Barzin and Errera (1927), we discuss only Church's (1928).[87] In "On the law of the excluded middle" Church discussed, and rejected, the claims by Barzin and Errera by making essentially three points. First, he points out that the easiest alternative to a system that includes the law of the excluded middle is a system in which the excluded middle is not assumed "without assertion of any contrary principle." Thus, because this is a subsystem of the original one, no contradictions can be derived that could not be derived in the original system. To generate a contradiction we must admit a new principle that is not consistent with the law of the excluded middle. Second, one can drop the principle of the excluded middle and "introduce the middle ground between true and false as an undefined term" in which case it might be that "making the appropriate set of assumptions about the existence and properties of tierce propositions, we can produce a system of logic which is consistent with itself but which becomes inconsistent if the law of the excluded middle be added."[88] This possibility had already been proven by Łukasiewicz in developing many-valued logics (see later discussion), but Church does not mention Łukasiewicz. Third, the argument by Barzin and Errera fails because they introduce the tierce propositions by defining them as being neither true nor false, and this leads to an inconsistency. The argument by Barzin and Errera works only if one admits the faulty definition of a tierce (rather than leaving the notion undefined) and the principle of the excluded fourth, which again is defended using the faulty definition. Finally, Church argued that Barzin and Errera's argument is ineffective against those who simply drop the principle of the excluded middle, as "the insistence that one who refuses to accept a proposition must deny it can be justified only by an appeal to the law of the excluded middle."

7.1.6. The Formalization and Interpretation of Intuitionistic Logic

Glivenko (1928) contributed an article on intuitionistic logic in which he showed that Brouwerian logic could not admit a tierce. But of great technical interest is Glivenko (1929), which contains the following two theorems:

1. If a certain expression in the logic of propositions is provable in classical logic, it is the falsity of the falsity of this expression that is provable in Brouwerian logic.
2. If the falsity of a certain expression in the logic of propositions is provable in classical logic, that same falsity is provable in Brouwerian logic (Glivenko 1929, 301)

Although Glivenko's results do not yet amount to a translation of classical logic into intuitionistic logic, they certainly paved the way for the later results by Gödel and Gentzen (see Troelstra 1990; van Atten 2005). By far the most important contribution in this period is the work of Heyting on the formalization of intuitionistic logic. Heyting's contributions were motivated by a prize question published in 1927 by the Dutch Mathematical Society on the formalization of the principles of intuitionism. Heyting was awarded the prize in 1928, but his result appeared in print only in 1930. Heyting (1930a) contains a formalization of the laws of intuitionistic propositional logic; (1930b) moves on to intuitionistic predicate logic and arithmetic; and finally, (1930c) investigates intuitionistic principles in analysis.

Heyting distilled the principles of intuitionistic logic by going through the list of axioms in *Principia Mathematica* and retaining only those that admitted of an intuitionist justification (letter to Becker, September 23, 1933; see Troelstra 1990; van Atten 2005). The axioms for the propositional part were the following.

1. $A \supset (A \wedge A)$.
2. $A \wedge B \supset B \wedge A$.
3. $(A \supset B) \supset ((A \wedge C) \supset (B \wedge C))$.
4. $((A \supset B) \wedge (B \supset C)) \supset (A \supset C)$.
5. $B \supset (A \supset B)$.
6. $(A \wedge (A \supset B)) \supset B$.
7. $A \supset A \vee B$.
8. $A \vee B \supset B \vee A$.
9. $((A \supset C) \wedge (B \supset C)) \supset (A \vee B \supset C)$.
10. $\neg A \supset (A \supset B)$.
11. $((A \supset B) \supset (A \supset \neg B)) \supset \neg A$.

In the appendix, Heyting proves that all the axioms are independent, exploiting a technique used by Bernays for proving the independence of the propositional axioms of *Principia* (see Section 5.3). Heyting (1930b) also gives an axiomatization for principles acceptable in intuitionistic first-order logic. He (1930a) only states the admissible principles and proved theorems from them, but he was not explicit on the meaning of the logical connectives in intuitionistic logic. However, he (1930d) did provide an interpretation for intuitionistic negation and disjunction. The interpretation depends on interpreting propositions as problems or expectations:

> A proposition p like, for example, "Euler's constant is rational" expresses a problem, or better yet, a certain expectation (that of finding two integers a and b such that $C = a/b$), which can be fulfilled or disappointed. (Heyting 1930d, 307)

This interpretation is influenced by Becker's treatment of intuitionism in *Mathematische Existenz* (1927) where, appealing to distinctions found in Husserl's *Logical Investigations*, Becker distinguishes between the fulfillment of an intention (say a proof of "a is B"), the frustration of an intention (a proof of "a is not B") and the nonfulfillment of an intention (i.e., the lack of a fulfillment). Indeed, Heyting (1931) explicitly refers to the phenomenological interpretation and claims that "the affirmation of a proposition is the fulfillment of an intention" (1931, 59). He mentions Becker in connection with the interpretation of intuitionistic negation:

> A logical function is a process for forming another proposition from a given proposition. Negation is such a function. Becker, following Husserl, has described its meaning very clearly. For him negation is something thoroughly positive, viz., the intention of a contradiction contained in the original intention. The proposition "C is not rational" therefore, signifies the expectation that one can derive a contradiction from the assumption that C is rational. (Heyting 1931, 59)

Disjunction is interpreted as the expectation of a mathematical construction that will prove one of the two disjuncts. In Heyting (1934), it is specified that the mathematical construction fulfilling a certain expectation is a proof. Under this interpretation, $A \supset B$ signifies "the intention of a construction that leads from each proof of A to a proof of B." This interpretation of the intuitionistic connectives is now known as the Brouwer–Heyting–Kolmogorov interpretation. The presence of Kolmogorov stems from his interpretation of the intuitionistic calculus as a calculus of problems (1932). In this interpretation, for instance, $\neg A$ is interpreted as the problem "to obtain a contradiction, provided the solution of A is given." Although the two interpretations are distinct, they were later on treated as essentially the same, and Heyting (1934, 14) speaks of Kolmogorov's interpretation as being closely related to his.[89]

7.1.7. Gödel's Contributions to the Metatheory of Intuitionistic Logic

Glivenko's work had shown that classical propositional logic could be interpreted as a subsystem of intuitionistic logic, and thus be intuitionistically consistent. We have also seen that Kolmogorov (1925) implicitly claimed that classical mathematics is intuitionistically consistent. A more modest, but extremely important, version of this unsupported general claim was proved by Gödel and Gentzen in 1933. Gödel states:

> The goal of the present investigation is to show that something similar [to the translation of classical logic into intuitionistic logic] holds also *for all of arithmetic and number theory*, delimited in scope by, say, Herbrand's axioms. Here, too, we can give an interpretation of the classical notions in terms of the intuitionistic ones *so that all*

> *propositions provable from the classical axioms hold for intuitionism as well.* (Gödel 1933c, 287–289)[90]

Gödel distinguished the classical connectives from the intuitionistic connectives: $\neg$, $\supset$, $\vee$, $\wedge$ are the intuitionistic connectives; the corresponding classical connectives are $\sim$, $\rightarrow$, $\vee$, $\cdot$. Gödel's translation $'$ from classical propositional logic into intuitionistic logic is defined as follows: $p' = p$, if p is atomic; let $(\sim p)' = \neg p'$, $(p \cdot q)' = p' \wedge q'$; $(p \vee q)' = \neg(\neg p' \wedge \neg q')$; $(p \rightarrow q)' = \neg(p' \wedge \neg q')$.

He then shows that classical propositional logic proves a sentence A if and only if intuitionistic propositional logic proves the translation A'. The result is then extended to first-order arithmetic by first extending the translation to cover the universal quantifier so that $(\forall x\, P)' = \forall x\, P'$. Letting H' stand for intuitionistic first-order arithmetic and Z for first-order arithmetic (in Herbrand's formulation), then Gödel showed that a sentence A is provable in Z iff its translation A' is provable in H'.

From the philosophical point of view, the importance of the result consists in showing that under a somewhat deviant interpretation, classical arithmetic is already contained in intuitionistic arithmetic. Therefore, this amounts to an intuitionistic proof of the consistency of classical arithmetic. This result once and for all brought clarity into a systematic confusion between finitism and intuitionism, which had characterized the literature on the foundation of mathematics in the 1920s.[91] Gödel's result makes clear that intuitionistic arithmetic is much more powerful than finitistic arithmetic.

Two more results by Gödel on the metatheory of intuitionistic logic have to be mentioned. The first (1933a) consists in an interpretation of intuitionistic propositional logic into a system of classical propositional logic extended by an operator B ("provable," from the German *beweisbar*). It is essential that provability here be taken to mean "provability in general" rather than provability in a specified system. The logic of the system B turns out to coincide with the modal propositional logic S4. The system S4 is characterized by the following axioms:

1. $Bp \rightarrow p$,
2. $Bp \rightarrow (B(p \rightarrow q) \rightarrow Bq)$,
3. $Bp \rightarrow BBp$.

The translation † works as follows: Atomic sentences are sent to atomic sentences; $(\neg p)^\dagger = \sim Bp^\dagger$; $(p \supset q)^\dagger = Bp^\dagger \rightarrow Bq^\dagger$; $(p \vee q)^\dagger = Bp^\dagger \vee Bq^\dagger$; $(p \wedge q)^\dagger = p^\dagger \cdot q^\dagger$. Gödel showed that if A is provable in intuitionistic propositional logic, then $A^\dagger$ is provable in S4. This result was important in that it showed the connections between modal logic and intuitionistic logic and paved the way for the development of Kripke's semantics for intuitionistic logic, once the semantics for modal logic had been worked out.

One final result by Gödel concerns intuitionistic logic and many-valued logic. Gödel (1932b) proved that intuitionistic propositional logic cannot be identified with a system of many-valued logic with finitely many truth values. Moreover, he showed that there is an infinite hierarchy of finite-valued logics between intuitionistic and classical propositional logic.[92]

7.2. Many-Valued Logics

The systematic investigation of systems of many-valued logics goes back to Jan Łukasiewicz.[93] Łukasiewicz arrived at many-valued logics as a possible way out of a number of philosophical puzzles he had been worrying about. The first concerns the very foundation of classical logic, that is, the principle that every proposition p is either true or false. This he called the law of bivalence (1930, 53). The principle had already been the subject of debate in ancient times, and Aristotle himself expressed doubts as to its applicability for propositions concerning future contingents ("there will be a sea battle tomorrow"). The wider philosophical underpinnings of such debates had to do with issue of determinism and indeterminism, which Łukasiewicz explored at length (see, for instance, Łukasiewicz 1922). In all such issues, the notion of possibility and necessity are obviously central. Indeed, in his presentation of many-valued logic, Łukasiewicz motivates the system by a reflection on modal operators, such as "it is possible that p." The first presentation of the results goes back to two lectures given in 1920: "On the concept of possibility" (1920b) and "On three valued-logic" (1920a). Let us follow these lectures. In the first lecture, Łukasiewicz considers the relationship between the following sentences:

i. S is P.

ii. S is not P.

iii. S can be P.

iv. S cannot be P.

v. S can be non-P.

vi. S cannot be non-P (i.e., S must be P).

He distinguishes three positions that can be held with respect to the logical relationship between the above sentences:

a. If S must be P (vi), then S is P (i).

b. If S cannot be P (iv), then S is not P (ii).

When no further relationships hold between (i)–(vi), this corresponds to the point of view of traditional logic. The second position, corresponding to ontological determinism, consists of theses (a) and (b) plus the implications

c. If S is P (i), then S must be P (vi).

d. If S is non-P (ii), then S cannot be P (iv).

Finally, the third position, corresponding to ontological indeterminism, consists of (a), (b), and the implications

e. If S can be P (iii), then S can be non-P (v).

f. If S can be non-P (v), then S can be P (iii).

All these theses have, according to Łukasiewicz, a certain intuitive obviousness. However, he shows that if one reasons within the context of classical logic, there is no way to consistently assign truth values 0 and 1 to (i)–(vi) so that all of (a)–(f) will get value 1. However, this becomes possible if one introduces a new truth value, 2, which stands for "possibility." This gives rise to the need for the study of "three-valued logic."

In the second lecture, Łukasiewicz defines three-valued logic as a system of non-Aristotelian logic and defines the truth tables for equivalence and implication based on three values in such a way that the tables coincide with classical logic when the values are 1 and 0 but satisfy the following laws when the value 2 occurs. For the biconditional, one stipulates that the values for 02, 20, 21, and 12 is going to be 2; for the material conditional, the value is 1 for 02, 21, and 22 and it is 2 for 20 and 12. From the general analysis, it is also clear that for negation the following holds: If p is assigned value 2 then $\sim p$ is also 2.

While all tautologies of three valued-logic are tautologies of classical propositional (two-valued) logic, the converse is not true. For instance, $p \vee \sim p$ is not a tautology in three-valued logic, because if p is assigned the value 2, the value of $p \vee \sim p$ is also 2.

In Post (1921) we also find a study of many-valued logics. However, Post studies these systems purely formally, without attempting to give them an intuitive interpretation. It is perhaps on account of this fact that he was the first to develop tables for negation known as "cyclic commutation" tables. In the case of Łukasiewicz's system, negation is always defined by a "mirror" truth table, that is, the value of negation is that of its opposite in the order of truth (the value of $\sim p$ is 1 minus the value of p). In the case of Post, the truth table for negation is defined by permuting the truth values cyclically. Here is a comparison of the tables for the two types of negations in three-valued logic:

Łukasiewicz		Post	
p	$\sim p$	p	$\sim p$
0	1	0	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
1	0	1	0

Post was motivated by issues of functional completeness and in fact one of the results (1921) is that the system of m-valued logic he introduces, with a "cyclic commutation" table for negation, and a disjunction table obtained by giving the disjunction the maximum of the truth values of the disjuncts, is truth-functionally complete. The table for negation, with values 1 to m, is as follows:

p	1	2	...	m
$\sim p$	2	3	...	1

Łukasiewicz generalized his work from three-valued logics to many-valued logics (1922). At first he looked at logics with n truth values and later he considered logics with $\aleph_0$ values. All these systems can be expressed as follows. Let n be a natural number or $\aleph_0$. Assume that p and q range over a set of n numbers from the interval $[0, 1]$. As usual at the time, let us standardize the values to be $k/(n-1)$ for $0 \leq k \leq n-1$ when n is finite and k/l $(0 \leq k \leq l)$ when n is $\aleph_0$. Define $p \rightarrow q$ to have value 1 whenever $p \leq q$ and value $1-p+q$ whenever $p > q$. Let $\sim p$ have value $1-p$. If we select only 0 and 1 we are back in the classical two-valued logic. If we add to 0 and 1 the value $\frac{1}{2}$ we get three-valued logic. In similar fashion, one can create systems of n-valued logic. If p and q range over a countable set of values one obtains an infinite-valued propositional calculus. Many Polish logicians investigated the relationships between systems of many-valued logic (see Woleński 1989). One of the first problems was to study how the sequence of logics L_n $(n > 1)$ behaves. It was soon shown that all tautologies of L_n are also tautologies of L_2, but the converse does not hold. While $L_{\aleph_0}$ turns out to be contained in all finite L_n, the relationship between any two finite L_m and L_n is more complicated. Łukasiewicz and Tarski (1930) attribute to Lindenbaum the following result (theorem 19): For $2 \leq m$ and $2 \leq n$ (m, n finite) we have: L_m is included in L_n iff $n-1$ divides $m-1$. Among the early results concerning the axiomatization of many-valued logics one should mention Wajsberg (1931), which contains a complete and independent axiomatization of three-valued logic. However, the system is not truth-functionally complete. Słupecki (1936) proved that if one adds to the connectives $\supset$ and $\sim$ in three-valued logic, the operator T such that Tp is always $\frac{1}{2}$ (for $p = 1$, 0, or $\frac{1}{2}$), then the system is truth-functionally complete. To provide an axiomatization one needs to add some axioms for T to the axioms given by Wajsberg. Thus, the axiomatization provided by Słupecki is given by the following six axioms:

1. $p \supset (q \supset p)$.
2. $(p \supset q) \supset ((q \supset r) \supset (p \supset r))$.
3. $(\sim p \supset \sim q) \supset (q \supset p)$.
4. $((p \supset \sim p) \supset p) \supset p$.
5. $Tp \supset \sim Tp$.
6. $\sim Tp \supset Tp$.[94]

The axiomatizability of $L_{\aleph_0}$ was conjectured by Łukasiewicz in 1930, who put forth the (correct) candidate axioms, but a proof of the result was only given by Rose and Rosser (1958).

Let us conclude this exposition on many-valued logic in the twenties and the early thirties by mentioning some relevant work on the connection between intuitionistic logic and many-valued logic. We have seen that Gödel in 1932 showed that intuitionistic logic did not coincide with any finite many-valued logic. More precisely, he showed that no finitely valued matrix characterizes intuitionistic logic. Theorem I of Gödel (1932b) reads:

> There is no realization with finitely many elements (truth values) for which the formulas provable in H [intuitionistic propositional logic], and only those, are satisfied (i.e., yield designated truth values for an arbitrary assignment). (Gödel 1932b, 225)

In the process he identified an infinite class of many-valued logic, now known as Gödel logics. This is captured in the second theorem of the paper:

> Infinitely many systems lie between H and the system A of the ordinary propositional calculus, that is, there is a monotonically decreasing sequence of systems all of which include H as a subset and are included in A as subsets. (Gödel 1932b, 225)

The previous result gave the first examples of logics that are now studied under the name of intermediate logics. One important result that should be mentioned in this connection was obtained by Jaśkowski (1936), who provided an infinite truth value matrix appropriate for intuitionistic logic.

8. Itinerary VIII. Semantics and Model-Theoretic Notions

8.1. Background

During the previous itineraries, we have come across the implicit and explicit use of semantic notions (interpretation, satisfaction, validity, truth, etc.). In this section we retrace, in broad strokes, the main contexts in which these notions occurred in the first two decades of the twentieth century. This will provide the background for an understanding of the gradual emergence of the formal discipline of semantics (as part of metamathematics) and, much later, model theory.

The first context we have encountered in which semantical notions make their appearance is that of axiomatics (see itinerary I). A central notion in the analysis of axiomatic theories is that of "interpretation," which of course has its roots in nineteenth-century work on geometry and abstract algebra (see Guillaume 1994; Webb 1995). The development of analysis, algebra, and geometry in the nineteenth century had led to the idea of an uninterpreted formal axiomatic system. We have seen that Pieri (1901) emphasized that the primitive notions of any deductive system "must be capable of arbitrary interpretations," with the only restriction that the primitive sentences are satisfied by the particular interpretation. The axioms are verified, or made true, by particular interpretations. Interpretations are essential for proofs of consistency and independence of the axioms. However, as we said, the semantical notions involved (satisfaction, truth in a system) are used informally. Moreover, all these developments took place without a formal specification of the background logic. With minor modifications from case to case, these remarks apply to Peano's school, Hilbert, and the American postulate theorists.

8.1.1. The Algebra of Logic Tradition

A second tradition in which semantic notions appear quite frequently is that of the algebra of logic. To this tradition we owe what is considered the very first important result in model theory (as we understand it today, i.e., a formal study of the relationship between a language and its interpretations). This is the Löwenheim–Skolem theorem. As stated by Skolem:

> In volume 76 of *Mathematische Annalen*, Löwenheim proved an interesting and very remarkable theorem on what are called "first-order expressions" [*Zählausdrücke*]. The theorem states that every first-order expression is either contradictory or already satisfiable in a denumerably infinite domain. (Skolem 1920, 254)

As we have already seen in itinerary IV, the basic problem is the satisfaction of (first-order) equations on certain domains. Domain and satisfaction are the key terminological concepts used by Löwenheim and Skolem (who do not talk of interpretations). However, all these semantical notions are used informally.

It can safely be asserted that the clarification of semantic notions was not seen as a goal for mathematical axiomatics. In 1918, Weyl gestures toward an attempt at clarifying the meaning of "true judgment," but he does so by delegating the problem to philosophy (Fichte, Husserl). An exception here is Ajdukiewicz (1921), who was only accessible to those who read Polish. Ajdukiewicz stressed the issues related to a correct interpretation of the notions of satisfaction and truth in the axiomatic context. This was to leave a mark on Tarski, who was thoroughly familiar with this text (see section 8.7).

8.1.2. Terminological Variations (Systems of Objects, Models, and Structures)

Throughout the 1910s, the terminology for interpretations of axiomatic systems remains rather stable. Interpretations are given by systems of objects with certain relationships defined on them. Bôcher (1904) suggests the expression "mathematical system" to "designate a class of objects associated with a class of relations between these objects" (128). Nowadays, however, we speak just as commonly of models or structures. When did the terminology become common currency in axiomatics?

"Model," as an alternative terminology for interpretation, makes its appearance in the mathematical foundational literature in von Neumann (1925), where he talks of models of set theory. However, the new terminology owes its influence and success to Weyl's "Philosophy of Mathematics and Natural Science" (1927). In introducing techniques for proving independence, Weyl describes the techniques of "construction of a model [*Modell*]" (18) and described both Klein's construction of a Euclidean model for non-Euclidean geometry and the construction of arithmetical models for Euclidean geometry (or subsystems thereof) given by Hilbert.[95] Once introduced in the axiomatical literature by Weyl, the word "model" finds a favorable reception. It occurs in Carnap (1927, 2000 [1927–1929], 1930), Kaufmann (1930), and in articles by Gödel (1930b), Zermelo (1929, 1930), and Tarski (1936a). The usage is, however, not universal. The word "model" is not used in Hilbert and Ackermann (1928) (but it is found in Bernays 1930). Fraenkel (1928) speaks about realizations or models (353), as does Tarski (1936a). The latter do not follow Carnap in drawing a distinction between realizations (concrete, spatiotemporal interpretations) and models (abstract interpretations). "Realization" is also used by Baldus (1924) and Gödel (1929).

As for "structure," it is not used in the twenties as an equivalent of "mathematical system." Rather, mathematical systems have structure. In *Principia Mathematica* (Whitehead and Russell 1912, part iv, *150ff.) and then in Russell (1919, ch. 6) we find the notion of two relations "having the same structure."[96] In Weyl (1927, 21), two isomorphic systems of objects are said to have the same structure. This process will eventually lead to the idea that a "structure" is what is captured by an axiom system: "An axiom system is said to be monomorphic when exactly one structure belongs to it [up to isomorphism]" (Carnap 2000 [1927–1929], 127; see also Bernays 1930).

Here it should be pointed out that the use of the word "structure" in the algebraic literature was not yet widespread, although the structural approach was. It seems that "structure" was introduced in the algebraic literature in the early 1930s by Øystein Ore to denote what we nowadays call a lattice (see Vercelloni 1988; Corry 2004).

8.1.3. Interpretations for Propositional Logic

A major step forward in the development of semantics is the clarification of the distinction between syntactical and semantic notions made by Bernays in *Habilitationsschrift* of 1918 (see itinerary V). We have seen that Bernays clearly distinguished between the syntax of the propositional calculus and its interpretations, a distinction that was not always clear in previous writers. This allowed him to properly address the problem of completeness for the propositional calculus. Bernays distinguished between provable formulas (obtainable from the axioms by means of the rules of inference) from the valid formulas (which yield true propositions for any substitution of propositions for the variables) and stated the completeness problem as follows: "Every provable formula is a valid formula and conversely." It would be hard to overestimate the importance of this result, which formally shows the equivalence of a syntactic notion (provable formula) with a semantic one (valid formula) (in section 8.4

we will look at the emergence of the corresponding notions for first-order logic). Post (1921) also made a clear distinction between the formal system of propositional logic and the semantic interpretation in terms of truth-table methods, and he also established the completeness of the propositional calculus (see section 8.3).

In this way, logic becomes an object of axiomatic investigation for which one can pose all the problems that had traditionally been raised about axiomatic systems. To get a handle on the problems, researchers first focused on the axiomatic systems for the propositional calculus and then moved on to wider systems (such as the "restricted functional calculus," that is, first-order predicate logic). Here we focus on the metatheoretical study of systems of axiomatic logic rather than the developments of mathematical axiomatic theories (models of set theory, arithmetic, geometry, various algebraic structures, etc.).

8.2. Consistency and Independence for Propositional Logic

We have seen that the use of interpretations to provide independence results was exploited already in the nineteenth century in several areas of mathematics. Hilbert, Peano and his students, and also the American postulate theorists put great value in showing the independence of the axioms for any proposed axiomatic system. Most of these applications concern specific mathematical theories. Applications to logic appear first in the tradition of the algebra of logic. For instance, in "Sets of independent postulates for the algebra of logic" (1904), Huntington studied the "algebra of symbolic logic" as an independent calculus, as a purely deductive theory. The object of study is given by a set K satisfying the axioms of what we would now call a Boolean algebra. Huntington provides three different axiomatizations of the "algebra of logic," of which we present the first, built after Whitehead's presentation in *Universal Algebra* (1898). Possible interpretations for the system are the algebra of classes and the algebra of propositions. Huntington claims originality in the extensive investigation of the independence of the axioms. The first axiomatization states the properties of a class K of objects on which are defined two operations, $\oplus$ and $\otimes$, satisfying the following axioms:

- Ia. $a \oplus b$ is in the class whenever a and b are in the class;
- Ib. $a \otimes b$ is in the class whenever a and b are in the class;
- IIa. There is an element $\bigwedge$ such that $a \oplus \bigwedge = a$, for every element a;
- IIb. There is an element $\bigvee$ such that $a \otimes \bigvee = a$, for every element a;
- IIIa. $a \oplus b = b \oplus a$ whenever a, b, $a \oplus b$, and $b \oplus a$ are in the class;
- IIIb. $a \otimes b = b \otimes a$ whenever a, b, $a \otimes b$, and $b \otimes a$ are in the class;
- IVa. $a \oplus (b \otimes c) = (a \oplus b) \otimes (a \oplus c)$ whenever a, b, c, $a \oplus b$, $a \oplus c$, $b \otimes c$, $a \oplus (b \otimes c)$, and $(a \oplus b) \otimes (a \oplus c)$ are in the class;

IVb. $a \otimes (b \oplus c) = (a \otimes b) \oplus (a \otimes c)$ whenever a, b, c, $a \otimes b$, $a \otimes c$, $b \oplus c$, $a \otimes (b \oplus c)$, and $(a \otimes b) \oplus (a \otimes c)$ are in the class;

V. If the elements $\bigwedge$ and $\bigvee$ in postulates IIa and IIb exist and are unique, then for every element a there is an element $\overline{a}$ such that $a \oplus \overline{a} = \bigvee$ and $a \otimes \overline{a} = \bigwedge$;

VI. There are at least two elements, x and y, in the class such that $x \neq y$.

The consistency of the set of axioms is given by a finite table consisting of two objects, 0 and 1, satisfying the following.

$\oplus$	0	1
0	0	1
1	1	1

$\otimes$	0	1
0	0	0
1	0	1

The reader will notice that if we interpret $\oplus$ as conjunction of propositions and $\otimes$ as disjunction, we can read the table as the truth table for conjunction and disjunction of propositions (letting 0 stand for true and 1 for false). Similar tables are used by Huntington to prove the independence of each of the axioms from the remaining ones. In every case one provides a class and tables for $\oplus$ and $\otimes$ that verify all of the axioms but the one to be shown independent. For instance, IIIa can be shown to be independent by taking two objects 0 and 1 with the following tables.

$\oplus$	0	1
0	0	0
1	1	1

$\otimes$	0	1
0	0	0
1	0	1

All the axioms are satisfied, but $a \oplus b = b \oplus a$ fails by letting $a = 0$ and $b = 1$. Similarly for $a \otimes b$.

These techniques were not new and were already used in connection with the algebra of propositions by Peirce and Schröder. An application of this algebraic approach to the propositional calculus of *Principia Mathematica* was given by Sheffer (1913). Sheffer showed that one could study an algebra on a domain K with a binary K-rule of combination $\mid$ satisfying the following axioms.

1. There are at least two distinct elements of K.
2. $a \mid b$ is in K whenever a and b are in K.
3. $(a \mid a) \mid (a \mid a) = a$ whenever a is an element of K and all the indicated combinations of a are in K.
4. $a \mid (b \mid (b \mid b)) = a \mid a$ whenever a and b are elements of K and all the indicated combinations of a and b are in K.
5. $(a \mid (b \mid c)) \mid (a \mid (b \mid c)) = ((b \mid b) \mid a) \mid ((c \mid c) \mid a)$ whenever a, b, and c are elements of K and all the indicated combinations of a, b, and c are in K.

Sheffer showed that this set of postulates implies Huntington's set by letting $\overline{a} = a \mid a$; $a \oplus b = (a \mid b) \mid (a \mid b)$ and $a \otimes b = (a \mid a) \mid (b \mid b)$. Conversely, by defining $a \mid b$ as $\overline{a} \otimes \overline{b}$, Huntington's set implies Sheffer's set of axioms. The application to *Principia* is now immediate. One can substitute a single connective $p \mid q$ defined as $\sim(p \vee q)$.

This work leads us to Bernays's (1918, 1926) studies of the independence of the axioms of the propositional fragment of *Principia*. Actually, Bernays was unaware of Sheffer's work until Russell mentioned it to him in 1920 (see Mancosu 2003). Bernays's (1926) formulation of the propositional logic ("theory of deduction") of *Principia* is given by

Taut. $\vdash :p \vee p . \supset .p$

Add. $\vdash :q . \supset .p \vee q$

Perm. $\vdash :p \vee q . \supset .q \vee p$

Assoc. $\vdash :p \vee (q \vee r) . \supset .q \vee (p \vee r)$

Sum. $\vdash :.q \supset r . \supset :p \vee q . \supset .p \vee r$

One also has rules of substitution and modus ponens.

The proof of independence of the axioms of the propositional calculus of *Principia*, with the exclusion of associativity, shown by Bernays to be derivable from the others, was given by appropriate interpretations in the style of the independence proofs we have looked at in the work of Huntington. However, one also has to show that the inference rules, and in particular modus ponens, preserve the right value. The technique is that of exhibiting "finite systems" consisting in the assignment of three or four finite values to the variables. One (or several) of these values are then singled out as distinguished value(s).

The proof of consistency of the calculus is given by letting propositions range over $\{0, 1\}$ and interpreting $\sim p$ as the numerical operation $1 - p$ and $p \vee q$ (disjunction) as the numerical operation $p \times q$. It is easy to check that the axioms always have value 0 and that substitution and modus ponens lead from formulas with value 0 to other formulas with value 0. This shows the calculus to be consistent, for were a contradiction provable, say $(p \,\&\, \neg p)$, then it would take the value 1.

The technique of proving independence of the axioms is similar (*Methode der Aufweisung*). Consider the axiom Taut. We give the following table with three values a, b, c with a distinguished value, say a.

$\vee$	a	b	c	$\sim$
a	a	a	a	b
b	a	b	c	a
c	a	c	a	c

It is easy to check that Add, Perm, and Sum always have value a, but not Taut as $(c \vee c) \supset c$ has value c $(\neq a)$. Bernays also proved completeness by using the technique of normal forms (see section 5.3 for details on this and Bernays's independence proofs in 1918). Since Bernays's work did not appear in print until 1926, Post's paper (1921) contained the most advanced published results on the metatheory of the propositional calculus by the early 1920s. Similar results were also obtained by Łukasiewicz around 1924 (see Tarski 1983, 43).

8.3. Post's Contributions to the Metatheory of the Propositional Calculus

Post (1921) represent a qualitative change with respect to the previous studies of axiomatic systems for the propositional calculus by Russell, Sheffer, and Nicod. Post begins by explicitly stating the difference between proving results in a system and proving results about a system. He emphasizes that his results are about the system of propositional logic, which he takes in the version offered in *Principia* but regards it as a purely formal system to be investigated.[97] A basic concept introduced by Post is that of a truth table development. Post claims no originality for the concept, which he attributes to previous logicians. He denotes the truth value of any proposition p by $+$ if p is true and by $-$ if p is false.

The notion of truth table is then applied to arbitrary functions of the form $f(p_1, p_2, \ldots, p_n)$ of n propositions built up from p_1, p_2, $\ldots$, p_n by means of arbitrary applications of $\sim$ and $\vee$. Because each of the proposition can assume either $+$ or $-$ as values, there are 2^n possible truth configurations for $f(p_1, p_2, \ldots, p_n)$. In general there will be 2^{2^n} possible truth tables for functions of n arguments. Let us call such truth tables of order n. Post proves first of all that for any n, to every truth table of order n there is at least one function $f(p_1, p_2, \ldots, p_n)$ which has it for its truth table. He then distinguishes three classes of functions: positive, negative, and mixed. Positive functions are those that always take $+$ (this is the equivalent of Wittgenstein's propositional tautologies as defined in the *Tractatus* (1921, 1922), say, $p \vee {\sim}p$, negative functions are those that always take $-$ (say, ${\sim}(p \vee {\sim}p)$), and mixed are those functions that take both $+$'s and $-$'s (e.g., $p \vee p$).

Post's major theorem then proves that a necessary and sufficient condition for a function $f(p_1, p_2, \ldots, p_n)$ to be a theorem of the propositional system of *Principia* is that $f(p_1, p_2, \ldots, p_n)$ be positive (i.e., all its truth values be $+$). In our terminology, $f(p_1, p_2, \ldots, p_n)$ is a theorem of propositional logic if and only if $f(p_1, p_2, \ldots, p_n)$ is a tautology. The proof makes use of the possibility of transforming sentences of the propositional calculus into special normal forms. Post emphasizes that the proof of his theorem gives a method both for deciding whether a function $f(p_1, p_2, \ldots, p_n)$ is positive and for actually writing down a derivation of the formula from the axioms of the calculus. Nowadays the property demonstrated by Post is called (semantic) completeness, but Post uses the word "completeness" in a different sense. He uses the word to discuss the adequacy of a system of functions to express all the possible truth tables (this is nowadays called truth-functional completeness). In this way he shows not only that through the connectives of *Principia* ($\sim$ and $\vee$) one can generate all possible truth tables but also that there are only two connectives which can, singly, generate all the truth tables. One is the Sheffer stroke, and the other is the binary connective that is always false except in the case when both propositions are false. The techniques used by Post are now standard, and we will not rehearse them here. Rather, we would like to mention another important concept introduced by Post. Post needed to introduce a concept of consistency for arbitrary systems of connectives (which therefore might not have negation as a basic connective). Because an inconsistent system brings about the assertion of every proposition, he defined a system to be inconsistent if it yields the assertion of the variable p (which is equivalent to the derivability of every proposition if the substitution rule is present). From this notion derives our notion of Post-completeness: A system of logic is Post-complete if every time we add to it a sentence unprovable in it, we obtain an inconsistent system. Post proved that the propositional system of *Principia* is thus both semantically complete and Post-complete.

Another powerful generalization was offered by Post in the last part of his article. There he defines m-valued truth systems, that is, system of truth values where instead of two truth values ($+$ and $-$), we have finitely many values. This development is, together with (Łukasiewicz 1920b), one of the first studies of many-valued logics (see itinerary VII).

One final point about Post. Although the truth table techniques he developed belong squarely to what we call semantics, this does not mean that Post was after an analysis of logical truth or a "semantics." Rather, his interest seems to have been purely formal and aimed at finding a decision procedure for provability (see Dreben and van Heijenoort 1986, 46).

To sum up: By 1921 the classical propositional calculus has been shown to be consistent, semantically complete, Post-complete, and truth-functionally complete. Moreover, Bernays improved the presentation of the calculus given in *Principia* by showing that if one deletes associativity from the system, one obtains an axiomatic systems, all of whose axioms are independent.

8.4. Semantical Completeness of First-Order Logic

With the work by Bernays (1918, 1926) and Post (1921), the notions of Post-completeness and semantic completeness had been spelled out with the required precision. After the recognition of first-order logic ("functional calculus" or "restricted functional calculus") as an important independent fragment of logic, due in great part to Hilbert's 1917–1918 lectures and Hilbert and Ackermann (1928), the axiomatic investigation of first-order logic could also be carried out.

Chapter 3 of Hilbert and Ackermann (1928) became the standard exposition of the calculus. In section 9 of the chapter, Hilbert and Ackermann show that the calculus is consistent (by giving an arithmetical interpretation with a domain of one element). Then it is shown, crediting Ackermann for the proof, that the system is not Post-complete. To pose the completeness problem for first-order logic, it was necessary to identify the appropriate notion of validity [*Allgemeingültigkeit*]. This notion seems to be have been defined for the first time by Behmann (1922). It turns out that Behmann's approach to the decision problem led to the notion of validity for first-order formulas (with variables for predicates) and for second-order formulas. This is well captured in Bernays's concise summary of the work:

> In the decision problem we have to distinguish between a narrower and a wider formulation of the problem. The narrower problem concerns logical formulas of the "first order," that is, those in which the signs for all and exist (universal and existential quantifiers) refer only to individuals (of the assumed individual domain); the logical functions occurring here are variables, with the exception of the relation of identity ("x is identical with y"), which is the only individual [constant] relation admitted. The task consists in finding a general procedure which allows to decide, for any given formula, whether it is valid [*allgemeingültig*], that is, whether it yields a correct assertion [*richtige Aussage*] for arbitrary substitutions of determinate logical functions.
>
> One arrives at the wider problem by applying the universal and the existential quantifiers in connection to function variables. Then one considers formulas of the "second order" in which all variables are bound by universal and existential quantifiers, in whose meaning therefore nothing remains undetermined except for the number of individuals which are taken as given at the outset. For an arbitrary given formula of this sort one must now decide whether it is correct or not, or for which domains it is correct. (Bernays 1928a, 1119–1120)

A logical formula, in this context, is one that is expressible only by means of variables (both individual and functional), connectives, and quantification over individual variables, that is, there are no constants (see Hilbert and Ackermann 1928, 54). With this in place, the problem of completeness is posed by Hilbert and Ackermann as the request for a proof that every logical formula (of the restricted functional calculus) which is correct for every domain of individuals (*Individuenbereich*) be shown to be derivable from the axioms by finitely many applications of the rules of logical inference (68).[98]

Hilbert and Ackermann also posed the problem to show the independence of the axioms for the restricted functional calculus. Both problems were solved in 1929 by Kurt Gödel in his dissertation and published in "The completeness of the axioms of the functional calculus of logic" (1929, 1930a). The solution to the completeness problem is the most important one. As there exist already several expositions of the proof (Kneale and Kneale 1962; Dreben and van Heijenoort 1986) we can simply outline the main steps of the demonstration. Let us begin with the axioms for the system:

1. $X \vee X \to X$,
2. $X \to X \vee Y$,
3. $X \vee Y \to Y \vee X$,
4. $(X \to Y) \to (Z \vee X \to Z \vee Y)$,
5. $(x)F(x) \to F(y)$,
6. $(x)[X \vee F(x)] \to X \vee (x)F(x)$.

Rules of inference:

1. From A and $A \to B$, B may be inferred.
2. Substitution for propositional and functional variables.
3. From $A(x)$, $(x)A(x)$ may be inferred.
4. Individual variables (free or bound) may be replaced by any others (with appropriate provisos).

A valid formula (*allgemeingültige Formel*) is one that is satisfiable in every domain of individuals. Gödel's completeness theorem is stated as:

Theorem I Every valid formula of the restricted functional calculus is provable.

If a formula A is valid, then $\overline{A}$ is not satisfiable. By definition "A is refutable" means "$\overline{A}$ is provable." This leads Gödel to restate the theorem as follows:

Theorem II Every formula of the restricted functional calculus is either refutable or satisfiable (and, moreover, satisfiable in the denumerable domain of individuals).

Suppose in fact we have shown Theorem II. To prove Theorem I, assume that A is universally valid. Then $\overline{A}$ is not satisfiable. By Theorem II, it is refutable, that is, it is provable that $\overline{\overline{A}}$. Thus, it is also provable that A.

We can thus focus on the proof of Theorem II and, without loss of generality, talk about sentences rather than formulas. The first step of the proof consists in reducing the complexity of dealing with arbitrary sentences to a special class in normal form. The result is an adaptation of a result given by Skolem in 1920. Gödel appeals to the result (from Hilbert and Ackermann 1928) that for each sentence S there is an associated normal sentence S^* such that S^* has all the quantifiers at the front of a quantifier-free matrix, and it is provable that $S^* \leftrightarrow S$. Gödel then focuses on sentences that in addition to being in prenex normal form are such that the prefix of the sentence begins with a universal quantifier and ends with an existential quantifier. Let us call such sentences K-sentences.

Theorem III establishes that if every K-sentence is either refutable or satisfiable, so is every sentence. This reduces the complexity of proving Theorem II to the following.

Every K-sentence is either satisfiable or refutable. The proof is by induction on the degree of the K-sentence, where the degree of a K-sentence is defined by counting the number of blocks in its prefix consisting of universal quantifiers that are separated by existential quantifiers. The inductive step is quite easy (Theorem IV). The real core of the proof is showing the result for K-sentences of degree 1:

Theorem V Every K-sentence of degree 1 is either satisfiable or refutable.

Proof: Assume we have a K-sentence of degree 1 of the form

$$(P)M = (x_1)\dots(x_r)(Ey_1)\dots(Ey_s)M(x_1,\dots,x_r,y_1,\dots,y_s).$$

For the sake of simplicity, let us fix $r = 3$ and $s = 2$.

Select a denumerable infinity of fresh variables z_0, z_1, $z_2,\dots$. Consider all 3-tuples of z_0, z_1, $z_2,\dots$ obtained by allowing repetitions of the variables and ordered according to the following order: $\langle z_{k_1}, z_{k_2}, z_{k_3}\rangle < \langle z_{t_1}, z_{t_2}, z_{t_3}\rangle$ iff $(k_1 + k_2 + k_3) < (t_1 + t_2 + t_3)$ or $(k_1 + k_2 + k_3) = (t_1 + t_2 + t_3)$ and $\langle k_1, k_2, k_3\rangle$ precedes $\langle t_1, t_2, t_3\rangle$ in the lexicographic ordering. In particular, the enumeration begins with $\langle z_0, z_0, z_0\rangle$, $\langle z_0, z_0, z_1\rangle$, $\langle z_0, z_1, z_0\rangle$, and so on. Let $\mathbf{w}_n$ be the nth triple in the enumeration.

We now define an infinite sequence of formulas from our original sentence as follows:

$$\begin{aligned} M_1 &= M(z_0, z_0, z_0; z_1, z_2)\\ M_2 &= M(z_0, z_0, z_1; z_3, z_4)\ \&\ M_1\\ &\ \vdots\\ M_n &= M(\mathbf{w}_n; z_{2(n-1)+1}, z_{2n})\ \&\ M_{n-1}. \end{aligned}$$

(Recall that our example works with $s = 2$).

Notice that the variables appearing after the semicolon are always fresh variables that have neither appeared before the semicolon nor in previous M_i's. Moreover, in each M_i except M_1 all the variables appearing before the semicolon have also appeared previously.

Now define $(P_n)M_n$ to be $(Ez_0)(Ez_1)\dots(Ez_{2n})M_n$. Thus, $(P_n)M_n$ is a sentence all of whose variables are bound by the existential quantifiers in its prefix.

With the above in place, Gödel proves (Theorem VI) that for every n, $(P)M$ implies $(P_n)M_n$. The proof, which we omit, is by induction on n and exploits the specific construction of the M_n's. The important point here is that the structure of the M_n's is purely propositional. Thus each M_n will be built out of functional variables $P_1(x_{p_1},\dots,x_{q_1})$, …, $P_k(x_{p_k},\dots,x_{q_k})$ (of different arity) and propositional variables X_1, …, X_l, (the elementary components, all of which are already in M) by use of "or" and "not." At this point we associate with every M_n a formula B_n of the propositional calculus obtained by replacing all the elementary components by propositional variables in such a way that to different components we associate different propositional variables. Thus, we can exploit the completeness theorem for the propositional calculus. B_n is either satisfiable or refutable.

Case 1. B_n is refutable. Then $(P_n)M_n$ is also refutable and so is

$$(x_1)\dots(x_r)(Ey_1)\dots(Ey_s)M(x_1,\dots,x_r;y_1,\dots,y_s).$$

Case 2. No B_n is refutable. Thus they are all satisfiable. Thus for each n, there are systems of predicates defined on the integers $\{0,\dots,ns\}$ and truth values t_0, ..., t_l for the propositional variables such that a true proposition results if in B_n we replace the P_i's by the system of predicates, the variables z_i by the natural numbers i, and X_i by the corresponding t_i.

Thus, for each M_n we have been able to construct an interpretation, with finite domain on the natural numbers, which makes M_n true. The step that clinches the proof consists in showing that since there are only finitely many alternatives at each stage n (given that the domain is finite) and that each interpretation that satisfies M_{n+1} makes true the previous M_n's, it follows that there is an infinite sequence of interpretations S_1, S_2, and so on. such that S_{n+1} contains all the preceding ones. This follows from an application of König's lemma, although Gödel does not explicitly appeal to König's result. From this infinite sequence of interpretations it is then possible to define a system satisfying the original sentence $(x_1)\dots(x_r)(Ey_1)\dots(Ey_s)M(x_1,\dots,x_r;$ $y_1,\dots,y_s)$ by letting the domain of interpretation be the natural numbers (hence a denumerable domain!) and declaring that a certain predicate appearing in M is satisfied by an n-tuple of natural numbers if and only if there is at least an n such that in S_n the predicate holds of the same numbers. Similarly the propositional variables occurring in M are given values according to whether they are given those values for at least one S_n. This interpretation satisfies $(P)M$.

This concludes the proof. Gödel generalizes the result to countable sets of sentences and to first-order logic with identity. The former result is obtained as a corollary to Theorem X, which is what we now call the compactness theorem: For a denumerably infinite system of formulas to be satisfiable, it is necessary and sufficient that every finite subsystem be satisfiable.[99]

8.5. Models of First-Order Logic

Although we have already discussed the notion of *Allgemeingültigkeit* in the presentation of the narrow functional calculus in Hilbert, it will be useful to go back to it to clarify how models are specified for such languages.

One first important point to notice is that both in Hilbert and Ackermann (1928) and in Bernays and Schönfinkel (1928), the problem of *Allgemeingültigkeit* is that of determining for logical expressions that have no constants whether a correct expression results for arbitrary substitution of values for the (predicate) variables. As a result, an interpretation for a logical formula becomes the assignment of a domain together with a system of individuals and functions. For instance $(x)(F(x) \vee \overline{F(x)})$ is, according to Bernays-Schönfinkel, *allgemeingültig* for every domain of individuals (i.e., by substituting a logical function for F one obtains a correct sentence). Tarski (1933b, 199, n. 3) points out that what is at stake here is not the notion of "correct or true sentence in an individual domain a" because the central concept in Hilbert-Ackermann and Bernays-Schönfinkel is that of sentential functions with free variables and not that of sentence (Tarski implies that one can properly speak of truth of sentences only; this is also in Ajdukiewicz 1921). For this reason, Tarski says, these authors use *allgemeingültig*, as opposed to *richtig* or *wahr*. This is, however, misleading in that *richtig* and *wahr* are used by the above-mentioned authors all over the place. Tarski is nevertheless right in pointing out that when, for a specific individual domain, we assign an interpretation to F, say X (a subset of the domain), we are still not evaluating the truth of $(x)(F(x) \vee \overline{F(x)})$, because the latter expression is not a sentence as F is free in it.[100]

In Gödel's dissertation we find the following presentation of the notion of satisfaction in an interpretation:

> Let A be any logical expression that contains the functional variables $F_1, F_2, \ldots, F_k$, the free individual variables $x_1, x_2, \ldots, x_l$, the propositional variables $X_1, X_2, \ldots, X_m$, and otherwise, only bound variables. Let S be a system of functions $f_1, f_2, \ldots, f_k$ (all defined in the same universal domain), and of individuals (belonging to the same domain), $a_1, a_2, \ldots, a_l$, as well as propositional constants, $A_1, A_2, \ldots, A_m$.
>
> We say that this system, namely $(f_1, f_2, \ldots, f_k, a_1, a_2, \ldots, a_l, A_1, A_2, \ldots, A_m)$ satisfies the logical expression if it yields a proposition that is true (in the domain in question) when it is substituted in the expression. (Gödel 1929, 69)[101]

We see that also in Gödel's case the result of substituting objects and functions into the formula is seen as yielding a sentence, although properly speaking one does not substitute objects into formulas. Unless what he means is that symbols denoting the objects in the system have to be substituted in the formula. Lack of clarity on this issue is typical of the period.

8.6. Completeness and Categoricity

In the introductory remarks to his "Untersuchungen zur allgemeinen Axiomatik," written around 1927–1929, Carnap wrote:

> By means of the new investigations on the general properties of axiomatic systems, such as, among others, completeness, monomorphism (categoricity), decidability [*Entscheidungsdefinitheit*], consistency and on the problems of the criteria and mutual relationships between these properties, it has become more and more clear that

> the main difficulty lies in the insufficient precision of the concepts applied. (Carnap 2000 [1927–1929], 59)

Carnap's work remained unpublished at the time, except for the programmatic (1930), but the terminological and conceptual confusion reigning in logic had been remarked by other authors. Let us first pursue the development of the notions of completeness and categoricity in the 1920s and early 1930s.

Recall the notion of completeness found in the postulate theorists (see section 1.4): A complete set of postulates is one such that its postulates are consistent, independent of each other, and sufficient, where "sufficiency" means that only one interpretation is possible.

According to contemporary terminology, a system of axioms is categorical if all its interpretations (or models) are isomorphic. In the early part of the twentieth century it was usually mentioned, for example, that Dedekind had shown that every two interpretations of the axiom system for arithmetic are isomorphic. One thing on which there was already clarity is that two isomorphic interpretations make the same set of sentences true. We know today that issues of categoricity are extremely sensitive to the language and logic in which the theory is expressed. Thus the set of axioms for first-order Peano arithmetic is not categorical (an immediate consequence of the Löwenheim–Skolem theorem and/or of Gödel's incompleteness theorem) but second-order arithmetic is categorical (at least with respect to standard second-order models). This sheds light on some of the early confusions. One such confusion was the tendency to infer the possibility of incompleteness results from the existence of nonisomorphic interpretations. Consider Skolem (1922): "Since Zermelo's axioms do not uniquely determine the domain B, it is very improbable that all cardinality problems are decidable by means of these axioms."

As an example he mentions the continuum problem.[102] The implicit assumption here is that if a system is not categorical, then there must be sentences A and $\neg A$ such that one of the interpretations makes A true and the other makes $\neg A$ true. That the situation is not as simple became clear only very late. In 1934, Skolem proved that there are nonisomorphic countable models of first-order Peano arithmetic which make true exactly the same (first-order) sentences. In later developments, the notion of elementary equivalent models was introduced to capture the phenomenon (see following discussion).

To gauge what the issues were surrounding a proper understanding of categoricity, let us look at how von Neumann deals with categoricity (1925). In the first part of his article, von Neumann discusses the Löwenheim–Skolem theorem, which shows that every set of first-order sentences that is satisfied by an infinite domain can also be satisfied in a denumerable domain. This immediately implies that no first-order theory which admits a nondenumerable interpretation can be categorical (in our sense). This should settle the problem of categoricity for the axioms being discussed by von Neumann. Indeed, von Neumann draws the right conclusion concerning the system of set theory: "We now know that, if it is at all possible to find a system S satisfying the axioms, we can also find such system in which there are only denumerably many I-objects and denumerably many II-objects" (von Neumann 1925, 409). Why then, in the following section (§6), does he discuss the issue of categoricity again? A careful reading shows that he is appealing to categoricity as nondisjunctiveness (see Veblen 1904), that is, an axiom system is categorical if it is not possible to add independent axioms to it.

An early attempt to provide a terminological clarification concerning different meanings of completeness is found in the second edition of *Einleitung in die Mengenlehre* (1923), where Fraenkel distinguishes between completeness in the sense of categoricity and completeness as decidability (*Entscheidungsdefinitheit*).[103] Both concepts of completeness are also discussed in Weyl (1927), but Weyl rejects completeness as decidability (for every sentence A, one should be able to derive from the axioms either A or $\neg A$) as a "philosopher's stone."[104] The only meaning of completeness that he accepts is the following: "The final formulation is thus the following: An axiom system is complete when two (contentual) interpretations of it are necessarily isomorphic" (Weyl 1927, 22). In this sense, he adds, Hilbert's axiomatization of geometry is complete.

In the third edition of *Einleitung in die Mengenlehre* (1928), Fraenkel adds a third notion of completeness, the notion of *Nichtgabelbarkeit* ("nonforkability"), meaning essentially that every two interpretations satisfy the same sentences (today we call this "semantic completeness"). Carnap (1927) claims that the first two notions are identical, and in 1930, he claimed to have proved the equivalence of all three notions (which he calls monomorphism, decidability, and nonforkability). The proofs were supposed to be contained in his manuscript "Untersuchungen zur allgemeinen Axiomatik," but his approach there is marred by his failure to distinguish between object language and metalanguage, and between syntax and semantics, and thus to specify exactly to which logical systems the proofs are supposed to apply (for an analysis of these issues see Awodey and Carus 2001; Carnap's unpublished investigations on general axiomatics are now edited in Carnap 2000 [1927–1929]). Gödel, however, had access to the manuscript and, in fact, his (1929) dissertation acknowledges the influence of Carnap's investigations (as does Kaufmann 1930). Awodey and Carus (2001, 23) also point out that Gödel's first presentation of the incompleteness theorem in Königsberg in 1930 (see Gödel 1995, 29 and the introduction by Goldfarb) was aimed specifically at Carnap's claim. Indeed, when speaking of the meaning of the completeness theorem for axiom systems, he pointed out that in first-order logic monomorphicity (Carnap's terminology) implies (syntactic) completeness (*Entscheidungsdefinitheit*). If syntactic completeness also held of higher-order logic then (second-order) Peano arithmetic, which by Dedekind's classical result is categorical, would also turn out to be syntactically complete. But, and here is the first announcement of the incompleteness theorem, Peano's arithmetic is incomplete (Gödel 1930a, 28–30).

An important result concerning categoricity was obtained by Tarski in work done in Warsaw between 1926 and 1928. He showed that if a consistent set of first-order propositions does not have finite models, then it has a nondenumerable model (upward Löwenheim–Skolem). This shows that no first-order theory that admits of an infinite domain can be categorical (*kategorisch*). The result was mentioned publicly for the first time in 1934 in the editor's remarks at the end of Skolem (1934). A proof by Malcev stating that, under the assumptions, the theory has models of every infinite cardinality was published in (1936);[105] this result was apparently also obtained by Tarski in his Warsaw seminar (see Vaught 1974, 160). Other results that Tarski obtained in the period (1927–1929) include the result that a first-order theory that contains as an extra-logical symbol "$<$" and that is satisfied in the order type ω is also satisfied in every set of order type $\omega + (\omega^* + \omega)\tau$, where ω^* is the reverse of the standard ordering on ω, and τ is an arbitrary order type. This was eventually to lead to the notion of *elementary equivalence*, defined for order types in the appendix to Tarski (1936a). This allowed Tarski to give a number of nondefinability results. In the same appendix he shows that, using η for the order type of the rationals, every order type of dense order is elementarily equivalent to one of the following types: η, $1+\eta$, $\eta+1$, and $1+\eta+1$ (which are not elementary equivalent to each other). He thus concluded that properties of order types such as continuity or nondenumerability cannot be expressed in the language of the elementary theory of order. Moreover, using the elementary equivalence of the order types ω and $\omega + (\omega^* + \omega)$, he also showed that the property of well ordering is not expressible in the elementary theory of order (Tarski 1936a, 380).

One of the techniques investigated in Tarski's seminar in Warsaw was what he called the elimination of quantifiers. The method was originally developed in connection to decidability problems by Löwenheim (1915) and Skolem (1920). It basically consists in showing that one can add to the theory certain formulas, perhaps containing new symbols, so that in the extended theory it is possible to demonstrate that every sentence of the original theory is equivalent to a quantifier-free sentence of the new theory. This idea was cleverly exploited by Langford to obtain, for instance, decision procedures for the first-order theories of linear dense orders without endpoints, with first but no last element and with first and last element (1927a) and for the first-order theory of linear discrete orders with a first but no last element (1927b). As Langford emphasizes (1927a), he is concerned with "categoricalness," that is, that the theories in question determine the truth value of all their sentences (something he obtains by showing that the theory is syntactically complete). Many such results were obtained afterward, such as Presburger's (1930) elimination of quantifiers for the additive theory of the integers and Skolem's (1929b) for the theory of order and multiplication (but without addition!) on the natural numbers. Tarski himself announced in 1931 to have obtained, by similar techniques, a decision procedure for elementary algebra and geometry (published however only in 1948). Moreover, he extended the results by Langford to the first-order theory of discrete order without a first or last element and to the first-order theory of discrete order with first and last element. This work is relevant to the study of models in that it allows the study of all the complete extensions of the systems under consideration and leads naturally to the notion of elementary equivalence between relational structures (for order types) that Tarski developed in his seminar. This work also dovetails with Tarski's "On certain fundamental concepts of metamathematics" (1930b), where for instance he proves Lindenbaum's result that every consistent set of sentences has a complete consistent extension. For reason of space, Tarski's contributions to metamathematics during this period cannot be discussed in their full extent, and we limit ourselves here to his definition of truth.[106]

Another important result concerning categoricity, or lack thereof, was obtained by Skolem (1933, 1934) (Skolem speaks of "complete characterizability"). The results we have mentioned so far, the upward and downward Löwenheim–Skolem theorems are consistent with the possibility that, for instance, there is only one countable model, up to isomorphism, for first-order Peano arithmetic. What Skolem showed was, in our terminology, that there exist countable models of Peano arithmetic that are not isomorphic. He constructed a model N^* of (classes of equivalence of) definable functions (hence the countability of the new model) which has all the constant functions ordered with the order type of the natural numbers and followed by nonstandard elements, which eventually majorize the constant functions, for instance, the identity function (for details see also Zygmunt 1973). Indeed, Skolem's result states that no finite (in 1933) or countable (in 1934) set of first-order sentences can characterize the natural numbers. The 1934 result implies that N^* can be taken to make true exactly those sentences that are true in N.

8.7. Tarski's Definition of Truth

The most important contribution to semantics in the early thirties was made by Alfred Tarski. Although his major work on the subject, "The concept of truth in formalized languages," came out in 1933 in Polish (1935 in German), Tarski said that most of the investigations contained in it date from 1929. However, the seeds of Tarski's reflection on truth were planted early on by the works of Ajdukiewicz (1921) and the lectures of Leśniewski.[107]

Tarski specifies the goal of his enterprise at the outset:

> The present article is almost wholly devoted to a single problem—*the definition of truth.* Its task is to construct—with reference to a given language—*a materially adequate and formally correct definition of the term "true sentence."* (Tarski 1933a, 152)

A materially adequate definition is one that for each sentence specifies under what conditions it must be considered true. A formally correct definition is one that does not generate a contradiction and uses only certain concepts and rules specified in advance. One should not expect the definition to give a criterion of truth. It is not the role of the definition to tell us whether "Paris is in France" is true but only to specify under what conditions the sentence is true.

Tarski begins by specifying that the notion of truth he is after is the one embodied in the classical conception of truth, where a sentence is said to be true if it corresponds with reality. According to Tarski, the definition of truth should avoid appeal to any semantical concepts, which have not been previously defined in terms of nonsemantical concepts. In Tarski's construction, truth is a predicate of sentences. The extension of such a predicate depends on the specific language under consideration; thus the inquiry is to take the form of specifying the concept of truth for specific individual languages. The first section of the paper describes at length the prospects for defining truth for a natural language and concludes that this is a hopeless task. Let us see what motivates this negative conclusion. Tarski first proposes a general scheme of what might count as a first approach toward a definition of the expression "x is a true sentence":

(∗) x is a true sentence if and only if p.

Concrete definitions are obtained by substituting for "p" any sentence and for "x" the name of the sentence. Quotation marks are one of the standard devices for creating names (but not the only one). If p is a sentence, we can use quotation marks around p to form a name for p. Thus, a concrete example of (∗) could be

(∗∗) "It is snowing" is a true sentence if and only if it is snowing.

The first problem with applying such a scheme to natural language is that although (∗) looks innocuous, one needs to be wary of the possibility of the emergence of paradoxes, such as the liar paradox. Tarski rehearses the paradox and notices that at a crucial point one substitutes in (∗) for "p" a sentence, which itself contains the term "true sentence." Tarski does not see a principled reason that such substitutions should be excluded, however. In addition, more general problems stand in the way of a general account. First of all, Tarski claims that if one treats quotation mark names as syntactically simple expressions the attempt to provide a general account soon runs into nonsense. Therefore, he points out that quotation mark names have to be treated as complex functional expressions, where the argument is a sentential variable, p, and the output is a quotation mark name. The important fact in this move is that the quotation mark name "p" now can be seen to have structure. According to Tarski, however, even in this case new problems emerge, for example, one ends up with an intensional account, which might be objectionable (even if p and q are equivalent, their names, "p" and "q," will not be). This leads Tarski to try a new strategy by attempting to provide a structural definition of true sentence which would look roughly as follows:

> *A true sentence is a sentence which possesses such and such structural properties* (i.e., properties concerning the form and arrange-

> ment in sequence of the single part of the expression) *or which can be obtained from such and such structurally described expressions by means of such and such structural transformations.* (Tarski 1933a, 163)

The major objection to this strategy is that we cannot, due to the open nature of natural languages, specify a structural definition of sentence, let alone of true sentence. Moreover, natural languages are "universal," that is, they contain such terms as "true sentence," "denote," "name," and so on, which allow for the emergence of self-reference such as the one leading to the liar antinomy. Tarski concluded:

> If these observations are correct, *then the very possibility of a consistent use of the expression "true sentence" which is in harmony with the laws of logic and the spirit of everyday language seems to be very questionable, and consequently the same doubt attaches to the possibility of constructing a correct definition of this expression.* (Tarski 1933a, 165)

Thus, the foregoing considerations explain a number of essential features of Tarski's account. First, the account will be limited to formal languages. For such languages it is in fact possible to specify the syntactic rules that define exactly what a well-formed sentence of the language is. Moreover, such languages are not universal, that is, one can keep the level of the object language and that of the metalanguage (which is used to describe the semantic properties of the object language) separate. When we talk about theories specified in a certain language, then we distinguish between the theory and the metatheory, where the latter is used to study the syntactic and semantic properties of the former.

Tarski provides then the definition of truth for a specific language, that is, the calculus of classes, but the treatment is extended in the later sections of the essay to provide a definition of truth for arbitrary languages of finite type. One important point stressed by Tarski is that the definition of truth is intended for "concrete" deductive systems, specifically, deductive systems that are interpreted. For purely formal systems, Tarski claims that the problem of truth cannot be meaningfully raised.

The calculus of classes is a subtheory of mathematical logic that deals with the relationships between classes and the operation of union, intersection and complement. There are also two special classes, the universal class and the empty class. The intuitive interpretation of the theory that Tarski has in mind is the standard one with the individual variables ranging over classes of individuals. In the following we give an (incomplete) sketch of the structure of the language L of the calculus of classes (with only instances of the axioms) and of the metalanguage, ML, in which the definition of truth is given. It should be pointed out that Tarski does not completely axiomatize the metalanguage, which is presented informally, and he uses the Polish notation in his presentation.

The Language of the Calculus of Classes

Variables: $x_{\prime}$, $x_{\prime\prime}$, $x_{\prime\prime\prime}$, ...

Logical constants: N [negation], A [disjunction], Π [universal quantifier]; relational constant: I [inclusion]

Expressions and formulas are defined as usual.

Logical axioms: $ANAppp$ [$\sim(p \vee p) \vee p$], etc.

Proper axioms: $\Pi\, x_1 I x_1 x_1$ [every class is included in itself]; $\Pi\, x_{\prime}x_{\prime\prime}x_{\prime\prime\prime}ANIx_{\prime}x_{\prime\prime}ANIx_{\prime\prime}x_{\prime\prime\prime}Ix_{\prime}x_{\prime\prime\prime}$ [transitivity of I], etc.

Rules of inference: substitution, modus ponens, introduction, and elimination of Π.

The Metalanguage

Logical constants: not, or, for all

Relational constants: $\subseteq$

Class theoretical terms: $\in$, individual, identical ($=$), class, cardinal number, domain, etc.

Terms of the logic of relations: ordered n-tuple, infinite sequence, relation, etc.

Terms of a structural descriptive kind: ng [for N]; sm [for A], un [for Π], v_k [the kth variable], $x \frown y$ [the expression that consists of x followed by y], etc. These form names of object-language expressions in the metalanguage.

Auxiliary symbols are introduced to give metatheoretical shorthands for whether an expression is an inclusion, a negation, a disjunction, or a universal quantification. They are: $x = \iota_{k,l}$ iff $x = (\text{in} \frown v_k) \frown v_l$, $x = \overline{y}$ iff $x = \text{ng} \frown y$; $x = y + z$ iff $x = (\text{sm} \frown y) \frown z)$; $x = \cap_k y$ iff $x = (\text{un} \frown v_k) \frown y$.

Variables:

1. a, b [names for classes of an arbitrary character]
2. f, g [sequences of classes]
3. k, l, m, n [natural numbers and sequences of natural numbers]
4. t, u, w, x, y, z [expressions]
5. X, Y [sequences of expressions]

The Metatheory

Logical axioms: not (p or p) or p, etc.

Axioms of the theory of classes: for all a, $a \subseteq a$ etc.

Proper axioms: several axioms characterizing the notion of expression. Intuitively, this is the smallest class X containing ng, sm, $\cap$, ι, v_k, such that if x, y are in X then $x \frown y$ is in X.

With this in place, we can give names in *ML* to every expression in L. For instance, $NIx_{\prime}x_{\prime\prime}$ is named in *ML* by $((\mathrm{ng} \frown \mathrm{in}) \frown v_1) \frown v_2$ or $\overline{\iota_{12}}$. We can now define the notions of the following.

Sentential function (Definition 10): Sentential functions are obtained by the closure of expressions of the form ι_{ik} under negation, disjunction, and universal quantification.

Sentence: A sentential function with no free variables is a sentence.

Axioms: A sentence is an axiom if it is the universal closure of either a logical axiom or of an axiom of the theory of classes.

Theorems: A sentence is a theorem if it can be derived from the axioms using substitution, modus ponens, introduction, and elimination rules for universal quantifier.

With this machinery in place (all of which is purely syntactical), Tarski proceeds to give a definition of truth for the calculus of classes. The richness of the metalanguage provides us both with a name of the sentence and a sentence with the same meaning (a translation into the metalanguage) for every sentence of the original calculus of classes. For instance, to '$\Pi\, v_{\prime}Iv_{\prime}v_{\prime}$' in L corresponds the name $\cap_1\iota_{11}$ and the sentence "for all a, $a \subseteq a$." The schema (∗) should now be recaptured in such a way that for any sentence of the calculus of classes its name in the metalanguage appears in place of x and in place of p we have the equivalent sentence in the metalanguage:

$$\cap_1\iota_{11} \text{ is a true sentence if and only if for all } a,\ a \subseteq a.$$

What is required of a satisfactory truth definition is that it contains all such equivalences in its extension. More precisely, let *Tr* denote the class of all true sentences and S the class of sentences. Then *Tr* must satisfy the following convention.

Convention T A formally correct definition of the symbol "*Tr*" formulated in the metalanguage, will be called an *adequate definition of truth* if it has the following consequences:

α. all sentences obtained from the expression "$x \in Tr$ if and only if p" by substituting for the symbol "x" a structural-descriptive name of any sentence of the language in question and for the symbol "p" the expression that forms the translation of this sentence into the metalanguage;

β. the sentence "for any x, if $x \in Tr$ then $x \in S$" (in other words, $Tr \subseteq S$). (Tarski 1933a, 188)

Ideally, one would like to proceed in the definition of truth by recursion on the complexity of sentences. Unfortunately, on account of the fact that sentences are in general not obtained from other sentences but rather from formulas (which, in general, may contain free variables), a recursive definition of "true sentence" cannot be given directly. However, complex formulas are obtained from formulas of smaller complexity, and here the recursive method can be applied. For this reason Tarski defines first what it means for a formula to be satisfied by given objects. Actually, for reasons of uniformity, Tarski defines what it means for an infinite sequence of objects to satisfy a certain formula. Definition of satisfaction (Definition 22): Let f be an infinite sequence of classes, and f_i the ith coordinate. Satisfaction is defined inductively on the complexity of formulas (denoted by x, y, z).

Atomic formulas: f satisfies the sentential function $(\iota_{k,l})$ iff $(f_k \subset f_l)$.

Molecular formulas:

a. for all f, y: f satisfies $\overline{y}$ iff f does not satisfy y;

b. for all f, y, z: f satisfies $y + z$ iff f satisfies y or f satisfies z;

c. for all f, y, k: f satisfies $\cap_k y$ iff every sequence of classes that differs from f at most in the kth place satisfies the formula y.

This definition is central to Tarski's semantics, since through it one can define the notions of denotation (the name "c" denotes a, if a satisfies the propositional function $c = x$), definability, and truth. A closer look at the definition of satisfaction shows that whether a sequence satisfies a formula depends only on the coordinates of the sequence corresponding to the free variables of the formula. When the formula is a sentence, there are no free variables, and thus either all sequences satisfy it or no sequence satisfies it. Correspondingly, we have the definition of truth and falsity for sentences given in Definition 23: x is a true sentence iff x is a sentence and every infinite sequence of classes satisfies x. Tarski then argues that the definition given is formally correct and satisfies Convention T.

Among the consequences Tarski draws from the precise definition of the class of true sentences is the fact that the theorems of the calculus of classes are a proper subset of the truths of the calculus (under the intended interpretation).

Nowadays such definitions of satisfaction and truth are given by first specifying what the domain of the interpretation is, but Tarski does not do that. He speaks of infinite sequences of classes as if these sequences were taken from a universal domain. Indeed, on p. 199 of his essay, Tarski contrasts his approach with the relativization of the concept of truth to that of "correct or true sentence in an individual domain a." This is the approach, he points out, of the Hilbert school in Göttingen and contains his own approach as a special case. Of course, Tarski claims to be able to give a precise meaning of the notions (Definitions 24 and 27) that were used only informally by the Hilbert school.[108]

The remaining part of the essay sketches how to generalize the approach to theories of finite order (with a fixed finite bound on the types) and points out the limitations in extending the approach to theories of infinite order. However, even in the latter case Tarski establishes that "the consistent and correct use of the concept of truth is rendered possible by including this concept in the system of primitive concept of the metalanguage and determining its fundamental properties by means of the axiomatic method" (266).

By far the most important result of the final part of the essay is Tarski's celebrated theorem of the undefinability of truth, which he obtained after reading Gödel's paper on incompleteness.[109] Basically, the result states that there is no way to express $Tr(x)$ as a predicate of object languages (under certain conditions) without running into contradictions. In particular, for systems of arithmetic such as Peano arithmetic, this says there is no arithmetical formula $Tr(x)$ such that $Tr(x)$ holds of a code of a sentence just in case that sentence is true in the natural numbers.

We have seen that Tarski emphasized that through the notion of satisfaction other important semantic notions, such as truth and definability, can be also defined. Thus, the work on truth also provided an exact foundation for (1930a) and (1931), on definable sets of real numbers and the connection between projective sets and definable sets, and to the general investigation on the definability of concepts carried out by Tarski in the mid-1930s.

One of the most important applications of the new semantic theory was the notion of logical consequence (1936b). Starting from the intuitive observation that a sentence X follows from a class of sentences K if "it can never happen that both the class K consists only of true sentences and the sentence X is false" (414), Tarski made use of his semantical machinery to give a definition of the notion of logical consequence. First he defined the notion of model. Starting with a class L of sentences, Tarski replaces all nonlogical constants by corresponding variables, obtaining the class of propositional sentences L'. Then he says:

> An arbitrary sequence of objects which satisfies every sentential function [formula] of the class L' will be called a *model* or *realization* of the class L of sentences (in just this sense one usually speaks of an axiom system of a deductive theory). (Tarski 1936b, 417)

From this he obtains the notion of logical consequence: "The sentence X *follows logically* from the sentence of the class K if and only if every model of the class K is also a model of the sentence X" (Tarski 1936b, 417). There are several controversial issues concerning the exact interpretation of Tarski's theory of truth and logical consequence; these cannot be treated adequately within the narrow limits of this exposition.[110]

In any case, the result of Tarski's investigations for logic and philosophy cannot be overestimated. The standard expositions of logic nowadays embody, in one form or another, the definition of truth in a structure, which ultimately goes back to Tarski's article. Tarski's article marks also an explicit infinitistic attitude to the metatheoretical investigations, in sharp contrast to the finitistic tendencies of the Hilbert school. As a consequence the definition of truth is often nonconstructive. Often, but not always: In the particular case of the calculus of classes, Tarski shows that from the definition of truth one can also extract a criterion of truth; but he also remarks that this depends on the specific peculiarities of the theory and in general this is not so. Finally, Tarski's definition of truth and logical consequence have shaped the discussion of these notions in contemporary philosophy and are still at the center of current debates.

Notes

We have received comments on an earlier draft of this paper from Mark van Atten, José Ferreiros, Johannes Hafner, Ignasi Jané, Bernard Linsky, Enrico Moriconi, Chris Pincock, and Bill Tait. Their help is gratefully acknowledged.

1. Each author has been responsible for specific sections of the essay: PM for I–III, VII, and VIII; RZ for itineraries V and VI; and CB for itinerary IV. While responsibility for the content of each section rests with its author, for the sake of uniformity of style we use "we" rather than "I" throughout. A book-length treatment of the topics covered in itinerary IV is Badesa (2004). Itinerary V contains passages from Richard Zach, "Completeness before Post: Bernays, Hilbert, and the development of propositional logic," *The Bulletin of Symbolic Logic* 5 (1999) 331–366, © 1999, Association for Symbolic Logic, which appear here with the kind permission of the Association for Symbolic Logic. Itinerary VI contains passages from Richard Zach, "The practice of finitism: Epsilon calculus and consistency proofs in Hilbert's program," *Synthese* 137 (2003), 79–94, © 2003, Kluwer Academic Publishers, which appear here with the kind permission of Kluwer Academic Publishers.

2. On Zermelo's contribution to mathematical logic during this period see Peckhaus (1990, chapter 4); see also Peckhaus (1992).

3. In 1914, Philip Jourdain drew the same distinction but related it to two different conceptions of logic:

> We can shortly but very accurately characterize the dual development of the theory of symbolic logic during the last sixty years as follows: The calculus ratiocinator aspect of symbolic logic was developed by Boole, De Morgan, Jevons, Venn, C. S. Peirce, Schröder, Mrs Ladd Franklin and others; the lingua characteristica aspect was developed by Frege, Peano and Russell. (Jourdain 1914, viii)

Couched in the Leibnizian terminology, we thus find the distinction of logic as calculus versus logic as language, which van Heijenoort (1967b) made topical in the historiography of logic.

4. On Peano's contributions to logic and the foundations of mathematics and that of his school the best source is Borga et al. (1985), which also contains a rich bibliography. For Peano's contributions to logic and the axiomatic method, see especially Borga (1985), Grattan-Guinness (2000), and Rodriguez-Consuegra (1991). See also Quine (1987).

5. This idea of Padoa is at the root of a widespread interpretation of axiomatic systems as propositional functions, which yield specific interpreted theories when the variables are replaced by constants with a definite meaning. This view is defended in Whitehead (1907), Huntington (1913), Korselt (1913), Keyser (1918b, 1922), and Ajdukiewicz (1921). Such an interpretation also influences the development of the notion of model in Carnap and Tarski.

6. A similar result is stated which shows that the set of basic propositions of a system is irreducible, that is, that no one of them follows for the others: "To prove that the system of unproved propositions [P] is irreducible it is necessary and sufficient to find, for each of these propositions, an interpretation of the system of undefined symbols that verifies the other unproved propositions but not that one" (1901, 123).

7. See also Hilbert's lectures on geometry, Hilbert (2004).

8. On the various meanings of completeness in Hilbert, see Awodey and Reck (2002, 8–15) and Zach (1999).

9. On the debate that opposed Hilbert and Frege on this and related issues, see Demopoulos (1994).

10. Padoa later criticizes Hilbert for claiming that there might be other ways of proving the consistency of an axiom system. After Hilbert's talk in 1900, Peano claimed that Padoa's lecture would give a solution to Hilbert's second problem. Hilbert was not present at the, lecture but the only proof of consistency given by Padoa for his system of integers was by interpreting the formal system in its natural way on the domain of positive and negative integers. It is hard to believe that this led to an acrimonious article in which Padoa (1903) attacked Hilbert for not acknowledging that his second problem was only a "trifle." After a refusal to buy into the hierarchical conception of mathematics displayed by the reduction of the consistency of geometry to arithmetic, Padoa stated that Hilbert could modify at will all the methods used in the theory of irrational numbers but that this would never give him a consistency proof. Indeed, only statements of inconsistency and dependence could be solved by means of deductive reasonings, not issues of consistency or independence. According to Padoa, a consistency proof could only be obtained by displaying a specific interpretation satisfying the statements of the theory. Hilbert never replied to Padoa; in a way, the problem Padoa had raised was also a result of the vague way in which Hilbert had conjectured how it could be solved. It should be pointed out that Pieri (1904) takes position against Padoa on this issue remarking that perhaps one could find a direct proof of consistency for arithmetic by means of pure logic.

11. On the relationship between the axiom of completeness and the metalogical notion of completeness, see section 5.3.

12. We will follow, for consistency, Awodey and Reck (2002) when providing the technical definitions required in the discussion. An axiomatic theory T is called categorical (relative to a given semantics) iff all models of T are isomorphic.

13. An axiomatic theory is called semantically complete (relative to a given semantics) if any of the following four equivalent conditions hold:

1. For all formulas φ and all models M, N of T, if $M \models \varphi$, then $N \models \varphi$.
2. For all formulas φ, either $T \models \varphi$ or $T \models \neg\varphi$.
3. For all formulas φ, either $T \models \varphi$ or $T \cup \{\varphi\}$ is not satisfiable.

4. There is no formula φ such that both $T \cup \{\varphi\}$ and $T \cup \{\neg\varphi\}$ are satisfiable.

14. This idea is expressed quite clearly in Bôcher (1904, 128).

15. "Suppose we express a law by a formal sentence S, and A is a structure. Different writers have different ways of saying that the structure A obeys the law. Some say that A satisfies S, or that A is a model of S. Many writers say that the sentence S is true in the structure A. This is the notion in the title of my talk. This use of the word *true* seems to be a little over fifty years old. The earliest occurrence I find is 'wahr in N^*' in a paper of Skolem (1933) on non-standard models of arithmetic (Padoa in [1901] has 'vérifie' [p. 136])" (Hodges 1986, 136).

16. A few more examples. "The assignment of an admissible meaning, or value, to each of the undefined elements of a postulate system will be spoken of as an interpretation of the system. By 'admissible' meanings are meant meanings that satisfy the postulates or that, in other words, render them true propositions" (Keyser 1918a, 391).

"The logical structure of axiomatic geometry in Hilbert's sense—analogously to that of group theory—is a purely hypothetical one. If there are anywhere in reality three systems of objects, as well as determined relationships between these objects, such that the axioms hold of them (this means that by an appropriate assignment of names to the objects and relations the axioms turn into true statements [die Axiome in wahre Behauptungen übergehen]), then all theorems of geometry hold of these objects and relationships as well" (Bernays 1922, 192).

17. For Russell's abandonment of idealism, see Hylton (1990).

18. For recent work on reconstructing Frege's system without Axiom V, see Demopoulos (1995) and Hale and Wright (2001).

19. For an overview of the role of paradoxes in the history of logic see Cantini (2008). See the references given in section 2.1 for extensive analyses of the paradoxes.

20. For a survey of the history of predicativity, see Feferman (2004a).

21. For Poincaré on predicativity, see Heinzmann (1985).

22. See Chihara (1973), de Rouilhan (1996), and Thiel (1972) for detailed analyses of the various versions of the vicious circle principle.

23. There is even disagreement as to whether the types are linguistic or ontological entities and on the issue of whether the type distinction is superimposed on the orders or vice versa; see Landini (1998) and Linsky (1999).

24. On Russell's reasons for ramification, see also Goldfarb (1989).

25. See the extensive treatment in Grattan-Guinness (2000), and also Potter (2000) and Giaquinto (2002). Recent work has also been directed at studying the differences between the first and second edition of *Principia*; see Linsky (2004) and Hazen and Davoren (2000). The reader is also referred to the classic treatment by Gödel (1944). Hazen (2004) has pursued Gödel's suggestion that there is a new theory of types in the second edition.

26. We disagree with those who claim that metatheoretical questions could not be posed by Russell on account of his "universalistic" conception of logic. However, a detailed discussion of this issue cannot be carried out here. For this debate, see van Heijenoort (1967b), Dreben and van Heijenoort (1986), Hintikka (1988), Goldfarb (1979, 2001), de Rouilhan (1991), Tappenden (1997), Rivenc (1993).

27. On the development of set theory see, among others, Dauben (1971), Ferreiros (1999), Garciadiego (1992), Grattan-Guinness (2000), Kanamori (2003), Hallett (1984), and Moore (1982).

28. On Zermelo's role in the development of set theory and logic, see also Peckhaus (1990).

29. It should be pointed out that Russell had independently formulated a version of the axiom of choice in 1904.

30. The best treatment of the debate about the axiom of choice and related debates is Moore (1982).

31. On the antinomy see Garciadiego (1992). The antinomy is a transformation of an argument of Burali-Forti, made by Russell. If there were a set Ω of all ordinals, then it can be well ordered. Thus it is itself an ordinal, that is, it belongs and it does not belong to itself.

32. On the connection between Weyl (1910) and (1918), see Feferman (1988).

33. On Zermelo's reaction to Skolem's paradox, see van Dalen and Ebbinghaus (2000).

34. Studies on the independence of the remaining axioms of set theory were actively pursued. See for instance Fraenkel (1922a).

35. On Mirimanoff, see the extended treatment in Hallett (1984).

36. On replacement, see Hallett (1984).

37. On von Neumann's system and its extensions, see Hallett (1984) and Ferreiros (1999).

38. Zermelo investigated the metatheoretical properties of his system, especially issues of categoricity (see Hallett 1996).

39. In 1870, Peirce used the word "relative" in place of "relation" employed by De Morgan. In 1903 (367n3), Peirce called De Morgan his "master," and regretted his change of terminology.

40. To our knowledge, van Heijenoort was the first to grasp the real historical interest of Löwenheim's paper. In "Logic as Calculus and Logic as Language" (1967b) he noted the elements in Löwenheim's paper that made it a pioneering work, deserving a place in the history of logic alongside Frege's *Begriffsschrift* and Herbrand's thesis. For the history of model theory, see Mostowski (1966), Vaught (1974), Chang (1974), the historical sections of Hodges (1993), and Lascar (1998).

41. For a detailed exposition and defense of the thesis presented in this contribution, see Badesa (2004).

42. On Tarski's suggestion, McKinsey (1940) had given an axiomatization of the theory of atomic algebras of relations. The 45 years that Tarski mentions is the time elapsed between the publication of the third volume of *Vorlesungen* and McKinsey's paper. A brief historical summary of the subsequent developments can be found in Jónsson (1986) and Maddux (1991).

43. It cannot be said to be totally algebraic, given the absence of an algebraic foundation of the summands and productands that range over an infinite domain.

44. Traditionally, "logic of relatives" is used to refer to the calculus or, depending on the context, to the theory of relatives. Our use of this expression is not standard.

45. Schröder showed how to develop the logic of predicates within the logic of binary relatives in his *Vorlesungen* 1895, §27. The proof that every relative equation is logically equivalent to a relative equation in which only binary relatives occur is due to Löwenheim (1915, Theorem 6).

46. Quantifiers were introduced in the algebraic approach to logic by Peirce (1883, 464). The word *quantifier* was also introduced by him in 1885 (183).

47. Expressions of the form $A \mathrel{=\!\!\!\in} B$ (called *subsumptions*) are also used as formulas, but the canonical statements are the equations. Depending on the context,

the subsumption symbol ($\Subset$) denotes the inclusion relation, the usual ordering on $\{0, 1\}$, or the conditional. Löwenheim does not consider this symbol to belong to the basic language of the logic of relatives; this explains why he does not take it into account in the proof of his theorem.

48. In 1920 Skolem used *Zählaussage* instead of Löwenheim's *Zählausdruck*. Gödel erroneously attributes the term *Zählaussage* to Löwenheim (Gödel 1929, 61–62).

49. In fact, Skolem (1922, 294) used the term *Lösung* (solution) to refer to the assignments of truth values to the relative coefficients that satisfy a given formula in a domain.

50. He probably intended not only to simplify the proof but also to make it more rigorous, but he did not doubt its correctness. See, for example, Skolem (1920, 254; 1922, 293; and 1938, 455–456).

51. Löwenheim also generalized (1) to the case of formulas with multiple quantifiers, but this generalization is trivial. For typographical reasons, we use $\underline{\Sigma}$ in place of Löwenheim's double sigma.

52. See van Heijenoort (1967a, 230), Wang (1970, 27), Vaught (1974, 156), Goldfarb (1979, 357), and Moore (1988, 122).

53. See van Heijenoort (1967a, 229–230) and Moore (1988, 121).

54. Which the possible systems are depends on whether the fleeing indices are functional terms. More exactly, certain alternatives are only possible when fleeing indices are not functional terms. For example, a system of equalities in which $1 = 2$ and $3 \neq 4$ is not compatible with a functional interpretation of the fleeing indices, because $3 = k_1$ and $4 = k_2$. Löwenheim repeatedly insists that two different numerals can denote the same element without placing restrictions on this, but he does not explicitly clarify which systems of equalities are admissible.

55. Skolem (1929a) proved again the weak version of the theorem. In this paper, Skolem corrects some deficiencies of his previous proof in 1922 (Wang 1974, 20ff.) and introduces the functional form. As it is well known, the functional form of a formula such as $\forall x \exists y \forall z \exists u A(x, y, z, u)$ is $\forall x \forall z A(x, f(x), z, g(x, z))$. Skolem (1929a) states explicitly the informal procedure to which Gödel refers, but some of his assertions reveal that he lacks a clear understanding of the completeness problem.

56. The use of substitution is indicated at the beginning of *2. A substitution rule was explicitly included in the system of Russell (1906b), and Russell also acknowledged its necessity later (e.g., in the introduction to the second edition of *Principia*). For a discussion of the origin of the propositional calculus of *Principia* and the tacit inference rules used there, see O'Leary (1988).

57. This becomes clear from Bernays (1918), who makes a point of distinguishing between correct and provable formulas, "to avoid a circle." In Hilbert (1920a, 8), we read: "It is now the first task of logic to find those combinations of propositions, which are always, that is, without regard for the content of the basic propositions, *correct*."

58. This connection between the completeness theorem and the completeness axiom is tenuous: Hilbert's completeness axioms do not in general guarantee the categoricity of the axiom systems, nor its completeness in the sense that the system proves or disproves every statement. See Baldus (1928) for a counterexample and Awodey and Reck (2002) for more detailed discussion.

59. Note that here, as indeed in Post (1921), syntactic completeness only holds if the rule of substitution is present.

60. Post (1921) gives the same definition and establishes similar results; see section 8.3.

61. The interested reader may consult Kneale and Kneale (1962, 689–694) and, of course, Bernays (1926). The method was discovered independently by Łukasiewicz (1924), who announced results similar to those of Bernays. Bernays's first system defines Łukasiewicz's three-valued implication.

62. Gödel (1932b) quotes the independence proofs given by Hilbert (1928a).

63. These results extend the method of the previous sections insofar as the independence of rules is also proved. To do this, it is shown that an instance of the premise(s) of a rule always takes designated values, but the corresponding instance of the conclusion does not. This extension of the matrix method for proving independence was later rediscovered by Huntington (1935).

64. This is not stated explicitly, but is evident from the derivation on p. 11.

65. Paul Bernays, notes to "Mathematische Logik," lecture course held winter semester 1929–30, Universität Göttingen. Unpublished shorthand manuscript. Bernays Nachlaß, ETH Zürich Archive, Hs 973.212. The signs & and ∨ were first used as signs for conjunction and disjunction in Hilbert and Bernays (1923b). The third axiom of group I and the second axiom of group V are missing from the system given in Hilbert and Bernays (1934). The first (*Simp*), third (*Comm*), and fourth axiom (*Syll*) of group I are investigated in the published version of the *Habilitationsschrift* (Bernays 1926), but not in the original version (1918).

66. Hilbert (1905a, 249); see Zach (1999, 335–336) for discussion.

67. See Mancosu (1999a) for a discussion of this talk.

68. For extensive historical data as well as an annotated bibliography on the decision problem, both for classes of logical formulas as well as mathematical theories, see Börger et al. (1997).

69. On Curry's work, see Seldin (1980).

70. For more details on the work of Hertz and Gentzen, see Abrusci (1983) and Schröder-Heister (2002).

71. On the ε-calculus, see Hilbert and Bernays (1939) and Avigad and Zach (2002).

72. Hilbert (1920b, 39–40). Almost the same passage is found in Hilbert (1922c, 1127–1128).

73. In a letter to Hilbert dated June 27, 1905, Zermelo mentions that he is still working on a "theory of proofs" which, he writes, he is trying to extend to " 'indirect proofs', 'contradictions' and 'consistency' " (Hilbert Papers, NSUB Göttingen, Cod Ms Hilbert 447:2). Unfortunately, no further details on Zermelo's theory are available, but it seems possible that Zermelo was working on a direct consistency proof for Hilbert's axiomatic system for the arithmetic of the reals as discussed by Hilbert (1905a).

74. Hilbert developed a second approach to eliminating ε-operators from proofs around the same time, but the prospects of applying this method to arithmetic were less promising. The approach was eventually developed by Bernays and Ackermann and was the basis for the proof of the first ε-theorem in Hilbert and Bernays (1939). On this, see Zach (2004).

75. See Zach (2004) for an analysis of this proof and a discussion of its importance.

76. Von Neumann (1927) is remarkable for a few other reasons. Not only is the consistency proof carried out with more precision than those of Ackermann, but so is the formulation of the underlying logical system. For instance, the set of well-formed formulas is given a clear inductive definition, application of a function to an argument

is treated as an operation, and substitution is precisely defined. The notion of axiom system is defined in very general terms, by a rule that generates axioms (additionally, von Neumann remarks that the rules used in practice are such that it is decidable whether a given formula is an axiom). Some of these features von Neumann owes to König (1914).

77. This is problem IV in Hilbert (1929).

78. See Gödel's recollections reported by Wang (1996, 82–84).

79. On the reception of Gödel's incompleteness theorems more generally, see Dawson (1989), and Mancosu (1999b, 2004).

80. On Brouwer's life and accomplishments see van Atten (2003), van Dalen (1999), and van Stigt (1990). For an account of the foundational debate between Brouwer and Hilbert see Mancosu (1998a) and the references contained therein.

81. A good account of the French intuitionists is found in Largeault (1993a, 1993b).

82. On the Kantian themes in Brouwer's philosophy see Posy (1974) and van Atten (2003, ch. 6).

83. Troelstra (1982) gives a detailed account of the origin of the idea of choice sequences.

84. On Brouwer's intuitionistic mathematics see van Atten (2003), van Dalen (1999), Dummett (1977), Franchella (1994), van Stigt (1990), and Troelstra and van Dalen (1988).

85. Indeed, in intuitionistic mathematics one can actually prove the negation of certain valid classical principles. For instance, one can prove in intuitionistic analysis that "it is not the case that every real number is either rational or irrational." These counterexamples are called strong counterexamples, and they are consequences of mathematical principles, such as the continuity principle, that are proper to intuitionism (as opposed to other forms of constructive mathematics or classical mathematics). Brouwer gave the above-mentioned counterexample in his 1928 publication. On the continuity principle in intuitionistic analysis, see van Atten (2003, ch. 3), and on the difference between weak and strong counterexamples, see van Atten (2003, chs. 2, 4, 5).

86. The best historical account of the debates surrounding intuitionism in the 1920s is Hesseling (2003).

87. We refer the reader to Thiel (1988), Mancosu and van Stigt (1998), and Hesseling (2003) for a more detailed treatment.

88. In Mancosu (1998a, 280) it was stated by mistake that Church had committed a faux pas at this juncture.

89. We should remark that Kolmogorov (1925) rejects the principle "ex falso sequitur quolibet," which he however accepts in 1932. There is some contemporary discussion on whether the principle is intuitionistically valid. For a first introduction see van Atten (2003, 24–25).

90. Gentzen (1933a) (in collaboration with Bernays) had arrived at the same result, but Gentzen withdrew the article from publication after Gödel's paper appeared in print. The similarity between Gödel's and Gentzen's articles is striking. This parallelism can be explained by noting that both of them relied on the formalization of intuitionistic logic given by Heyting (1930a) and the axiomatization of arithmetic given by Herbrand (1931a).

91. See Mancosu (1998b) on finitism and intuitionism in the 1920s.

92. On all these contributions, see the useful introductions by Troelstra in (Gödel 1986).

93. On Łukasiewicz's logical accomplishments and the context in which he worked see Woleński (1989).

94. Słupecki, like Łukasiewicz, used the Polish notation; for the reader's benefit, we have used the *Principia* notation in this section.

95. Among the few variations one can mention "concrete representation" (Veblen and Young 1910, 3; Young 1917, 43). It should be pointed out here that although the word "model" was widespread in physics (see, e.g., "dynamical models" in Hertz 1894) it is not as common in the literature on non-Euclidean geometry, where the terminology of choice remains "interpretation" (as in Beltrami's 1868 interpretation of non-Euclidean geometry). However, "*Modelle*," that is, desktop physical models, of particular geometrical surfaces adorned the German mathematics departments of the time. Many thanks to Jamie Tappenden for useful information on this issue.

96. Following Russell, structure-theoretic terminology is found all over the epistemological landscape. See, for instance, Carnap's *Der logische Aufbau der Welt* (1928).

97. A similar approach is found in Lewis (1918, 355).

98. See Dreben and van Heijenoort (1986, 47–48) for a clarification of some delicate points in Hilbert and Ackermann's statement of the completeness problem.

99. In the 1929 dissertation, the result for countable sentences is obtained directly and not as a corollary to compactness. For the history of compactness, see Dawson (1993).

100. The notion of *allgemeingültig* can be relativized to specific types of domains. So, for instance, $(Ex)F(x) \vee (x)F(x)$ is *allgemeingültig* for those domain consisting of only one element. See Bernays and Schönfinkel (1928, 344).

101. Gödel did not provide the foregoing explanations in the published version of the thesis (1930a), but the same definition occurs in later published works (Gödel 1933b, 307), where the same idea is used to define the notion of a model over I (a domain of individuals).

102. An early case is Weyl (1910) and concerns the continuum problem. Weyl says (p. 304) that the continuum problem will not admit a solution until one adds to the system of set theory an analog of the opposite of Hilbert's completeness axiom: From the domain of Zermelo's axioms one cannot cut out a subdomain which already makes all the axioms true.

103. Nowadays we call the second notion "syntactic completeness." As the notion of categoricity as isomorphism is already found, among other places, in Bôcher (1904), Huntington (1906–1907), and Weyl (1910) (also Weyl 1927), we cannot agree with Howard (1996, 157), when he claims that Carnap (1927) is "the first place where the modern concept of categoricity, or monomorphism in Carnap's terminology, is clearly defined and its relation to issues of completeness and decidability clearly expounded. Moreover, it was through Carnap's relations with Kurt Gödel and Alfred Tarski that the concept of categoricity later made its way into formal semantics." The first conjunct is made false by the references just given, the second by the fact that Carnap's claims as to the equivalence of categoricity and decidability turned out to be unwarranted. As for Carnap's influence, it is certainly the case that Tarski was familiar with the concept of categoricity before he knew of Carnap's investigations (see Tarski 1930b, 33). Howard's article is to be recommended for exploring the relevance of the issue of categoricity for the natural sciences. On completeness and categoricity see Awodey and Carus (2001), Awodey and Reck (2002), and also Read (1997).

104. Weyl's reflection on *Entscheidungsdefinitheit* are related to the great attention given to this notion in the phenomenological literature, including Husserl, Becker, Geiger, London, and Kaufmann.

105. See the review by Rosser (1937).

106. Scanlan (2003) deals with the influence of Langford's work on Tarski. See Zygmunt (1990) on Presburger's life and work. Tarski's early results are discussed by Feferman (2004b), who uses them to reply to some points by Hodges (1986). On Tarski's quantifier elimination result for elementary algebra and geometry, see the extensive study by Sinaçeur (2007). For a treatment of the main concepts of the methodology of deductive sciences according to Tarski, see Czelakowski and Malinowski (1985) and Granger (1998).

107. One should also not forget the possible influence of Łukasiewicz; see Woleński (1994). On the Polish school see Woleński (1989, 1995).

108. For the interpretation of the differences between the original article (1933b) and the claims made in the postscript in 1935, see de Rouilhan (1998).

109. Gödel was aware of the result before Tarski published it; see the discussion in Murawski (1998). However, the author makes heavy weather of Gödel's use of the word *richtig* as opposed to *wahr*. To this it must be remarked that *richtig* is used in opposition to *falsch* throughout the writings of the Hilbert school. Moreover, Gödel himself speaks of *wahr* in his dissertation (Gödel 1929, 68–69). See also Feferman (1984).

110. On the issue of whether Tarski defines truth in a structure, see Hodges (1986) and Feferman (2004b). On logical consequence, see, among the many contributions, Etchemendy (1988, 1990), Ray (1996), Gomez-Torrente (1996), Bays (2001), and Mancosu (2006).

References

Abrusci, Vito Michele. 1983. Paul Hertz's logical works: Contents and relevance. In *Atti del Convegno Internazionale di Storia della Logica, Le teorie delle modalità. 4–8 December 1982, San Gimignano*, 369–374. Bologna: CLUEB.

Ackermann, Wilhelm. 1924. Begründung des "tertium non datur" mittels der Hilbertschen Theorie der Widerspruchsfreiheit. *Mathematische Annalen* 93: 1–36.

Ackermann, Wilhelm. 1928a. Über die Erfüllbarkeit gewisser Zählausdrücke. *Mathematische Annalen* 100: 638–649.

Ackermann, Wilhelm. 1928b. Zum Hilbertschen Aufbau der reellen Zahlen. *Mathematische Annalen* 99: 118–133.

Ackermann, Wilhelm. 1940. Zur Widerspruchsfreiheit der Zahlentheorie. *Mathematische Annalen* 117: 162–194.

Ajdukiewicz, Kazimierz. 1921. *Z metodologii nauk dedukcyjnych*. Lwów. English translation by J. Giedymin, *From the Methodology of the Deductive Sciences*, *Studia Logica*, 19, 1966, pp. 9–46.

Andrews, Peter B. 2003. Herbrand award acceptance speech. *Journal of Automated Reasoning* 31: 169–187.

Avigad, Jeremy, and Richard Zach. 2002. The epsilon calculus. *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/epsilon-calculus.

Awodey, Steve and Andre W. Carus. 2001. Carnap, completeness, and categoricity: the Gabelbarkeitssatz of 1928. *Erkenntnis* 54: 145–172.

Awodey, Steve, and Erich Reck. 2002. Completeness and categoricity. Part I: Nineteenth-century axiomatics to twentieth-century metalogic. *History and Philosophy of Logic* 23: 1–30.

Badesa, Calixto. 2004. *The Birth of Model Theory. Löwenheim's Theorem in the Frame of the Theory of Relatives*. Princeton, N.J.: Princeton University Press.

Baire, René-Louis, Émile Borel, Jacques Hadamard, and Henri Lebesgue. 1905. Cinq lettres sur la théorie des ensembles. *Bulletin de la Société Mathématique de France* 33: 261–273. English translation in Moore (1982, appendix 1).

Baldus, Richard. 1924. *Formalismus und Intuitionismus in der Mathematik*. Karlsruhe: Braun.

Baldus, Richard. 1928. Zur Axiomatik der Geometrie I: Über Hilberts Vollständigkeitsaxiom. *Mathematische Annalen* 100: 321–333.

Barzin, Marcel and Alfred Errera. 1927. Sur la logique de M. Brouwer. *Académie Royale de Belgique, Bulletin* 13: 56–71.

Bays, Timothy. 2001. On Tarski on models. *Journal of Symbolic Logic* 66: 1701–1726.

Becker, Oskar. 1927. Mathematische Existenz. *Jahrbuch für Philosophie und phänomenologische Forschung* 8: 439–809.

Behmann, Heinrich. 1922. Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem. *Mathematische Annalen* 86: 163–229.

Beltrami, Eugenio. 1868. Saggio di interpretazione della geometria non-euclidea. *Giornale di Matematiche* 6: 284–312.

Benacerraf, Paul, and Hilary Putnam, eds. 1983. *Philosophy of Mathematics*. Cambridge: Cambridge University Press, 2nd ed.

Bernays, Paul. 1918. Beiträge zur axiomatischen Behandlung des Logik-Kalküls. Habilitationsschrift, Universität Göttingen. Bernays Nachlaß, WHS, ETH Zürich Archive, Hs 973.192.

Bernays, Paul. 1922. Die Bedeutung Hilberts für die Philosophie der Mathematik. *Die Naturwissenschaften* 10: 93–99. English translation in Mancosu (1998a, 189–197).

Bernays, Paul. 1926. Axiomatische Untersuchungen des Aussagen-Kalkuls der "Principia Mathematica." *Mathematische Zeitschrift* 25: 305–320.

Bernays, Paul. 1927. Probleme der theoretischen Logik. *Unterrichtsblätter für Mathematik und Naturwissenschaften* 33: 369–377.

Bernays, Paul. 1928a. Review of: Heinrich Behmann, "Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem." *Jahrbuch für die Fortschritte der Mathematik* 48(8): 1119–1120.

Bernays, Paul. 1928b. Zusatz zu Hilberts Vortrag über "Die Grundlagen der Mathematik." *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg* 6: 88–92. English translation in van Heijenoort (1967a, 485–489).

Bernays, Paul. 1930. Die Philosophie der Mathematik und die Hilbertsche Beweistheorie. *Blätter für deutsche Philosophie* 4: 326–367. Reprinted in Bernays (1976, 17–61). English translation in Mancosu (1998a, 234–265).

Bernays, Paul. 1976. *Abhandlungen zur Philosophie der Mathematik*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Bernays, Paul and Moses Schönfinkel. 1928. Zum Entscheidungsproblem der mathematischen Logik. *Mathematische Annalen* 99: 342–372.

Beth, Evert W. 1953. On Padoa's method in the theory of definition. *Indagationes Mathematicae* 15: 330–339.

Bôcher, Maxime. 1904. The fundamental conceptions and methods of mathematics. *Bulletin of the American Mathematical Society* 10: 115–135.

Borel, Émile. 1905. Quelques remarques sur les principes de la thèorie des ensembles. *Mathematische Annalen* 60: 194–195.

Borga, Marco. 1985. La logica, il metodo assiomatico e la problematica metateorica. In Borga et al. (1985), 11–75.

Borga, Marco, Paolo Freguglia, and Dario Palladino, eds. 1985. *I Contributi Fondazionali della Scuola di Peano*. Milan: Franco Angeli.

Börger, Egon, Erich Grädel, and Yuri Gurevich. 1997. *The Classical Decision Problem*. Berlin: Springer.

Brouwer, Luitzen Egbertus Jan. 1908. De onbetrouwbaarheid der logische principes. *Tijdschrift voor Wijsbegeerte* 2: 152–158. English translation in Brouwer (1975, 107–111).

Brouwer, Luitzen Egbertus Jan. 1912a. Intuitionism and formalism. *Bulletin of the American Mathematical Society* 20: 81–96. Reprinted in Benacerraf and Putnam (1983, 77–89).

Brouwer, Luitzen Egbertus Jan. 1912b. *Intuitionisme en Formalisme*. Amsterdam: Clausen. English translation in Brouwer (1912a).

Brouwer, Luitzen Egbertus Jan. 1918. Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten. Erster Teil: Allgemeine Mengenlehre. *KNAW Verhandelingen 1e Sectie* XII(5): 1–43.

Brouwer, Luitzen Egbertus Jan. 1921. Intuitionistische Verzamelingsleer. *KNAW Verslagen* 29: 797–802. English translation in Mancosu (1998a, 23–27).

Brouwer, Luitzen Egbertus Jan. 1923a. Intuitionistische splitsing van mathematische grondbegrippen. *KNAW Verslagen* 32: 877–880. English translation in Mancosu (1998a, 286–289).

Brouwer, Luitzen Egbertus Jan. 1923b. Über die Bedeutung des Satzes vom ausgeschlossenene Dritten in der Mathematik, insbesondere in der Funktionentheorie. *Journal für die reine und angewandte Mathematik* 154: 1–7. English translation in van Heijenoort (1967a, 334–341).

Brouwer, Luitzen Egbertus Jan. 1928. Intuitionistische Betrachtungen über den Formalismus. *KNAW Proceedings* 31: 374–379. English translation in Mancosu (1998a, 40–53).

Brouwer, Luitzen Egbertus Jan. 1975. *Collected Works*, vol. 1, eds. Arend Heyting and Hans Freudenthal. Amsterdam: North-Holland.

Cantini, Andrea. 2008. Paradoxes, self-reference and truth in the 20th century. In *Handbook of the History of Logic*, vol 6. *The Emergence of Classical Logic*, eds. Dov Gabbay and John Woods. Amsterdam: Elsevier.

Carnap, Rudolf. 1927. Eigentliche und uneigentliche Begriffe. *Symposion* 1: 355–374.

Carnap, Rudolf. 1928. *Der Logische Aufbau der Welt*. Berlin: Weltkreis-Verlag.

Carnap, Rudolf. 1930. Bericht über Untersuchungen zur allgemeinen Axiomatik. *Erkenntnis* 1: 303–310.

Carnap, Rudolf. 2000. [1927–1929]. *Untersuchungen zur allgemeinen Axiomatik*, eds. Thomas Bonk and Jesus Mosterin. Darmstadt: Wissenschaftliche Buchgesellschaft.

Chang, Chen Chung. 1974. Model theory 1945–1971. In Henkin et al. (1974), 173–186.

Chihara, Charles S. 1973. *Ontology and the Vicious-Circle Principle*. Ithaca, N.Y., and London: Cornell University Press.

Church, Alonzo. 1928. On the law of the excluded middle. *Bulletin of the American Mathematical Society* 34: 75–78.

Church, Alonzo. 1932. A set of postulates for the foundation of logic. *Annals of Mathematics* 33: 346–366.

Church, Alonzo. 1933. A set of postulates for the foundation of logic (second paper). *Annals of Mathematics* 34: 839–864.

Church, Alonzo. 1936a. A note on the Entscheidungsproblem. *Journal of Symbolic Logic* 1: 40–41.

Church, Alonzo. 1936b. An unsolvable problem of elementary number theory. *American Journal of Mathematics* 58: 345–363.

Cohen, Paul J. 1966. *Set Theory and the Continuum Hypothesis*. New York: W. A. Benjamin.

Corry, Leo. 2004. *Modern Algebra and the Rise of Mathematical Structures*. Basel: Birkhäuser, 2nd ed.

Curry, Haskell B. 1929. An analysis of logical substitution. *American Journal of Mathematics* 51: 363–384.

Curry, Haskell B. 1930. Grundlagen der kombinatorischen Logik. *American Journal of Mathematics* 52: 509–536, 789–834.

Czelakowski, Janusz, and Grzegorz Malinowski. 1985. Key notions of Tarski's methodology of deductive systems. *Studia Logica* 44: 321–351.

Dauben, Joseph W. 1971. *Georg Cantor: His Mathematics and Philosophy of the Infinite*. Boston: Harvard University Press.

Dawson, John W. 1989. The reception of Gödel's incompleteness theorems. In *Gödel's Theorem in Focus*, ed. Stuart G. Shanker, 74–95. London: Routledge.

Dawson, John W. 1993. The compactness of first-order logic: From Gödel to Lindström. *History and Philosophy of Logic* 14: 15–38.

Dedekind, Richard. 1888. *Was sind und was sollen die Zahlen?* Braunschweig: Vieweg. English translation in Ewald (1996, 787–833).

Demopoulos, William. 1994. Frege, Hilbert, and the conceptual structure of model theory. *History and Philosophy of Logic* 15: 211–225.

Demopoulos, William, ed. 1995. *Frege's Philosophy of Mathematics*. Cambridge: Cambridge University Press.

de Rouilhan, Philippe. 1991. De l'universalité de la logique. In *L'âge de la science*, ed. Jacques Bouveresse, vol. 4, part I, 93–119. Paris: Jacob.

de Rouilhan, Philippe. 1996. *Russell et le Cercle de Paradoxes*. Paris: Presses Universitaires de France.

de Rouilhan, Philippe. 1998. Tarski et l'universalité de la logique. In *Le Formalisme en Question. Le Tournant des Années 30*, eds. F. Nef and D. Vernant, 85–102. Paris: Vrin.

Dreben, Burton, Peter Andrews, and Stål Aanderaa. 1963. False lemmas in Herbrand. *Bulletin of the American Methematical Society* 69: 699–706.

Dreben, Burton and Jean van Heijenoort. 1986. Introductory note to Gödel 1929, 1930 and 1930a, 44–59. Vol. 1 of Gödel (1986).

Dumitriu, Anton. 1977. *History of Logic*. Tunbridge Wells: Abacus Press.

Dummett, Michael. 1977. *Elements of Intuitionism*. Oxford: Oxford University Press. Second edition 2000.

Etchemendy, John. 1988. Tarski on truth and logical consequence. *Journal of Symbolic Logic* 53: 51–79.

Etchemendy, John. 1990. *The Concept of Logical Consequence*. Cambridge, MA: Harvard University Press.

Ewald, William Bragg, ed. 1996. *From Kant to Hilbert. A Source Book in the Foundations of Mathematics*, vol. 2. Oxford: Oxford University Press.

Feferman, Solomon. 1984. Kurt Gödel: conviction and caution. *Philosophia naturalis* 21: 546–562. Reprinted in Feferman (1998, 150–164).

Feferman, Solomon. 1988. Weyl vindicated: "Das Kontinuum" 70 years later. In *Atti del Congresso Temi e prospettive della logica e della filosofia della scienza contemporanee*. Cesena 7–10 gennaio 1987, vol. I, 59–93. Bologna: CLUEB. Reprinted in Feferman (1998, 249–283).

Feferman, Solomon. 1998. *In the Light of Logic*. New York and Oxford: Oxford University Press.

Feferman, Solomon. 2004a. Predicativity. In *The Oxford Handbook of the Philosophy of Mathematics and Logic*, ed. Stewart Shapiro, 590–624. New York: Oxford University Press.

Feferman, Solomon. 2004b. Tarski's conceptual analysis of semantical notions. In *Sémantique et épistémologie*, ed. Ali Benmakhlouf, 79–108. Casablanca and Paris: Editions Le Fennec and J. Vrin.

Ferreiros, José. 1999. *Labyrinth of Thought. A History of Set Theory and its Role in Modern Mathematics*. Basel: Birkhäuser.

Ferreiros, José. 2001. The road to modern logic—an interpretation. *Bulletin of Symbolic Logic* 7: 441–484.

Fraenkel, Adolf Abraham. 1922a. Axiomatische Begründung der transfiniten Kardinalzahlen. I. *Mathematische Zeitschrift* 13: 153–188.

Fraenkel, Adolf Abraham. 1922b. Der Begriff "definit" und die Unabhängigkeit des Auswahlaxioms. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, physikalische-mathematische Klasse*, 253–257. English translation in van Heijenoort (1967a, 253–257).

Fraenkel, Adolf Abraham. 1922c. Zu den Grundlagen der Cantor-Zermeloschen Mengenlehre. *Mathematische Annalen* 86: 230–237.

Fraenkel, Adolf Abraham. 1923. *Einleitung in die Mengenlehre*. Berlin: Springer, 2nd ed.

Fraenkel, Adolf Abraham. 1928. *Einleitung in die Mengenlehre*. Berlin: Springer, 3rd ed.

Franchella, Miriam. 1994. *L. E. J. Brouwer Pensatore Eterodosso*. Milano: Guerini.

Frege, Gottlob. 1893. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*, vol. 1. Jena: Pohle. Partial translation in Frege (1964).

Frege, Gottlob. 1903. *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*, vol. 2. Jena: Pohle. Partial translation in Frege (1964).

Frege, Gottlob. 1964. *Basic Laws of Arithmetic*. Berkeley and Los Angeles: University of California Press. Translated and with an introduction by Montgomery Furth.

Freguglia, Paolo. 1985. Il calcolo geometrico ed i fondamenti della geometria. In Borga et al. (1985), 174–236.

Garciadiego, Alejandro R. 1992. *Bertrand Russell and the Origins of Set-theoretic "Paradoxes."* Basel: Birkhäuser.

Gentzen, Gerhard. 1933a. Über das Verhältnis zwischen intuitionistischer und klassischer Logik. First published in *Archiv für mathematische Logik und Grundlagenforschung* 16 (1974), 119–132. English translation in Gentzen (1969, 53–67).

Gentzen, Gerhard. 1933b. Über die Existenz unabhängiger Axiomensysteme zu unendlichen Satzsystemen. *Mathematische Annalen* 107: 329–350. English translation in Gentzen (1969, 29–52).

Gentzen, Gerhard. 1934. Untersuchungen über das logische Schließen I–II. *Mathematische Zeitschrift* 39: 176–210, 405–431. English translation in Gentzen (1969, 68–131).

Gentzen, Gerhard. 1935. Die Widerspruchsfreiheit der reinen Zahlentheorie. Published as "Der erste Widerspruchsfreiheitsbeweis für die klassische Zahlentheorie" in *Archiv für mathematische Logik und Grundlagenforschung* 16 (1974), 97–118. English translation in Gentzen (1969, 132–213).

Gentzen, Gerhard. 1936. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen* 112: 493–565. English translation in Gentzen (1969, 132–213).

Gentzen, Gerhard. 1969. *The Collected Papers of Gerhard Gentzen*, ed. M. E. Szabo. Amsterdam: North-Holland.

Giaquinto, Marcus. 2002. *The Search for Certainty*. Oxford: Oxford University Press.

Glivenko, Valerii. 1928. Sur la logique de M. Brouwer. *Académie Royale de Belgique, Bulletin* 14: 225–228.

Glivenko, Valerii. 1929. Sur quelques points de la logique de M. Brouwer. *Académie Royale de Belgique, Bulletin* 15: 183–188. English translation in Mancosu (1998a, 301–305).

Gödel, Kurt. 1929. Über die Vollständigkeit des Logikkalküls. Dissertation, Universität Wien. Reprinted and translated in Gödel (1986, 60–101).

Gödel, Kurt. 1930a. Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik* 37: 349–360. Reprinted and translated in Gödel (1986, 102–123).

Gödel, Kurt. 1930b. Einige metamathematische Resultate über Entscheidungsdefinitheit und Widerspruchsfreiheit. *Anzeiger der Akademie der Wissenschaften in Wien* 67: 214–215. Reprinted and translated in Gödel (1986, 140–143).

Gödel, Kurt. 1931. Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I. *Monatshefte für Mathematik und Physik* 38: 173–198. Reprinted and translated in Gödel (1986, 144–195).

Gödel, Kurt. 1932a. Ein Spezialfall des Entscheidungsproblem der theoretischen Logik. *Ergebnisse eines mathematischen Kolloquiums* 2: 27–28. Reprinted and translated in Gödel (1986, 130–235).

Gödel, Kurt. 1932b. Zum intuitionistischen Aussagenkalkül. *Anzeiger der Akademie der Wissenschaften in Wien, Mathematisch-Naturwissenschaftliche Klasse* 69: 65–66. Reprinted and translated in Gödel (1986, 222–225).

Gödel, Kurt. 1933a. Eine Interpretation des intuitionistischen Aussagenkalküls. *Ergebnisse eines mathematisches Kolloquiums* 4: 39–40. Reprinted and translated in Gödel (1986, 300–303).

Gödel, Kurt. 1933b. Zum Entscheidungsproblem des logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik* 40: 433–443. Reprinted and translated in Gödel (1986, 306–327).

Gödel, Kurt. 1933c. Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines mathematisches Kolloquiums* 4: 34–38. Reprinted and translated in Gödel (1986, 286–295).

Gödel, Kurt. 1940. *The Consistency of the Axiom of Choice and of the Generalized Continuum Hypothesis*. Annals of Mathematics Studies, vol. 3. Princeton, N.J.: Princeton University Press. Reprinted in Gödel (1990, 1–101).

Gödel, Kurt. 1944. Russell's mathematical logic. In *The Philosophy of Bertrand Russell*, ed. Paul Arthur Schilpp, 125–153. La Salle, Ill.: Open Court. Reprinted in Gödel (1990, 102–141).

Gödel, Kurt. 1986. *Collected Works*, vol. 1, eds. Solomon Feferman et al. Oxford: Oxford University Press.

Gödel, Kurt. 1990. *Collected Works*, vol. 2, eds. Solomon Feferman et al. Oxford: Oxford University Press.

Gödel, Kurt. 1995. *Collected Works*, vol. 3, eds. Solomon Feferman et al. Oxford: Oxford University Press.

Goldfarb, Warren D. 1979. Logic in the twenties: The nature of the quantifier. *Journal of Symbolic Logic* 44: 351–368.

Goldfarb, Warren D. 1989. Russell's reasons for ramification. In *Rereading Russell*, eds. C. W. Savage and C. A. Anderson. Minneapolis: University of Minnesota Press.

Goldfarb, Warren D. 1993. Herbrand's error and Gödel's correction. *Modern Logic* 3: 103–118.

Goldfarb, Warren D. 2001. Frege's conception of logic. In *Future Pasts: the Analytic Tradition in Twentieth-Century Philosophy*, eds. Juliet Floyd and Sanford Shieh, 25–41. New York: Oxford University Press.

Gomez-Torrente, Mario. 1996. Tarski on logical consequence. *Notre Dame Journal of Formal Logic* 37: 125–151.

Granger, Gilles Gaston. 1998. Le problème du fondement selon Tarski. In *Le Formalisme en Question. Le Tournant des Années 30*, eds. F. Nef and D. Vernant, 37–47. Paris: Vrin.

Grattan-Guinness, Ivor. 1977. *Dear Russell—Dear Jourdain*. London: Duckworth.

Grattan-Guinness, Ivor. 2000. *The Search for Mathematical Roots 1870–1940. Logics, Set Theories and the Foundations of Mathematics from Cantor through Russell to Gödel*. Princeton, N.J.: Princeton University Press.

Guillaume, Marcel. 1994. La logique mathématique en sa jeunesse. In *Development of Mathematics 1900–1950*, ed. J.-P. Pier, 185–367. Basel: Birkhäuser.

Hale, Bob, and Crispin Wright. 2001. *The Reason's Proper Study. Essays towards a Neo-Fregean Philosophy of Mathematics*. Oxford: Oxford University Press.

Hallett, Michael. 1984. *Cantorian Set Theory and Limitation of Size*. Oxford: Oxford University Press.

Hallett, Michael. 1996. Introduction to Zermelo 1930. In Ewald (1996), 1208–1218.

Hausdorff, Felix. 1904. Der Potenzbegriff in der Mengenlehre. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 13: 569–571.

Hazen, Allen P. 2004. A "constructive" proper extension of ramified type theory (The logic of *Principia Mathematica*, second edition, appendix B). In *One Hundred Years of Russell's Paradox*, ed. Godehard Link, 449–480. New York: de Gruyter.

Hazen, Allen P. and Jennifer Davoren. 2000. Russell's 1925 logic. *Australasian Journal of Philosophy* 78: 534–566.

Heinzmann, Gerhard. 1985. *Entre Intuition et Analyse. Poincaré et le Concept de Prédicativité*. Paris: Albert Blanchard.

Henkin, Leon, John Addison, Chen Chung Chang, Dana Scott, Robert L. Vaught, and William Craig, eds. 1974. *Proceedings of the Tarski Symposium: An International Symposium held to honor Alfred Tarski on the occasion of his seventieth birthday*, vol. 25 of *Proceedings of Symposia in Pure Mathematics*. Providence, R.I.: American Mathematical Society.

Herbrand, Jaques. 1930. Recherches sur la théorie de la démonstration. Doctoral dissertation, University of Paris. English translation in Herbrand (1971, 44–202).

Herbrand, Jaques. 1931a. Sur la non-contradiction de l'arithmétique. *Journal für die Reine und Angewandte Mathematik* 166: 1–8. English translation in van Heijenoort (1967a, 618–628).

Herbrand, Jaques. 1931b. Sur le probléme fondamental de la logique mathémathique. *Sprwozdania z posiedzeń Towarzystwa Naukowego Warszawskiego, Wydział III* 24: 12–56. English translation in Herbrand (1971, 215–271).

Herbrand, Jaques. 1971. *Logical Writings*, ed. Warren D. Goldfarb. Cambridge, Mass.: Harvard University Press.

Hertz, Heinrich. 1894. *Gesammelte Werke III: Die Prinzipien der Mechanik*. Leipzig: Barth. Translated as *The Principles of Mechanics*, New York: Dover, 1956.

Hertz, Paul. 1922. Über Axiomenssysteme für beliebige Satzmengen. I. Teil. *Mathematische Annalen* 87: 246–269.

Hertz, Paul. 1923. Über Axiomensysteme für beliebige Satzmengen. II. Teil. *Mathematische Annalen* 89: 76–102.

Hertz, Paul. 1928. Reichen die üblichen syllogistischen Regeln für das Schließen in der positiven Logik elementarer Sätze aus? *Annalen der Philosophie und philosophischen Kritik* 7: 272–277.

Hertz, Paul. 1929. Über Axiomensysteme für beliebige Satzmengen. *Mathematische Annalen* 101: 457–514.

Hesseling, Dennis E. 2003. *Gnomes in the Fog. The Reception of Brouwer's Intuitionism in the 1920s*. Boston: Birkhäuser.

Heyting, Arend. 1930a. Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der Preussischen Akademie der Wissenschaften* 42–56. English translation in Mancosu (1998a, 311–327).

Heyting, Arend. 1930b. Die formalen Regeln der intuitionistischen Mathematik. *Sitzungsberichte der Preussischen Akademie der Wissenschaften* 57–71.

Heyting, Arend. 1930c. Die formalen Regeln der intuitionistischen Mathematik. *Sitzungsberichte der Preussischen Akademie der Wissenschaften* 158–169.

Heyting, Arend. 1930d. Sur la logique intuitionniste. *Académie Royale de Belgique, Bulletin* 16: 957–963. English translation in Mancosu (1998a, 306–310).

Heyting, Arend. 1931. Die intuitionistische Grundlegung der Mathematik. *Erkenntnis* 2: 106–115. English translation in Benacerraf and Putnam (1983, 52–61).

Heyting, Arend. 1934. *Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie*. Berlin: Springer. Extended French translation in Heyting (1955).

Heyting, Arend. 1955. *Les Fondements des Mathématiques. Intuitionisme, Théorie de la Démonstration*. Paris: Gauthier-Villars.

Hilbert, David. 1899. Grundlagen der Geometrie. In *Festschrift zur Feier der Enthüllung des Gauss-Weber-Denkmals in Göttingen*, 1–92. Leipzig: Teubner.

Hilbert, David. 1900a. Mathematische Probleme. *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Math.-Phys. Klasse* 253–297. Lecture given at the International Congress of Mathematicians, Paris, 1900. Partial English translation in Ewald (1996, 1096–1105).

Hilbert, David. 1900b. Über den Zahlbegriff. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 8: 180–184. English translation in Ewald (1996, 1089–1096).

Hilbert, David. 1902. *Foundations of Geometry*. Chicago: Open Court.

Hilbert, David. 1905a. Logische Principien des mathematischen Denkens. Vorlesung, Sommer-Semester 1905. Lecture notes by Ernst Hellinger. Unpublished manuscript, 277 pp. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, David. 1905b. Über die Grundlagen der Logik und der Arithmetik. In *Verhandlungen des dritten Internationalen Mathematiker-Kongresses in Heidelberg vom 8. bis 13. August 1904*, ed. A. Krazer, 174–185. Leipzig: Teubner. English translation in van Heijenoort (1967a, 129–138).

Hilbert, David. 1918a. Axiomatisches Denken. *Mathematische Annalen* 78: 405–415. Lecture given at the Swiss Society of Mathematicians, September 11, 1917. Reprinted in Hilbert (1935, 146–56). English translation in Ewald (1996, 1105–1115).

Hilbert, David. 1918b. Prinzipien der Mathematik. Lecture notes by Paul Bernays. Winter-Semester 1917–1918. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, David. 1920a. Logik-Kalkül. Vorlesung, Winter-Semester 1920. Lecture notes by Paul Bernays. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, David. 1920b. Probleme der mathematischen Logik. Vorlesung, Sommer-Semester 1920. Lecture notes by Paul Bernays and Moses Schönfinkel. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, David. 1922a. Grundlagen der Mathematik. Vorlesung, Winter-Semester 1921–1922. Lecture notes by Helmut Kneser. Unpublished manuscript, three notebooks.

Hilbert, David. 1922b. Grundlagen der Mathematik. Vorlesung, Winter-Semester 1921–1922. Lecture notes by Paul Bernays. Unpublished typescript. Bibliothek, Mathematisches Institut, Universität Göttingen.

Hilbert, David. 1922c. Neubegründung der Mathematik: Erste Mitteilung. *Abhandlungen aus dem Seminar der Hamburgischen Universität* 1: 157–177. Series of talks given at the University of Hamburg, July 25–27, 1921. Reprinted with notes by Bernays in Hilbert (1935, 157–177). English translation in Mancosu (1998a, 198–214) and Ewald (1996, 1115–1134).

Hilbert, David. 1923. Die logischen Grundlagen der Mathematik. *Mathematische Annalen* 88: 151–165. Lecture given at the Deutsche Naturforscher-Gesellschaft, September 1922. Reprinted in Hilbert (1935, 178–191). English translation in Ewald (1996, 1134–1148).

Hilbert, David. 1926. Über das Unendliche. *Mathematische Annalen* 95: 161–190. Lecture given in Münster, June 4, 1925. English translation in van Heijenoort (1967a, 367–392).

Hilbert, David. 1928a. Die Grundlagen der Mathematik. *Abhandlungen aus dem Seminar der Hamburgischen Universität* 6: 65–85. English translation in van Heijenoort (1967a, 464–479).

Hilbert, David. 1928b. Probleme der Grundlegung der Mathematik. In *Atti del Congresso Internazionale dei Matematici.* September 3–10, 1928, Bologna, ed. Nicola Zanichelli, 135–141.

Hilbert, David. 1929. Probleme der Grundlegung der Mathematik. *Mathematische Annalen* 102: 1–9. Lecture given at the International Congress of Mathematicians, September 3, 1928. English translation in Mancosu (1998a, 266–273).

Hilbert, David. 1935. *Gesammelte Abhandlungen*, vol. 3. Berlin: Springer.

Hilbert, David. 2004. *David Hilbert's Lectures on the Foundations of Geometry, 1891–1902*, eds. Ulrich Majer and Michael Hallett. New York: Springer.

Hilbert, David, and Wilhelm Ackermann. 1928. *Grundzüge der theoretischen Logik.* Berlin: Springer.

Hilbert, David, and Paul Bernays. 1923a. Logische Grundlagen der Mathematik. Winter-Semester 1922–1923. Lecture notes by Helmut Kneser. Unpublished manuscript.

Hilbert, David, and Paul Bernays. 1923b. Logische Grundlagen der Mathematik. Vorlesung, Winter-Semester 1922–1923. Lecture notes by Paul Bernays, with handwritten notes by Hilbert. Hilbert-Nachlaß, Niedersächsische Staats- und Universitätsbibliothek, Cod. Ms. Hilbert 567.

Hilbert, David, and Paul Bernays. 1934. *Grundlagen der Mathematik*, vol. 1. Berlin: Springer.

Hilbert, David, and Paul Bernays. 1939. *Grundlagen der Mathematik*, vol. 2. Berlin: Springer.

Hintikka, Jaakko. 1988. On the development of the model-theoretic viewpoint in logical theory. *Synthese* 77: 1–36.

Hodges, Wilfrid. 1986. Truth in a structure. *Proceedings of the Aristotelian Society* 86: 135–151.

Hodges, Wilfrid. 1993. *Model Theory*. Cambridge: Cambridge University Press.

Howard, Don. 1996. Relativity, *Eindeutigkeit*, and monomorphism: Rudolf Carnap and the development of the categoricity concept. In *Origins of Logical Empiricism*, eds. Ronald N. Giere and Alan W. Richardson, Minnesota Studies in the Philosophy of Science 16, 115–164. Minneapolis: University of Minnesota Press.

Huntington, Edward V. 1902. A complete set of postulates for the theory of absolute continuous magnitude. *Transactions of the American Mathematical Society* 3: 264–279.

Huntington, Edward V. 1904. Sets of independent postulates for the algebra of logic. *Transactions of the American Mathematical Society* 5: 288–309.

Huntington, Edward V. 1905. A set of postulates for real algebra, comprising postulates for a one-dimensional continuum and for the theory of groups. *Transactions of the American Mathematical Society* 6: 17–41.

Huntington, Edward V. 1906–1907. The fundamental laws of addition and multiplication in elementary algebra. *Annals of Mathematics* 8: 1–44.

Huntington, Edward V. 1911. The fundamental propositions of algebra. In *Monographs on Topics of Modern Mathematics*, ed. J. W. A. Young. New York: Longmans, Green.

Huntington, Edward V. 1913. A set of postulates for abstract geometry, expressed in terms of the simple relation of inclusion. *Mathematische Annalen* 73: 522–599.

Huntington, Edward V. 1935. The inter-deducibility of the new Hilbert-Bernays theory and *Principia Mathematica*. *Annals of Mathematics* 36: 313–324.

Hylton, Peter W. 1990. *Russell, Idealism, and the Emergence of Analytic Philosophy*. Oxford: Oxford University Press.

Jaśkowski, Stanisław. 1936. Recherches sur le système de la logique intuitioniste. In *Actes du Congrès International de Philosophie Scientifique*, vol. 4, 58–61. Paris. English translation in McCall (1967, 259–263).

Jónsson, Bjarni. 1986. The contributions of Alfred Tarski to general algebra. *Journal of Symbolic Logic* 51: 883–889.

Jourdain, Philip. 1914. Preface. Luis Couturat, *The Algebra of Logic*. Chicago and London: Open Court, iii–xiii.

Kalmár, László. 1933. Über die Erfüllbarkeit derjenigen Zählausdrücke, welche in der Normalform zwei benachbarte Allzeichen enthalten. *Mathematische Annalen* 108: 466–484.

Kanamori, Akihiro. 2003. *The Higher Infinite. Large Cardinals in Set Theory from Their Beginnings*. Berlin: Springer, 2nd ed.

Kaufmann, Felix. 1930. *Das Unendliche in der Mathematik und seine Ausschaltung*. Vienna: Deuticke.

Keyser, Cassius J. 1918a. Concerning the number of possible interpretations of any axiom system of postulates. *Bulletin of the American Mathematical Society* 24: 391–393.

Keyser, Cassius J. 1918b. Doctrinal functions. *Journal of Philosophy* 15: 262–267.

Keyser, Cassius J. 1922. *Mathematical Philosophy. A Study of Fate and Freedom*. New York: Dutton.

Kleene, Stephen C. 1935. A theory of positive integers in formal logic. *American Journal of Mathematics* 57: 219–244.

Kleene, Stephen C., and J. Barkley Rosser. 1935. The inconsistency of certain formal logics. *Annals of Mathematics* 36: 630–636.

Kneale, William, and Martha Kneale. 1962. *The Development of Logic*. Oxford: Oxford University Press.

Kolmogorov, Andrei N. 1925. O principe tertium non datur. *Matematiceskij Sbornik* 32: 646–667. English translation in van Heijenoort (1967a, 416–437).

Kolmogorov, Andrei N. 1932. Zur Deutung der intuitionistischen Logik. *Mathematische Zeitschrift* 35: 58–65. English translation in Mancosu (1998a, 328–334).

König, Dénes. 1926. Sur les correspondances multivoques des ensembles. *Fundamenta Mathematicae* 8: 114–134.

König, Dénes. 1927. Über eine Schlussweise aus dem Endlichen ins Unendliche. *Acta litterarum ac scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae, Sectio scientiarum mathematicarum* 3: 121–130.

König, Julius. 1904. Zum Kontinuum-Problem. In *Verhandlungen des dritten internationalen Mathematiker-Kongresses in Heidelberg*, 144–147. Leipzig: Teubner.

König, Julius. 1914. *Neue Grundlagen der Logik, Arithmetik und Mengenlehre*. Leipzig: Veit.

Korselt, A. 1913. Was ist Mathematik? *Archiv der Mathematik und Physik* (*Series 3*) 21: 371–373.

Landini, Gregory. 1998. *Russell's Hidden Substitutional Theory*. Oxford: Oxford University Press.

Langford, Cooper Harold. 1927a. Some theorems on deducibility. *Annals of Mathematics* (Series 2) 28: 16–40.

Langford, Cooper Harold. 1927b. Some theorems on deducibility. *Annals of Mathematics* (Series 2) 28: 459–471.

Largeault, Jean. 1993a. *Intuition et Intuitionisme*. Paris: Vrin.

Largeault, Jean. 1993b. L'intuitionisme des mathématiciens avant Brouwer. *Archives de Philosophie* 56: 53–68.

Lascar, Daniel. 1998. Perspective historique sur les rapports entre la théorie des modèles et l'algebrè. *Revue d' Histoire des Mathématiques* 4: 237–260.

Lewis, Clarence Irving. 1918. *A Survey of Symbolic Logic*. Berkeley: University of California Press.

Linsky, Bernard. 1999. *Russell's Metaphysical Logic*. Stanford: CSLI Publications.

Linsky, Bernard. 2004. Leon Chwistek on the no-classes theory in *Principia Mathematica*. *History and Philosophy of Logic* 25: 53–71.

Lovett, Edgar O. 1900–1901. Mathematics at the International Congress of Philosophy, Paris, 1900. *Bulletin of the American Mathematical Society* 7: 157–183.

Löwenheim, Leopold. 1915. Über Möglichkeiten im Relativkalkül. *Mathematische Annalen* 447–470. Translation: van Heijenoort (1967a, 228–251).

Łukasiewicz, Jan. 1920a. O logice trówartościowej (On three-valued logic). *Ruch Filozoficzny* 6: 170–171. English translation in McCall (1967, 16–17).

Łukasiewicz, Jan. 1920b. O pojęciu możliwości (On the concept of possibility). *Ruch Filozoficzny* 6: 169–170. English translation in McCall (1967, 15–16).

Łukasiewicz, Jan. 1922. O determiniźmie (On determinism). English translation in McCall (1967, 19–39).

Łukasiewicz, Jan. 1924. Démonstration de la compatibilité des axiomes de la théorie de la déduction (abstract). *Annales de la Société Polonaise de Mathématique* 3: 149. Talk given June 13, 1924.

Łukasiewicz, Jan. 1930. Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls. *Comptes Rendus des Sèancs de la Société des Sciences et des Lettres de Varsovie. Classe III* 23: 51–77. English translation in McCall (1967, 40–65).

Łukasiewicz, Jan and Alfred Tarski. 1930. Untersuchungen über den Aussagenkalkül. *Comptes Rendus des Sèancs de la Société des Sciences et des Lettres de Varsovie. Classe III* 23: 30–50. English translation in Tarski (1983, 38–59).

Maddux, Roger D. 1991. The origin of relation algebras in the development and axiomatization of the calculus of relations. *Studia Logica* 50: 421–455.

Malcev, Anatoly Ivanovich. 1936. Untersuchungen aus dem Gebiet der mathematischen Logik. *Matematiceskij Sbornik* (n.s.) 1: 323–336. English translation in Anatoly I Maltsev, *The Metamathematics of Algebraic Systems: Collected Papers 1936–1967*, Amsterdam: North Holland, 1971, 1–14.

Mancosu, Paolo, ed. 1998a. *From Brouwer to Hilbert. The Debate on the Foundations of Mathematics in the 1920s.* Oxford: Oxford University Press.

Mancosu, Paolo. 1998b. Hilbert and Bernays on metamathematics. In Mancosu (1998a), 149–188.

Mancosu, Paolo. 1999a. Between Russell and Hilbert: Behmann on the foundations of mathematics. *Bulletin of Symbolic Logic* 5(3): 303–330.

Mancosu, Paolo. 1999b. Between Vienna and Berlin: The immediate reception of Gödel's incompleteness theorems. *History and Philosophy of Logic* 20: 33–45.

Mancosu, Paolo. 2002. On the constructivity of proofs. A debate among Behmann, Bernays, Gödel, and Kaufmann. In *Reflections on the Foundations of Mathematics. Essays in Honor of Solomon Feferman*, eds. Wilfried Sieg, Richard Sommer, and Carolyn Talcott, Lecture Notes in Logic 15, 346–368. Association for Symbolic Logic.

Mancosu, Paolo. 2003. The Russellian influence on Hilbert and his school. *Synthese* 137: 59–101.

Mancosu, Paolo. 2004. Review of Kurt Gödel, *Collected Works*, vols. IV and V, Solomon Feferman, et al., eds. Oxford: Oxford University Press, 2003. *Notre Dame Journal of Formal Logic* 45: 109–125.

Mancosu, Paolo. 2006. Tarski on models and logical consequence. In *The Architecture of Modern Mathematics*, eds. Jeremy Gray and José Ferreiros, 209–237. Oxford: Oxford University Press.

Mancosu, Paolo and Walter P. van Stigt. 1998. Intuitionistic logic. In Mancosu (1998a), 275–285.

Mangione, C. and S. Bozzi. 1993. *Storia della Logica.* Milano: Garzanti.

Marchisotto, Elena A. 1995. In the shadow of giants: The work of Mario Pieri in the foundations of mathematics. *History and Philosophy of Logic* 16: 107–119.

McCall, Storrs. 1967. *Polish Logic 1920–1939.* Oxford: Oxford University Press.

McKinsey, John Charles Chenoweth. 1940. Postulates for the calculus of binary relations. *Journal of Symbolic Logic* 5: 85–97.

Mirimanoff, Dmitry. 1917. Les antinomies de Russell et de Burali-Forti et le problème fondamental de la théorie des ensembles. *L'Enseignement Mathématique* 19: 37–52.

Moore, Gregory H. 1980. Beyond first-order logic: The historical interplay between mathematical logic and axiomatic set theory. *History and Philosophy of Logic* 1: 95–137.

Moore, Gregory H. 1982. *Zermelo's Axiom of Choice. Its Origins, Development and Influence.* New York and Heidelberg: Springer.

Moore, Gregory H. 1988. The emergence of first-order logic. In *History and Philosophy of Modern Mathematics*, ed. William Aspray and Philip Kitcher, Minnesota Studies in the Philosophy of Science 11, 95–135. Minneapolis: University of Minnesota Press.

Moore, Gregory H. 1994. Editor's introduction. In Russell (1994a).

Moore, Gregory H., and Alejandro R. Garciadiego. 1981. Burali-Forti's paradox: A reappraisal of its origins. *Historia Mathematica* 8: 319–350.

Mostowski, Andrzej. 1966. *Thirty Years of Fundational Studies.* Oxford: Basil Blackwell.

Murawski, Roman. 1998. Undefinability of truth. The problem of priority: Tarski vs Gödel. *History and Philosophy of Logic* 19: 153–160.

Nicod, Jean G. P. 1916–1919. A reduction in the number of the primitive propositions of logic. *Proceedings of the Cambridge Philosophical Society* 19: 32–41.

O'Leary, Daniel J. 1988. The propositional logic of *Principia Mathematica* and some of its forerunners. *Russell* 8: 92–115.

Padoa, Alessandro. 1901. Essai d'une théorie algébrique des nombre entiers, precede d'une introduction logique à une théorie deductive quelconque. In *Bibliothèque du Congrès international de philosophie, Paris 1900*, 309–365. Paris: A. Colin. Partially translated in van Heijenoort (1967a, 118–123).

Padoa, Alessandro. 1902. Un nouveau système irréductible de postulats pour l'algèbre. In *Compte rendu du deuxième congrès international des mathematicians tenu à Paris du 6 au 12 août 1900*, 249–256. Paris: Gauthiers-Villars.

Padoa, Alessandro. 1903. Le problème no 2 de M. David Hilbert. *L'Enseignement Mathématique* 5: 85–91.

Peano, Giuseppe. 1889. *I Principii di Geometria Logicamente Esposti.* Torino: Fratelli Bocca. Reprinted in Peano (1958, 56–91).

Peano, Giuseppe. 1958. *Opere Scelte*, vol. 2. Roma: Edizioni Cremonese.

Peckhaus, Volker. 1990. *Hilbertprogramm und Kritische Philosophie*. Göttingen: Vandenhoeck und Ruprecht.

Peckhaus, Volker. 1992. Hilbert, Zermelo und die Institutionalisierung der mathematischen Logik in Deutschland. *Berichte zur Wissenschaftsgeschichte* 15: 27–38.

Peckhaus, Volker. 1994. Hilbert's axiomatic programme and philosophy. In *The History of Modern Mathematics*, eds. E. Knobloch and D. E. Rowe, vol. 3, 91–112. Boston: Academic Press.

Peckhaus, Volker. 1995. Hilberts Logik: Von der Axiomatik zur Beweistheorie. *Internationale Zeitschrift für Geschichte und Ethik der Naturwissenschaften, Technik und Medizin* 3: 65–86.

Peirce, Charles Sanders. 1870. Description of a notation for the logic of relatives, resulting from an amplification of the conceptions of Boole's calculus on logic. *Memoirs of the American Academy of Arts and Sciences* 9: 317–378. Reprinted in Peirce (1984, 359–429).

Peirce, Charles Sanders. 1883. The logic of relatives. In *Studies in Logic by Members of the Johns Hopkins University*, 187–203. Philadelphia: John Benjamins. Reprinted in Peirce (1986, 453–466).

Peirce, Charles Sanders. 1885. On the algebra of logic: A contribution to the philosophy of notation. *American Journal of Mathematics* 7: 180–202. Reprinted in Peirce (1993, 162–190).

Peirce, Charles Sanders. 1903. Nomenclature and divisions of dyadic relations. In *Collected Papers*, eds. C. Hartshorne and P. Weiss. 1933, vol. 3, 366–387. Cambridge, Mass.: Harvard University Press.

Peirce, Charles Sanders. 1984. *Writings of Charles S. Peirce. Vol. 2: 1867–1871*, ed. C. J. W. Kloesel. Bloomington: Indiana University Press.

Peirce, Charles Sanders. 1986. *Writings of Charles S. Peirce. Vol. 4: 1879–1884*, ed. C. J. W. Kloesel. Bloomington: Indiana University Press.

Peirce, Charles Sanders. 1993. *Writings of Charles S. Peirce. Vol. 5: 1884–1886*, ed. C. J. W. Kloesel. Bloomington: Indiana University Press.

Pieri, Mario. 1901. Sur la géométrie envisagée comme un système purement logique. In *Bibliothèque du Congrès international de philosophie, Paris 1900*, 367–404. Paris: A. Colin. Reprinted in Pieri (1980, 235–272).

Pieri, Mario. 1904. Circa il teorema fondamentale di Staudt e i principi della geometria projettiva. *Atti della Reale academia delle scienze di Torino* 39: 313–331. Reprinted in Pieri (1980, 289–307).

Pieri, Mario. 1980. *Opere Scelte I: Opere sui fondamenti della matematica*. Bologna: UMI.

Poincaré, Henri. 1906. Les mathématiques et la logique. *Revue de métaphysique et de morale* 14: 294–317. English translation in Ewald (1996, 1038–1052).

Post, Emil L. 1921. Introduction to a general theory of elementary propositions. *American Journal of Mathematics* 43: 163–185. Reprinted in van Heijenoort (1967a, 264–283).

Posy, Carl J. 1974. Brouwer's constructivism. *Synthese* 27: 125–159.

Potter, Michael. 2000. *Reason's Nearest Kin*. Oxford: Oxford University Press.

Presburger, Mojėsz. 1930. Über die Vollständigkeit eines gewisse Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt. In *Comptes-rendus du I Congrès des Mathématiciens des Pays Slaves, Varsovie 1929*, ed. F. Leja, 92–101. Translated as "On the completeness of a certain system of arithmetic of whole numbers in which addition occurs as the only operation," *History and Philosophy of Logic* 12 (1991): 225–233.

Quine, Willard Van Orman. 1955a. On Frege's way out. *Mind* 64: 145–159.

Quine, Willard Van Orman. 1955b. A proof procedure for quantification theory. *Journal of Symbolic Logic* 20: 141–149.

Quine, Willard Van Orman. 1972. *The Methods of Logic*. New York: Holt, Rinehart & Winston, 3rd ed.

Quine, Willard Van Orman. 1987. Peano as logician. *History and Philosophy of Logic* 8: 15–24.

Ramsey, Frank P. 1925. The foundations of mathematics. *Proceedings of the London Mathematical Society* 25: 338–384. Reprinted in Ramsey (1990, 164–224).

Ramsey, Frank P. 1930. On a problem of formal logic. *Proceedings of the London Mathematical Society, 2nd Series* 30: 264–286.

Ramsey, Frank P. 1990. *Philosophical Papers*, ed. D. H. Mellor. Cambridge: Cambridge University Press.

Ray, Greg. 1996. Logical consequence: a defence of Tarski. *Journal of Philosophical Logic* 25: 617–677.

Read, Stephen. 1997. Completeness and categoricity: Frege, Gödel and model theory. *History and Philosophy of Logic* 18: 79–93.

Richard, Jules. 1905. Les principes des mathématiques et le problème des ensembles. *Revue Générale des Sciences Pures et Appliquées* 16: 541. Reprinted as "Lettre a Monsieur le rédacteur de la revue Générale des Sciences," *Acta Mathematica* 30 (1906): 295–296. English translation in van Heijenoort (1967a, 142–144).

Rivenc, F. 1993. *Recherches sur l'Universalisme Logique. Russell et Carnap*. Paris: Payot.

Rodriguez-Consuegra, Francisco A. 1991. *The Mathematical Philosophy of Bertrand Russell: Origins and Development*. Basel: Birkhäuser.

Rose, Alan and J. Barkley Rosser. 1958. Fragments of many-valued statement calculi. *Transactions of the American Mathematical Society* 87: 1–53.

Rosser, J. Barkley. 1936. Extensions of some theorems of Gödel and Church. *Journal of Symbolic Logic* 1: 87–91.

Rosser, J. Barkley. 1937. Review of Malcev (1936). *Journal of Symbolic Logic* 2: 84.

Russell, Bertrand. 1901a. Recent work on the principles of mathematics. *International Monthly* 4: 83–101. Reprinted in Russell (1994a, 363–369).

Russell, Bertrand. 1901b. Sur la logique des relations avec des applications à la théorie des series. *Revue des Mathématiques* 7: 115–148. Reprinted in Russell (1994a, 618–627).

Russell, Bertrand. 1902a. Letter to Frege, June 16, 1902. In van Heijenoort (1967a), 124–125.

Russell, Bertrand. 1902b. Théorie générale des series bien-ordonnées. *Revue des Mathématiques* 8: 12–43. Reprinted in Russell (1994a, 661–673).

Russell, Bertrand. 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press.

Russell, Bertrand. 1905. On denoting. *Mind* 14: 479–493. Reprinted in Russell (1994b, 414–427).

Russell, Bertrand. 1906a. Les paradoxes de la logique. *Revue de metaphysique et de morale* 14: 627–650.

Russell, Bertrand. 1906b. On some difficulties in the theory of transfinite numbers and order types. *Proceedings of the London Mathematical Society* 4: 29–53.

Russell, Bertrand. 1908. Mathematical logic as based on the theory of types. *Americal Journal of Mathematics* 30: 222–262. Reprinted in van Heijenoort (1967a, 150–182).

Russell, Bertrand. 1919. *Introduction to Mathematical Philosophy*. London: Allen and Unwin.

Russell, Bertrand. 1967. *The Autobiography of Bertrand Russell*, vol. 1. London: Allen and Unwin.

Russell, Bertrand. 1973. *Essays in Analysis*, ed. D. P. Lackey. London: Allen and Unwin.

Russell, Bertrand. 1989. My mental development. In *The Philosophy of Bertrand Russell*, ed. Paul Arthul Schilpp, 3–20. La Salle, Ill.: Open Court.

Russell, Bertrand. 1994a. *Collected Works*, vol. 3. *Toward the "Principles of Mathematics," 1900–02*, ed. Gregory H. Moore. London and New York: Routledge.

Russell, Bertrand. 1994b. *Collected Works*, vol. 4. *Foundations of Logic, 1903–05*, ed. Alasdair Urquhart. London and New York: Routledge.

Scanlan, Michael. 1991. Who were the American postulate theorists? *Journal of Symbolic Logic* 56: 981–1002.

Scanlan, Michael. 2003. American postulate theorists and Alfred Tarski. *History and Philosophy of Logic* 24: 307–325.

Schönfinkel, Moses. 1922. Zum Entscheidungsproblem der mathematischen Logik. Manuscript, 21 pp. Bernays Nachlaß, WHS, ETH Zürich Archive, Hs. 974.282.

Schönfinkel, Moses. 1924. Über die Bausteine der mathematischen Logik. *Mathematische Annalen* 92: 305–316. English translation in van Heijenoort (1967a, 355–366).

Schröder, Ernst. 1890. *Vorlesungen über die Algebra der Logik (exakte Logik)*, vol. 1. Reprinted 1966, New York: Chelsea.

Schröder, Ernst. 1891. *Vorlesungen über die Algebra der Logik*, vol. 2. Leipzig: Teubner. Reprinted 1966, New York: Chelsea.

Schröder, Ernst. 1895. *Vorlesungen über die Algebra der Logik*, vol. 3. Leipzig: Teubner. Reprinted 1966, New York: Chelsea.

Schröder, Ernst. 1898. On pasigraphy. Its presents state and the pasigraphic movement in Italy. *Monist* 9: 246–262, 320.

Schröder-Heister, Peter. 2002. Resolution and the origins of structural reasoning: early proof-theoretic ideas of Hertz and Gentzen. *Bulletin of Symbolic Logic* 8: 246–265.

Schütte, Kurt. 1934a. Über die Erfüllbarkeit einer Klasse von logischen Formeln. *Mathematische Annalen* 110: 572–603.

Schütte, Kurt. 1934b. Untersuchungen zum Entscheidungsproblem der mathematischen Logik. *Mathematische Annalen* 109: 572–603.

Seldin, Jonathan P. 1980. Curry's program. In *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, eds. J. P. Seldin and J. R. Hindley, 3–33. London: Academic Press.

Sheffer, Henry Maurice. 1913. A set of five independent postulates for Boolean algebras, with application to logical constants. *Transactions of the American Mathematical Society* 14: 481–488.

Sieg, Wilfried. 1999. Hilbert's programs: 1917–1922. *Bulletin of Symbolic Logic* 5(1): 1–44.

Sierpinski, Wacław. 1918. L'axiome de M. Zermelo et son rôle dans la théorie des ensembles. *Bulletin de l'Académie des Sciences de Cracovie, Série A* 99–152.

Sinaçeur, Hourya B. 2007. *Fields and Models. From Sturm to Tarski and Robinson.* Basel: Birkhäuser.

Skolem, Thoralf. 1912. Review of Weyl (1910). *Jahrbuch für die Fortschritte der mathematik* 41(1): 89–90.

Skolem, Thoralf. 1919. Untersuchungen über die Axiome des Klassenkalküls und über Produktations- und Summationsprobleme, welche gewisse Klassen von Aussagen betreffen. *Videnskasselskapets skrifter, I. Matematisk-naturvidenskabelig klasse* 3. Reprinted in Skolem (1970, 66–101).

Skolem, Thoralf. 1920. Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen. *Videnskasselskapets skrifter, I. Matematisk-naturvidenskabelig klasse* 4. Reprinted in Skolem (1970, 103–136). English translation of §1 in van Heijenoort (1967a, 252–263).

Skolem, Thoralf. 1922. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. In *Matematikerkongressen I Helsingfors*, 217–232. Helsinki: Akademiska Bokhandeln. Reprinted in Skolem (1970, 137–152). English translation in van Heijenoort (1967a, 290–301).

Skolem, Thoralf. 1928. Über die mathematische Logik. *Norsk Mathematisk Tidsskrift* 106: 125–142. Reprinted in Skolem (1970, 189–206).

Skolem, Thoralf. 1929a. Über einige Grundlagenfragen der Mathematik. *Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, I. Mathematisk-naturvidenskapelig klasse* 4. Reprinted in Skolem (1970, 227–273).

Skolem, Thoralf. 1929b. Über einige Satzfunktionen in der Arithmetik. *Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, I. Mathematisk-naturvidenskapelig klasse* 7: 1–28. Reprinted in Skolem (1970, 281–306).

Skolem, Thoralf. 1933. Über die Unmöglichkeit einer vollständigen Charakterisierung der Zahlenreihe mittels eines endlichen Axiomensystems. *Norsk matematisk forenings skrifter*, series 2 10: 73–82. Reprinted in Skolem (1970, 345–354).

Skolem, Thoralf. 1934. Über die Nicht-Charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschließlich Zahlenvariablen. *Fundamenta Mathematicae* 23: 150–161. Reprinted in Skolem (1970, 355–366).

Skolem, Thoralf. 1935. Über die Erfüllbarkeit gewisser Zählausdrücke. *Skrifter utgitt av Det Norske Videnskaps-Akademi i Oslo, I. Mathematisk-naturvidenskapelig klasse* 6: 1–12. Reprinted in Skolem (1970, 383–394).

Skolem, Thoralf. 1938. Sur la portée du théorème de Löwenheim-Skolem. In *Les Entretiens de Zurich sur les Fondaments et la Méthode des Sciences Mathematiques.* 6–9 December 1938, ed. Ferdinand Gonseth. Zurich: Leemann. Reprinted in Skolem (1970, 455–482).

Skolem, Thoralf. 1970. *Selected Works in Logic*, ed. Jens E. Fenstad. Oslo: Universitetsforlaget.

Słupecki, Jerzy. 1936. Der volle dreiwertige Aussagenkalkül. *Comptes Rendus des Sèancs de la Sociéte des Sciences et des Lettres de Varsovie, Classe III* 29: 9–11. English translation in McCall (1967, 335–337).

Tappenden, Jamie. 1997. Metatheory and mathematical practice in Frege. *Philosophical Topics* 25: 213–264.

Tarski, Alfred. 1930a. Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften, I. *Monatshefte für Mathematik und Physik* 37: 361–404. English translation with revisions in Tarski (1983, 60–109).

Tarski, Alfred. 1930b. Über einige fundamentale Begriffe der Metamathematik. *Sprawozdania z posiedzen Towarzystwa Naukowego Warszwskiego, wydzial III* 23: 22–29. English translation with revisions in Tarski (1983, 30–37).

Tarski, Alfred. 1931. Sur les ensembles définissables de nombres réels, I. *Fundamenta Mathematicae* 17: 210–239. English translation with revisions in Tarski (1983, 110–142).

Tarski, Alfred. 1933a. Einige Betrachtungen über die Begriffe der ω-Widerspruchsfreiheit und der ω-Vollständigkeit. *Monatshefte für Mathematik und Physik* 40: 97–112. English translation with revisions in Tarski (1983, 279–295).

Tarski, Alfred. 1933b. Pojęcie prawdy w językach nauck dedukcyjnych (The concept of truth in the languages of deductives sciences). *Prace Towarzystwa Naukowego Warszawskiego, wydzial III* 34. German translation in Tarski (1935). English translation in Tarski (1983, 152–278).

Tarski, Alfred. 1934–1935. Einige methodologische Untersuchungen über die Definierbarkeit der Begriffe. *Erkenntnis* 5: 80–100. English translation in Tarski (1983, 296–319).

Tarski, Alfred. 1935. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica (Lemberg)* 1: 261–405. English translation in Tarski (1983, 152–278).

Tarski, Alfred. 1936a. Grundzüge des Systemenkalküls. Zweiter Teil. *Fundamenta Mathematicae* 26: 283–301. English translation in Tarski (1983, 342–383).

Tarski, Alfred. 1936b. Über den Begriff der logischen Folgerung. In *Actes du Congrès International de Philosophie Scientifique 7, Actualités Scientifiques et Industrielles*, 1–11. Paris: Herman. English translation with revisions in Tarski (1983, 409–420).

Tarski, Alfred. 1941. On the calculus of relations. *Journal of Symbolic Logic* 6: 73–89.

Tarski, Alfred. 1948. *A Decision Method for Elementary Algebra and Geometry.* Berkeley: University of California Press. Second, rev. ed. 1951.

Tarski, Alfred. 1983. *Logic, Semantics, Metamathematics*, ed. John Corcoran, 2nd ed. Indianapolis: Hackett.

Thiel, Christian. 1972. *Grundlagenkrise und Grundlagenstreit.* Meisenheim: Anton Hain.

Thiel, Christian. 1988. Die Kontroverse um die intuitionistische Logik vor ihrer Axiomatisierung durch Heyting im Jahre 1930. *History and Philosophy of Logic* 9: 67–75.

Thiel, Christian. 1994. Schröders zweiter Beweis für die Unabhängigkeit der zweiten Subsumtion des Distributivgesetzes im logischen Kalkül. *Modern Logic* 4: 382–391.

Toepell, Michael-Markus. 1986. *Über die Entstehung von Hilbert's "Grundlagen der Geometrie."* Göttingen: Vandenhoek und Ruprecht.

Troelstra, Anne S. 1982. On the origin and development of Brouwer's concept of choice sequence. In *The L. E. J. Brouwer Centenary Symposium*, eds. Anne S. Troelstra and Dirk van Dalen, 465–486. Amsterdam: North-Holland.

Troelstra, Anne S. 1990. On the early history of intuitionistic logic. In *Mathematical Logic*, ed. Petio P. Petkov, 3–17. New York and London: Plenum Press.

Troelstra, Anne S. and Dirk van Dalen. 1988. *Constructivism in Mathematics. An Introduction*, vol. 1. Amsterdam: North-Holland.

Turing, Alan M. 1937. On computable numbers, with an application to the "Entscheidungsproblem." *Proceedings of the London Mathematical Society*, 2nd Series 42: 230–265.

van Atten, Mark. 2003. *On Brouwer*. Belmont: Wadsworth.

van Atten, Mark. 2005. The Becker-Heyting correspondence. In *Die Philosophie und die Mathematik: Oskar Becker in der mathematischen Grundlagendiskussion*, ed. Volker Peckhaus, 119–142. Munich: Wilhelm Fink Verlag.

van Dalen, Dirk. 1999. *Mystic, Geometer, Intuitionist. The Life of L. E. J. Brouwer.* Oxford: Oxford University Press.

van Dalen, Dirk and Hans-Dieter Ebbinghaus. 2000. Zermelo and the Skolem paradox. *Bulletin of Symbolic Logic* 6: 145–161.

van Heijenoort, Jean, ed. 1967a. *From Frege to Gödel. A Source Book in Mathematical Logic, 1897–1931*. Cambridge, Mass.: Harvard University Press.

van Heijenoort, Jean. 1967b. Logic as calculus and logic as language. *Boston Studies in the Philosophy of Science* 3: 440–446.

van Stigt, Walter P. 1990. *Brouwer's Intuitionism*. Amsterdam: North-Holland.

Vaught, Robert L. 1974. Model theory before 1945. In Henkin et al. (1974), 153–172.

Veblen, Oswald. 1904. A system of axioms for geometry. *Transactions of the American Mathematical Society* 5: 343–384.

Veblen, Oswald, and John Wesley Young. 1910. *Projective Geometry*. Boston: Ginn.

Vercelloni, Luca. 1988. *Filosofia delle Strutture*. Florence: La Nuova Italia.

von Neumann, Johann. 1923. Zur Einführung der transfiniten Zahlen. *Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae* 1: 199–208. English translation in van Heijenoort (1967a, 346–354).

von Neumann, Johann. 1925. Eine Axiomatisierung der Mengenlehre. *Journal für die reine und angewandte Mathematik* 154: 219–240. English translation in van Heijenoort (1967a, 393–413).

von Neumann, Johann. 1927. Zur Hilbertschen Beweistheorie. *Mathematische Zeitschrift* 26: 1–46.

von Neumann, Johann. 1928. Die Axiomatisierung der Mengenlehre. *Mathematische Zeitschrift* 27: 669–752.

Vuillemin, Jules. 1968. *Leçons sur la première philosophie de Russell*. Paris: Colin.

Wajsberg, Mordechaj. 1931. Aksjomatyzacja trjwartościowego rachnuku zdań (Axiomatization of the three-valued propositional calculus). *Comptes Rendus des Sèances de la Société des Sciences et des Lettres de Varsovie, Classe III* 24: 126–145.

Wang, Hao. 1967. Introductory note to Kolmogorov (1925). In van Heijenoort (1967a), 414–416.

Wang, Hao. 1970. A survey of Skolem's work in logic. In Skolem (1970), 17–52.

Wang, Hao. 1974. *From Mathematics to Philosophy*. London: Routledge & Kegan Paul.

Wang, Hao. 1996. *A Logical Journey*. Cambridge, Mass.: MIT Press.

Wavre, Rolin. 1926. Logique formelle et logique empiriste. *Revue de Métaphysique et de Morale* 33: 65–75.

Webb, Judson. 1995. Tracking contradictions in geometry: The idea of model from Kant to Hilbert. In *Essays on the Development of the Foundations of Mathematics*, ed. Jaakko Hintikka, 1–20. Dordrecht: Kluwer.

Weyl, Hermann. 1910. Über die Definitionen der mathematischen Grundbegriffe. *Mathematisch-naturwissenschaftliche Blätter* 7: 93–95, 109–113. Reprinted in Weyl (1968, 298–304).

Weyl, Hermann. 1918. *Das Kontinuum*. Leipzig: Veit. English translation: Weyl (1994).

Weyl, Hermann. 1919. Der circulus vitiosus in der heutigen Begründung der Analysis. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 28: 85–92. Reprinted in Weyl (1968, 43–50). English Translation in Weyl (1994).

Weyl, Hermann. 1921. Über die neue Grundlagenkrise der Mathematik. *Mathematische Zeitschrift* 10: 37–79. Reprinted in Weyl (1968, 143–180). English translation in Mancosu (1998a, 86–118).

Weyl, Hermann. 1927. *Philosophie der Mathematik und Naturwissenschaft*. Munich: Oldenbourg. Augmented and revised English translation by Olaf Helmer, *Philoso-*

phy of Mathematics and Natural Science (Princeton, N.J.: Princeton University Press, 1949).

Weyl, Hermann. 1968. *Gesammelte Abhandlungen*, vol. 1, ed. K. Chandrasekharan. Berlin: Springer Verlag.

Weyl, Hermann. 1994. *The Continuum*. New York: Dover.

Whitehead, Alfred North. 1898. *A Treatise on Universal Algebra, with Applications*. Cambridge: Cambridge University Press.

Whitehead, Alfred North. 1902. On cardinal numbers. *American Journal of Mathematics* 24: 367–394.

Whitehead, Alfred North. 1907. Introduction logistique à la géométrie. *Revue de Métaphysique et de Morale* 15: 34–39.

Whitehead, Alfred North and Bertrand Russell. 1910. *Principia Mathematica*, vol. 1. Cambridge: Cambridge University Press.

Whitehead, Alfred North and Bertrand Russell. 1912. *Principia Mathematica*, vol. 2. Cambridge: Cambridge University Press.

Whitehead, Alfred North and Bertrand Russell. 1913. *Principia Mathematica*, vol. 3. Cambridge: Cambridge University Press.

Wittgenstein, Ludwig. 1921. Logisch-philosophische Abhandlung. *Annalen für Naturphilosophie* 14: 198–262. Reprinted with English translation in Wittgenstein (1922).

Wittgenstein, Ludwig. 1922. *Tractatus logico-philosophicus*. London: Kegan Paul.

Woleński, Jan. 1989. *Logic and Philosophy in the Łvov-Warsaw School*. Dordrecht: Kluwer.

Woleński, Jan. 1994. Jan Łukasiewicz on the liar paradox, logical consequence, truth, and induction. *Modern Logic* 4: 392–399.

Woleński, Jan. 1995. Mathematical logic in Poland 1900–1939: People, circles, institutions, ideas. *Modern Logic* 5: 363–405.

Young, John Wesley. 1917. *Lectures on the Fundamental Concepts of Algebra and Geometry*. New York: Macmillan.

Zach, Richard. 1999. Completeness before Post: Bernays, Hilbert, and the development of propositional logic. *Bulletin of Symbolic Logic* 5(3): 331–366.

Zach, Richard. 2003. The practice of finitism. Epsilon calculus and consistency proofs in Hilbert's program. *Synthese* 137: 211–259.

Zach, Richard. 2004. Hilbert's "Verunglückter Beweis," the first epsilon theorem and consistency proofs. *History and Philosophy of Logic* 25: 79–94.

Zach, Richard. 2006. Hilbert's program then and now. In *Philosophy of Logic. Handbook of the Philosophy of Science*, vol. 5, ed. Dale Jacquette, 411–447. Amsterdam: Elsevier.

Zermelo, Ernst. 1904. Beweis, daß jede Menge wohlgeordnet werden kann. *Mathematische Annalen* 59: 514–516. English translation in van Heijenoort (1967a, 139–141).

Zermelo, Ernst. 1908a. Mathematische Logik. Vorlesung gehalten von Prof. Dr E. Zermelo zu Göttingen im S.S. 1908. Lecture notes by Kurt Grelling. Nachlaß Zermelo, Kapsel 4, Universitätsbibliothek Freiburg im Breisgau.

Zermelo, Ernst. 1908b. Neuer Beweis für die Möglichkeit einer Wohlordung. *Mathematische Annalen* 65: 107–.128. English translation in van Heijenoort (1967a, 183–198).

Zermelo, Ernst. 1908c. Untersuchungen über die Grundlagen der Mengenlehre I. *Mathematische Annalen* 65: 261–281. English translation in van Heijenoort (1967a, 199–215).

Zermelo, Ernst. 1929. Über den Begriff der Definitheit in der Axiomatik. *Fundamenta Mathematicae* 14: 339–344.

Zermelo, Ernst. 1930. Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundamenta Mathematicae* 16: 29–37.

Zermelo, Ernst. 1931. Über die Stufen der Quantifikation und die Logik des Undendlichen. *Jahresbericht der Deutschen Mathematiker-Vereinigung* 41: 85–88.

Zygmunt, Jan. 1973. On the sources of the notion of reduced product. *Reports on Mathematical Logic* 1: 53–67.

Zygmunt, Jan. 1990. Mojżes Presburger: Life and work. *History and Philosophy of Logic* 12: 211–223.

10

Set Theory, Model Theory, and Computability Theory

Wilfrid Hodges

This chapter surveys set theory, model theory, and computability theory: how they first emerged from the foundations of mathematics, and how they have developed since. There are any amounts of mathematical technicalities in the background, but I try to highlight those themes that have some philosophical resonance. Readers who find I said too little on some topic are encouraged to explore the references.

1. The Architecture of Modern Mathematics

The term "mathematical logic" apparently first saw the light of day in a footnote of a paper of De Morgan (1858, p. 78 of [1966]). De Morgan hoped that mathematicians would succeed in wresting the study of argument forms away from the philosophers, because mathematicians are better at detecting significant patterns. De Morgan's coinage went unnoticed; in any case it had little connection with the modern use of the phrase.

In the very different climate of the 1890s, the Italian mathematician Peano proudly put "mathematical logic" into the titles of several of his papers (for example 1891 and 1896–1897). This work of Peano was a direct ancestor of modern model theory and axiomatic set theory. The path from Peano to these two disciplines is worth tracing. Only part of it lies within logic; changes in the broad shape of mathematics are at least as important.

Peano's colleague Pieri said at the beginning of his analysis of projective geometry as "a logico-deductive system" (Pieri 1898, 6):

> The chief feature of the *primitive objects* of any hypothetico-deductive system is that they can be interpreted arbitrarily, within

> certain limits imposed by the *primitive propositions* (*axioms* or *postulates*). ... the Reader is allowed to assign to [the primitive] words and signs an *arbitrary* meaning, so long as this meaning is compatible with the general properties imposed on these entities by the primitive propositions.

Having said that much, Pieri proceeded to make deductions from his axioms without any regard for interpretations (though they briefly appeared again as a technical device for proving independence of axioms). Of course any formal statement deduced from the primitive propositions must be true in any allowed interpretation A. But this route yields no information about the relations between A and any other allowed interpretation B. (See also §1.2 in Mancosu, Zach, and Badesa 2009.)

Russell (1903, ch. xiv) was deeply impressed by Pieri's work. Unlike Pieri, Russell concentrated at once on the allowed interpretations, which he called "geometries." He showed in detail how one could regard a geometry as a set-theoretical object consisting of sets and relations built up from its set of points. This was prophetic but probably quite unmotivated at the time.

In group theory the situation was subtly different. The class of (abstract) groups is defined by a set of axioms, just as the class of geometries. But already Galois in 1832 studied relationships between one group and another, so Pieri's approach would have been useless for him. Late nineteenth-century group theorists agreed that a group is a "system of objects," but it wasn't the custom to explain in set-theoretic terms what a "system of objects" is. Burnside (1897, ch. ii), for example, evaded the question neatly by defining a group as a set of permutations of a given set, satisfying certain conditions. Since it was intuitively obvious how to compose two permutations or invert one, there was no cause for him to spend any time on the set-theoretic "relations" that occupied Russell.

During the first half of the twentieth century, the scenery changed slowly but relentlessly. It became the custom to study classes of structures and the homomorphisms between structures in a class. Some new branches of algebra, such as ring theory, were organized along these lines from the start. In geometry itself the conversion to the new style was complete by the time of Artin (1957). We probably owe commutative diagrams to lectures of Hurewicz around 1940. Category theory, which gives equal status to structures and to structure-preserving maps between them, followed in 1942 (Mac Lane 1971, 29).

Even after these changes, most working mathematicians saw set theory as irrelevant to their practice. However, it came to light that some questions within mathematical disciplines depended on unresolved problems about the universe of sets. Some notable examples were:

a. Vitali (1905) used the Axiom of Choice to show the existence of a non-Lebesgue-measurable set of real numbers. In the relatively new topic of infinite abelian groups, Kaplansky (1954, 6) described Zorn's Lemma (a form of the Axiom of Choice) as "an axiom like other axioms needed to set up the foundations of mathematics," but unlike the other axioms he was careful to indicate where he used this one.

b. In the late 1950s, Łoś found he was unable to complete a theorem on abelian groups without making the set-theoretic assumption that there are no measurable cardinals (a kind of large cardinal, see section 4). In 1983 Eda showed how to plug the gap. Eklof and Mekler (1990) discuss Łoś's and Eda's work.

c. The *Continuum Hypothesis* is the hypothesis that $\omega_1 = 2^\omega$ in cardinal arithmetic. Sierpiński (1934) proved a number of theorems on topological and combinatorial properties of the real numbers, assuming the Continuum Hypothesis. During the 1960s it was shown by forcing (section 3) that many of the results in his book are not provable without some form of this hypothesis.

To apply set-theoretic assumptions, one had to regard the structures involved as particular objects within the universe of sets. This didn't mean that the specialists in these various branches of mathematics had to adopt set-theoretic notation; but it did mean that they needed to be sure that at least someone else could formalize their work within axiomatic set theory.

Bernays (Bernays and Fraenkel 1958) described the required translations into set-theoretic language as "embodying" in set theory. For example (1958 vii.§5), he gave detailed examples of the set-theoretic translation in the case of algebraic structures. Thus he brought up to date the account in Russell (1903).

These developments led to a new picture of the logical structure of mathematics. The picture is not associated with anybody's name in particular; it just happened. In 1936 it was still possible for Tarski to present mathematics as a collection of "deductive systems," each with its own axioms. But from the 1950s onward, classical mathematics had just one deductive system, namely, first-order Zermelo–Fraenkel Set Theory with Choice (ZFC—one writes ZF for the theory without the Axiom of Choice). The axioms of the more algebraic branches of mathematics, for example, group theory or geometry, became set-theoretic definitions of classes of structures within the universe of set theory. A structure satisfying a set of axioms was called a *model* of these axioms. Epistemologically, all axioms except those of set theory became definitions, so the question of their truth completely vanished; at best one could ask whether they defined what they were intended to define. Ontologically, real numbers and all elements of structures were sets, so all problems about the nature of the subject matter reduced to questions about the nature of sets.

Not everybody accepted this set-theoretic picture of mathematics. Some constructive mathematicians preferred to work within their own styles of system; see von Plato (2009). But also some classical mathematicians experimented with alternatives, such as toposes (section 4).

2. A Catalogue of Sciences

The American Mathematical Society (the AMS) has for decades maintained a Subject Classification which catalogs the different areas of mathematics. We can learn the shape of mathematical logic by following the development of this catalog.

Logic appeared in the 1950 Subject Classification as "Foundations." In 1952 the first two categories were

Logic and Foundations

Set theory

What came under these two heads?

Broadly, Logic and Foundations covered the topics discussed in Mancosu, Zach, and Badesa (2009), so far as these topics were still mathematically active. What Mancosu et al. (2009) describe as "Zermelo's Axiomatization of Set Theory and Related Foundational Issues" was certainly an active area, but for the AMS it was not a part of set theory. Set theory was a mathematical discipline that included, for example, the contents of Hausdorff's text (1927). For Hausdorff, the heart of set theory lay in the theory of cardinal and ordinal numbers, the cardinalities of power sets, and descriptive set theory (essentially the topology and combinatorics of the real numbers—see Kechris 1995). Hausdorff mentions Zermelo just twice, once to introduce Zermelo's well-ordering principle and once to note that he will not explore Zermelo's axiomatics. By 1952, set theory also included Ramsey theory and other topics that today are commonly grouped under combinatorial set theory or infinitary combinatorics. It was possible to pursue all these topics as parts of *naive set theory*, that is, without formalized axioms (Halmos 1974).

Sometime in the early 1960s it became common for mathematical logicians to classify themselves into four main groups: model theorists, set theorists, recursion theorists, and proof theorists. This classification is still standard today. It first reached the ASL Subject Classification in 1973, when the section "Logic and Foundations" became subdivided at two levels. The higher level divisions were

Philosophical and critical

Classical logical systems

Nonclassical formal systems

Proof theory

Constructive mathematics

Recursion theory

Methodology of deductive systems

Model theory

Algebraic logic

Set theory

Proof theory and constructive mathematics overlapped each other; from 1980 they were combined in one subsection. Algebraic logic was a minority interest but significant enough to include. There were some relevant developments in other parts of the 1973 Subject Classification too: Group Theory sprouted a subsection "Metamathematical considerations," and Computer Science acquired subsections on "Algorithms," "Computational complexity and efficiency," and "Theorem proving." The 1980 Subject Classification replaced "Logic and Foundations" with "Mathematical Logic and Foundations," and included "Nonstandard models" as a new subsection.

The 2000 Subject Classification saw two further changes. "Recursion theory" was renamed "Computability and recursion theory." This was at least partly a response to a growing feeling among recursion theorists that their subject should have been called computability theory (see Soare 1996). The second change was that Set Theory disappeared as a separate section in the Classification, and survived only as a branch of logic. Presumably the reason was that it had become impossible to distinguish between logicians' set theory and anybody else's. Three examples will suffice. First, in 1975 the logician Martin (1975) proved that all Borel games are determined. This result is useful for logicians, but its proof needs no specifically logical methods and no abnormal axioms. Second, the logician Shelah (1994) published a range of results, mostly discovered by himself, about the cofinalities of cartesian products of ordinals. Though questions in logic had led him to this theory, he commented, "Cantor should have no problems understanding and (so I feel) appreciating the theorems and even most proofs in this book" (1994, xi). Third, various people found "logical" proofs of results that had been proved earlier in the nonlogical tradition of set theory; see, for example, Simpson's logical recasting of Nash-Williams theory in the appendix to Mansfield and Weitkamp (1985).

The main shaping of mathematical logic took place in the interval between two classic textbooks, Kleene's *Introduction to Metamathematics* (1952) and Shoenfield's *Mathematical Logic* (1967). Shoenfield's book is still suitable as an introduction; the Association for Symbolic Logic and A. K. Peters reissued it in 2001.

3. Set Theory I: Building Models

As the axioms of set theory came to carry a greater load of mathematics, it was increasingly necessary to justify them. They should be intuitively true and they should allow one to deduce anything needed in any acceptable branch of mathematics. Also, a traditional task of foundations has been to demonstrate the independence of an axiom from other axioms, that is, to show that it is not logically deducible from the other axioms.

Mancosu, Zach, and Badesa (2009) carry the story up to 1930. Most of the work after this date took the axioms of ZF as given; one wanted to know what else followed from ZF and what didn't. (Some authors used axiomatic systems that were equivalent to ZF but had two sorts of individual: sets and classes. Two-sorted systems largely dropped out of use in the 1960s.)

The first major advance after 1930 was Gödel's slim but deep monograph (1940) on the constructible universe. Gödel worked in his own two-sorted system, but it's conventional to state his results in terms of ZF. He showed how, within the formal language of ZF, one can define a class L of sets so that L together with the restriction to L of the membership relation $\in$ also satisfies the axioms of ZF. Given any model M of ZF, L (more precisely L^M, the interpretation in M of the definition of L) forms a definable substructure of M that is also a model of ZF. Definable substructures of a model of ZF which are provably models of ZF in their own right are known as *inner models*.

Gödel's inner model L, known as the *constructible universe*, has further important properties. (See Devlin 1984 for an introduction.) An ordinal^L (i.e., an ordinal in the sense of the model L) is the same thing as an ordinal. The model L^L is again L. There is a formula that defines a bijection between the class of ordinals and the whole of L. Using this formula, we can write an axiom "$V = L$" (the *Axiom of Constructibility*), which says that every set is in L; since $L^L = L$, L itself satisfies this axiom. From "$V = L$" one can deduce many other statements. Gödel himself deduced the Axiom of Choice (easily—the formula listing all the sets in L makes choices for us). He also deduced the Generalized Continuum Hypothesis (GCH for short), which says that for every infinite cardinal κ, κ^+ (the next cardinal after κ) is equal to 2^κ (the number of subsets of a set of cardinality κ). Gödel showed that if ZF is consistent, then so is ZF with "$V = L$" added: If we have a contradiction from ZF + "$V = L$," then we can write down a proof from ZF that the contradiction holds in L, and this shows that ZF itself is inconsistent. As a corollary, if ZF is consistent then so are ZFC and ZFC + GCH.

A subtle analysis by Jensen (1972) revealed several combinatorial principles that are true in L; the most famous are box and diamond ($\Box$ and $\Diamond$), and some combinatorial devices known as *morasses*. These principles settle many previously open mathematical questions, and in most cases we now know that the answers to these questions don't follow from ZFC alone, or even from ZFC + GCH. Two famous examples are that Suslin continua (which are distinct from but in many ways similar to the real line) exist and that Whitehead's problem in abelian groups has a positive solution. The Whitehead problem is only one example of the deep relations between abelian groups and infinitary combinatorics; see Eklof (1980).

The fact that L^L equals L puts severe limits on consistency proofs by inner models. For example, if we assume "$V = L$," then it follows that every inner model is the entire universe. So to prove that a set-theoretic statement ϕ is consistent with ZF, in cases where ϕ is inconsistent with "$V = L$," we need another method. It took over 20 years to find one. But Cohen (1963) finally made the breakthrough.

Levy, Solovay, Shoenfield, and others refined Cohen's method, and the following scheme emerged. We take a model M of ZFC (the *ground model*) and a partially ordered set $\mathbb{P}$, which is an element in M; $\mathbb{P}$ is called the *notion of forcing*, and its elements are called *conditions*. Together M and $\mathbb{P}$ determine a language $\mathcal{L}$ known as the *forcing language*; $\mathcal{L}$ is a language for set theory, and elements of M serve as constant symbols in $\mathcal{L}$. A relation "p forces ϕ," between conditions p and sentences ϕ of $\mathcal{L}$, is defined in M; the most significant clause in the definition is

> p forces $\neg\phi$ if and only if there is no condition $q \leqslant_{\mathbb{P}} p$ such that q forces ϕ.

Now if M itself is countable, then from outside M we can build up, in a countable sequence of steps, a set G of conditions so that (among other things) for every sentence ϕ of $\mathcal{L}$, exactly one of ϕ and $\neg\phi$ is forced by some condition in G. We take the set $\{\, \phi : \text{some condition in } G \text{ forces } \phi \,\}$, and we read it as a description of a structure N. All this can be arranged so that N is in fact a model of ZFC extending M (but not adding any new elements of any member of M), N has the same ordinals as M, and N contains G. We write N as $M[G]$, and we call it a *generic extension* of M; the set G is called the *generic set*. Because G is defined from outside M, there is no guarantee that G is a set in M, and in fact we easily ensure that it is not in M.

If M is a model of "$V = L$," then $M[G]$ has the same ordinals as M but extra elements (for example G), and so "$V = L$" is false in $M[G]$. This proves that "$V = L$" is not deducible from ZFC. By working a little harder, we can build a model of ZFC in which the Continuum Hypothesis fails, too. An extra twist, using permutation groups, yields modified generic extensions in which the Axiom of Choice fails. In short, Cohen had found a very flexible way of proving independence results in set theory. His method is known as *forcing*.

Cohen assumed he had a countable model of ZFC. This is a stronger assumption than the truth of ZFC itself; if ZFC entailed the existence of a countable model of ZFC, it would entail the consistency of ZFC, and hence by Gödel's Second Incompleteness Theorem, ZFC would be inconsistent. Cohen himself pointed out a way around this problem. To show the consistency of ZFC with the statement "$V \neq L$," by the Compactness Theorem for first-order logic (section 6) we need only show that this statement is consistent with every finite subset T of ZFC. But in ZFC we can prove the existence of a countable model M of U, where U is a finite subset of ZFC large enough to ensure that $M[G]$ is a model of T and "$V \neq L$." (Also in forcing we normally assume that M is a *transitive model*, i.e., that its elements are genuinely sets, the membership relation is genuine membership, and every member of an element of M is also an element of M. These stronger assumptions are justified in the same way.)

Scott discovered another way around the problem. Given M, we build in M a complete boolean algebra $\mathbb{B}$ and give each sentence ϕ of the forcing language a *boolean value* $|\phi|_{\mathbb{B}}$ in $\mathbb{B}$. Choosing a generic set is equivalent to finding a generic (i.e., suitably well-behaved) ultrafilter G on $\mathbb{B}$; the sentences true in $M[G]$ will be those whose boolean values lie in the ultrafilter. Cohen's construction is a special case; for each notion of forcing $\mathbb{P}$ there is a corresponding complete boolean algebra $\mathbb{B}$. But Scott's method is more general than Cohen's. If $|\phi|_{\mathbb{B}}$ is not the top element of $\mathbb{B}$, this fact already shows that ϕ is not deducible from ZFC, whether or not we can find a generic ultrafilter on $\mathbb{B}$. Scott's approach is known as *boolean-valued models*.

Among several extensions of Cohen's ideas, probably the most important is iteration. Instead of taking a sequence of generic extensions M, $M[G]$, $M[G][G']$, we reach $M[G][G']$ as $M[G^\star]$ in a single step, by finding the generic set $G^\star$ in a composite notion of forcing. More generally we can pack a transfinite sequence of generic extensions into a single step; see for example Jech (1986). The result is a battery of techniques that have allowed set theorists to prove the consistency and independence of most of the problems in descriptive set theory left open by the set theorists of the 1920s. One of the most powerful iteration techniques is the *proper forcing* of Shelah (1998).

Forcing axioms are set-theoretic statements that sum up the things made true by some form of forcing. The earliest examples were a cluster of statements known as *Martin's Axiom*, which sum up the effects of forcing with c.c.c. notions of forcing. For example, to prove that the statement "There are no Suslin continua" is consistent with ZFC, we need only deduce the statement from Martin's Axiom. *Martin's Maximum* is a stronger version of Martin's Axiom; we need large cardinals (see the next section) to prove its consistency. A related notion is the Ω-logic of Woodin; if ϕ is a theorem of Ω-logic in some model M of ZFC, then ϕ remains a theorem of Ω-logic in every generic extension of M. Woodin (1999) makes sophisticated applications of this range of ideas.

Jech (1997) is a general reference covering the constructible universe and forcing.

4. Set Theory II: Large Cardinal and Other Axioms

In general a proof that neither ϕ nor $\neg\phi$ follows from ZFC gives us no clue whether ϕ is in fact true. If we want to settle the question, we need new arguments.

Not all set theorists agree that the choice of set-theoretic axioms is a matter of truth or falsity. Of those who do, many have respected the advice given by Gödel (1947). First, Gödel urged set theorists to explore axioms that express that "very great cardinal numbers" exist. The intuitive basis for such axioms is that the universe of sets has no upper bound; the axioms "assert the existence of still further iterations of the operation 'set of.' " Second, he suggested that axioms can justify themselves by their "fruitfulness in consequences and in particular in 'verifiable' consequences, i.e., consequences demonstrable without the new axiom, whose proofs by means of the new axiom, however, are considerably simpler and easier to discover, and make it possible to condense into one many different proofs."

In the spirit of Gödel's first proposal, logicians began to explore the consequences of assuming that there exists a measurable cardinal (i.e., a cardinal whose power set carries a countably complete nonprincipal ultrafilter). Ulam had introduced measurable cardinals in 1930. By 1964 it was known that if there are any measurable cardinals at all, the smallest one κ must be very large; for example, there must be κ strongly inaccessible cardinals less than it. Scott (1961) showed, by considering an ultrapower of the universe by a countably complete ultrafilter (see section 6.1), that there are no measurable cardinals in Gödel's constructible universe L. Subsequent work of Gaifman, Rowbottom, Silver, and Solovay showed that if there is a measurable cardinal, then there is a nonconstructible Δ^1_3 set of natural numbers (known as $0^\sharp$, see section 8 for Δ^1_3) which encodes a description of L as an Ehrenfeucht-Mostowski model (see section 6.3) with spine containing all the uncountable cardinals. One consequence is that ω_1^L, the cardinal ω_1 in the sense of L, is tiny compared with ω_1 itself. Jech (1997) surveys all this work.

Jensen's *covering lemma* (see Dodd 1982) classified all models of ZFC into two classes. If a model of ZFC contains $0^\sharp$, then its cardinals are widely different from those of L, as we have just seen for ω_1. On the other hand, if the model M doesn't contain $0^\sharp$, then for every uncountable set X of ordinals in the model, there is in L^M a set of ordinals Y of the same cardinality as X, with $X \subseteq Y$; it follows that cardinalities and cofinalities are nearly the same in L^M as they are in M. Subsequent work by Jensen, Dodd, Mitchell, Steel, and others showed that there is a similar dichotomy for models of set theory containing a measurable cardinal. Further dichotomies of the same kind exist for stronger large cardinal assumptions, but they become harder to describe as the assumptions grow stronger. In each case one has to find an inner model called a *core model*, which plays a role analogous to L. See Zeman (2002).

Mycielski and Steinhaus (1962) suggested an axiom broadly in the spirit of Gödel's second proposal. Let X be a set of infinite sequences of natural numbers. Suppose two players take turns to choose natural numbers $n_0, n_1, n_2, \ldots$; suppose we reckon that the first player wins this game if the resulting infinite sequence is in X, otherwise the second player wins. Then we say that the game is *determined* if one of the players has a strategy guaranteeing that he will win regardless of how the other player plays. Mycielski and Steinhaus proposed the axiom that for every such set X, the game is determined. Their axiom is known as the *Axiom of Determinacy*, AD for short.

The axiom AD is inconsistent with the Axiom of Choice, but it has a number of very pleasant consequences, for example, that every set of real numbers has a Lebesgue measure. (ZFC entails that all analytic sets of reals are Lebesgue measurable, but in L there is a Δ^1_2 set of reals that is not Lebesgue measurable.) A weaker statement is $\mathrm{AD}^{L(\mathbb{R})}$, which states that AD is true in the universe of sets constructible in the sense of Gödel but starting from the set $\mathbb{R}$ of real numbers. There is no reason to think that this weaker axiom is inconsistent with ZFC. Mycielski and Steinhaus expressed the hope that descriptive set theorists might be able to assume $\mathrm{AD}^{L(\mathbb{R})}$ and then work inside $L(\mathbb{R})$, thus getting most of the benefits of AD.

Gödel's two proposals came together in a satisfying way during the 1980s. There is a type of large cardinal called a *Woodin cardinal*; Woodin cardinals have many measurable cardinals below them. Martin, Steel, and Woodin succeeded in showing from ZFC that if there are infinitely many Woodin cardinals with a measurable cardinal above them, then $\mathrm{AD}^{L(\mathbb{R})}$ is true. Woodin also showed that the theory ZFC + "There are infinitely many Woodin cardinals" is consistent if and only if ZF + AD is consistent. See Kanamori (1994, §32).

In line with Gödel's suggestions, one might read these results of Woodin et al. as reasons for adopting $\mathrm{AD}^{L(\mathbb{R})}$ as an axiom. Woodin (2001) has pressed the matter further. He argues that we are now in sight of being able to decide whether the Continuum Hypothesis is true or false. Gödel himself in 1970 submitted for publication a paper giving "considerations leading to the probable conclusion that the true power of the continuum is $\aleph_2$" (Gödel 1995, 405–422). He withdrew the paper when an error came to light, but some set theorists still hope that the intuition behind it can be rescued.

Most set theorists now accept ZFC as the standard framework. It includes the Axiom of Choice; in fact several well-known theorems of classical pure mathematics turned out to be equivalent to the Axiom of Choice, in the sense that each of them together with the axioms of ZF entails Choice. Examples are: Tikhonov's theorem that every product of compact topological spaces is compact; Krull's theorem that every commutative ring with $0 \neq 1$ has a maximal ideal; the theorem that every vector space has a basis. Details of these three results and others are in Rubin and Rubin (1985).

ZFC also includes the Axiom of Foundation, which guarantees that there are no infinite descending sequences of members:

$$\cdots \in a_2 \in a_1 \in a_0.$$

This useful axiom allows us to use induction on the ordinals, in analogy with induction on the natural numbers, to prove theorems about all sets. But the truth of the axiom is hardly an issue of principle. Within any model M of the other axioms of ZF, we can pick out an inner model satisfying Foundation, and (given Choice) this inner model contains an isomorphic copy of any mathematical structure in M. So the use of set theory with axioms contradicting Foundation, as in Aczel (1988), need not imply any opposition to ZFC. As Aczel shows, non–well-founded models have applications in computer science, where they provide intuitive representations of infinite processes.

Reinhardt (1974) explored axioms for set theory in a style proposed earlier by Ackermann. These axioms can be read as saying that the universe of sets has an invisible top: For example, there is no largest infinite cardinal, but the universe itself can't be the first place where this is true, so there must be within the universe a cardinal such that there is no largest cardinal below it, that is, a limit cardinal. Although this framework looks at first very different from ZFC, Reinhardt showed that natural versions of Ackermann's idea lead either to a theory equivalent to ZFC or to ZFC with some large cardinal axioms.

Quine's (1937) set theory New Foundations (NF for short) was much harder to digest. Its consistency relative to ZFC is still an open problem. Specker (1953) showed that NF implies the negation of the Axiom of Choice. See Forster (1995) for further information.

The notion of a topos arose first in category-theoretic work of Grothendieck and his colleagues, as part of a program to integrate algebraic geometry with the theory of sheaves. In 1970, Lawvere and Tierney extracted the first-order part of the Grothendieck definition of a topos, and described the models of their axioms as *elementary toposes*. A model of set theory can be construed as an elementary topos; but elementary toposes in general have no notion of element, and they encode an "internal logic" which need not be two-valued. Lawvere urged that elementary toposes are a better foundation than ZFC for mathematics. He pointed to the successes of category theory in homing in on mathematically interesting notions. More controversially he added (1971):

> [Developing and transforming] "set theory" requires taking account of the experience that the main pairs of opposing tendencies in mathematics take the form of adjoint functors, and frees us of the mathematically irrelevant traces ($\in$) left behind by the process of accumulating ($\cup$) the power set (P) at each state of a metaphysical "construction."

The axioms of elementary topos theory are equiconsistent with a form of set theory that has a weak comprehension axiom in place of Separation and Replacement. See Johnstone (1977, ch. 9).

5. Model Theory I: From Deductive Systems to Formal Theories

The name "theory of models" comes from Tarski (1954). But a better date for the birth of the new subject is 1950, when Abraham Robinson (1950) and Tarski addressed the International Congress of Mathematicians at Harvard. The theme of both their talks was that the general theory of deductive systems, as it had been built up by logicians working in the foundations of mathematics, had reached a point where it was ready for "effective application ... to mathematics proper" (to cite Robinson's opening paragraph). Mal'tsev (1941) in the Soviet Union had already launched the same program, but it took some time for his work to be recognized in the West.

The new discipline was a branch of pure mathematics with no philosophical content and no connection with foundations. One could hardly say the same of the theory of deductive systems that had given rise to it. In this section we discuss how the old foundational discipline metamorphosed into the new mathematical one. The next section reports the main technical achievements of model theory.

During the 1930s Tarski was working on a project to give definitions of central concepts in the foundations of mathematics, in terms of notions from set theory and syntax. In his textbook (1936), he explained that each deductive system has a set of "primitive terms" whose meanings "seem to us to be immediately understandable," and a set of "axioms" which "we accept as true without establishing them in any way" (p. 110 of the 1994 translation). How does one choose the axioms? Tarski's teacher Leśniewski had expressed the view that

> every language system, even the most formalized, says "something" "about something." ... any formalized theory consists of statements endowed with meaning, and not of "more or less picturesque formulas." (Woleński 1989, 145)

This was a key part of Leśniewski's "intuitionistic formalism," which Tarski (1983, 62) said he endorsed.

Tarski's view here had two major flaws. In the first place, we can't formalize the notions of "immediately understandable" and "accepting as true" in terms of notions from set theory and syntax. So the crucial notion of a deductive system remains formally undefined. In the second place, as we saw in section 1, group theory studies not just a single group at a time but also relations between several groups. The only appropriate way to treat a deductive system whose axioms are those of group theory (or indeed those of any algebraic theory) would be to agree with Pieri, and against Leśniewski, that there is no fixed interpretation. Tarski admitted this last point already in (1936, 119f. of the 1994 translation), but he offered no repair of his general framework to accommodate it.

This background explains the extraordinary definition of "model" that Tarski was using in the late 1930s, as reported by Mancosu, Zach, and Badesa (2009) on their final page. His problem was to formalize the idea of "changing the meanings of the primitive terms." Since he had no formal description of what counted as assigning the original meanings, he could hardly define how to change them. So he physically removed the primitive terms and replaced them by higher-order variables, and then invoked the notion of satisfaction. This elaborate rigmarole survived as late as Mostowski's textbook (1948).

In Hodges (2004) I analyzed how the progress of theorems made Tarski's definitions of the 1930s steadily less usable in practice. Around 1950, he consciously made the break. In future, deductive systems (now called simply *theories*) are sets of sentences of a formal language. A mathematician might choose to assign fixed meanings (and the theory would then have a *standard* or *intended model*), but the meanings are no part of the theory itself. The truth definition is recast with a distinguished set of *nonlogical constants* (the primitive terms). The notion of model is defined in terms of functions assigning set-theoretic objects to these constants. The notions of meaning and reference vanish into set theory. Tarski reports the final outcome in Tarski and Vaught (1957). In a footnote added to a translation of a paper in 1956 (1983, 62), Tarski stated that he no longer subscribed to Leśniewski's intuitionistic formalism.

Robinson had no such adjustments to make. His notion of model came from Carnap (1942). Carnap had long since rejected the view that mathematics consists of "statements endowed with meaning"; in 1935 (36) he had said, "The formal sciences do not have any objects at all; they are systems of auxiliary statements without objects and without content." Incidentally Robinson also took from Carnap's semantics the idea of adding constant symbols to the formal language as names of the elements of a structure. Robinson's technique of *diagrams*, which he used for constructing models of a theory with given structures as substructures, rests directly on this idea.

Philosophers studying model-theoretic work of the 1950s have sometimes been tempted to read into it some of the earlier foundational concerns. For example in Tarski, Mostowski, and Robinson (1953, 8) we learn that "A sentence Φ is said to be a *logical consequence* of a set A of sentences if it is satisfied in every realization $\mathcal{R}$ in which all sentences of A are satisfied." In other words, a formal sentence is a logical consequence of a formal theory if and only if every model of the theory is also a model of the sentence. One should be clear that Tarski is simply giving a convenient name to a useful model-theoretic relation. His definition is formally close to the characterization he proposed for logical consequence in the 1930s, but in 1953 he is not making a claim about any preexisting notion of logical consequence—any more than I express my view of the Royal Family by calling my dog Prince.

6. Model Theory II: The Technical Results

Model theory inherited several important mathematical results from the foundational research of the period 1920–1940. One result central to the later theory was the fact (proved by Gödel in his doctoral thesis in 1930) that if T is a first-order theory, ϕ is a first-order sentence and every model of T is a model of ϕ, then T has a finite subset U such that all models of U are models of ϕ. This became known as the Compactness Theorem. Gödel proved it only for languages with countably many symbols. Mal'tsev (1936) (and later but independently Abraham Robinson and Henkin) extended the result to first-order languages of arbitrary cardinality.

Robinson saw how to use the Compactness Theorem to find a structure that satisfies all the same first-order sentences as the ordered field of real numbers, but has extra elements that can't be identified with any real numbers. These extra elements are said to be *nonstandard*. Some of them must be *infinitesimal*, that is, greater than 0 but smaller than 1/2, 1/4, 1/8, and so on. Robinson also realized that one can use infinitesimals in very much the way that Leibniz had proposed, for proving theorems of differential and integral calculus. The theorems so proved would transfer back to the "genuine" field of real numbers, provided that they can be written as first-order sentences. See Bell (1998) for the resulting *Nonstandard Analysis.*

In the same spirit, one can study nonstandard models of first-order arithmetic or of any other first-order description of some known mathematical structure. Work in this area led Paris and Harrington (1977) to their supplement to Gödel's First Incompleteness Theorem: They found a previously known mathematical theorem that can be stated as a first-order sentence about the natural numbers but can't be proved from first-order Peano Arithmetic. See Kaye (1991) for more on nonstandard models of arithmetic.

Model theory also inherited from general algebra some set-theoretic constructions: cartesian product, substructure, generating set, unions of chains, isomorphism, embedding, and homomorphism. A theorem of Birkhoff (1935) said that a class of algebras is closed under homomorphic images, substructures, and cartesian products if and only if it is the class of models of a set of universally quantified equations. This was a model-theoretic result ahead of its time.

What follows is a sample of typical model-theoretic questions. Marker (2002) is a broad and up-to-date text; Hodges (1997) is a gentler introduction, and Chang and Keisler (1973) is the classic textbook. Here I concentrate on first-order model theory. Barwise and Feferman (1985) report on a wider range of logics; for work in progress on the model theory of Banach algebras one should try a Web search with the keywords "Henson" and "Iovino."

6.1. What Are the Significant Maps between Structures?

An *elementary embedding* from a structure A to a structure B is a homomorphism f such that if elements of A satisfy a first-order formula in A, then their images under f satisfy the same formula in B. If A here is a substructure of B and f is the inclusion map, we say that B is an *elementary extension* of A. Robinson used these notions implicitly in his doctoral thesis in 1949. Tarski and Vaught (1957) defined them explicitly and proved their fundamental properties. For example their version of the *Downward Löwenheim-Skolem Theorem* says, in its simplest case, that if A is an infinite structure whose first-order language has at most countably many nonlogical constants, then A is an elementary extension of some structure B with only countably many elements. Their *Theorem on Unions of Elementary Chains* is a tool for building up arbitrarily large elementary extensions of any infinite structure. The theorem stating the existence of these elementary extensions is known as the *Upward Löwenheim-Skolem Theorem* (though it was proved by Tarski, and Skolem rejected it for foundational reasons). In the mid-1950s another construction to prove the same theorem came to light: One took a cartesian power A^I of the structure A and used an ultrafilter D on the index set I to form a homomorphic image

A^I/D in which A was elementarily embedded. The structure A^I/D is known as an *ultrapower* of A; Chang and Keisler (1973) analyze it in detail.

Robinson observed that all embeddings between algebraically closed fields are elementary embeddings. This was the first of a long sequence of translations between model-theoretic and algebraic notions. One of the most influential was Macintyre's (1971) demonstration that an infinite field is algebraically closed if and only if its first-order theory is ω-stable (see section 6.4).

6.2. What Is Elementary Equivalence?

Two structures A, B are said to be *elementarily equivalent*, in symbols $A \equiv B$, if they have the same language and the same sentences of this language are true in both of them. Tarski introduced this notion informally but quite precisely in 1930 (before he had a truth definition adequate for formalizing it). In 1946 (Sinaceur 2000), he asked for a "theory" of elementary equivalence. By the 1950s, this request had hardened into the question of giving purely structural necessary and sufficient conditions for two structures to be elementarily equivalent.

Answers of two kinds were found. The first is the *back-and-forth criterion*. It was discovered by the French-Algerian logician Fraïssé (and later rediscovered by the Kazakh logician Taĭmanov). As Ehrenfeucht pointed out, it can be thought of as a game: Two players take turns to pick elements of the structures, the first from either structure and then the second from the other structure. At each stage the second player has to choose so that the elements chosen so far in one structure exactly match those chosen in the other; if she fails to do this, she loses. The two structures are elementarily equivalent if and only if the second player has a strategy that ensures she never loses. The back-and-forth criterion adapts to other languages besides first-order, and works equally well for finite and infinite structures. It has become an important tool of theoretical computer science. See Ebbinghaus and Flum (1999).

Lindström (1969) found an unexpected application of the back-and-forth criterion. He showed that if a logic contains first-order logic, is closed under first-order operations (such as conjunction and quantification), and obeys analogs of the Compactness Theorem and the Downward Löwenheim-Skolem Theorem (for example), then the logic expresses nothing that can't already be expressed with first-order formulas. It seemed to some that his results might form a mathematical proof that first-order logic is in some sense the basic logic; though closer reflection makes one wonder whether satisfying the Compactness Theorem is really a mark of "the basic logic," whatever that might mean. But Lindström did inspire valuable work on the comparison of different logics, much of which is reported in Barwise and Feferman (1985).

Ultrafilters gave a second answer to Tarski's question. Two structures A and B are elementarily equivalent if and only if there are a set I and an ultrafilter D on I such that the two ultrapowers A^I/D and B^I/D are isomorphic. This was proved first under special set-theoretic assumptions by Keisler, and then from ZFC by Shelah; see Chang and Keisler (1973, §6.1). In a precise technical sense known as *saturation*, ultrapowers tend to be "rich in elements." This richness of the ultrapowers A^I/D and B^I/D allowed Keisler and Shelah to find, for every element of one ultrapower, a corresponding element of the other.

6.3. Model-Theoretic Constructions

What methods of construction are available to produce a model of T, where T can be any (or any suitable) first-order theory? We have already mentioned three such methods: Robinson diagrams, unions of elementary chains, ultrapowers. Three others worth noting are omitting types, Fraïssé limits, and the Ehrenfeucht-Mostowski construction.

The opposite of a saturated structure is one that is missing some types of element; we say it *omits* them. Vaught (1961) showed how to construct models of a theory that have countably many elements and omit a given countable set of types. Some other model-theoretic constructions, notably the finite forcing of Abraham Robinson, are close analogs of Vaught's construction. One can express the common feature of these constructions as a game; see Hodges (1985).

Fraïssé observed that we can build a countable densely ordered set without endpoints by taking a family of finite ordered sets and slotting them into each other in all possible ways. His generalization of this construction is known as *Fraïssé limits*. Further generalizations are due to Jónsson, Shelah, and Hrushovski for a wide range of model-theoretic purposes.

The Ehrenfeucht-Mostowski construction was perhaps the most surprising discovery in early model theory. For any consistent theory T with infinite models, and any linearly ordered set $(I, <)$, we can build a model of T around the set I (the *spine* or *indiscernible sequence*) so that the behavior of each element of the model is completely determined by the way it is attached to the spine. For example every order-automorphism of $(I, <)$ induces an automorphism of the model. The resulting *Ehrenfeucht-Mostowski models* generally omit many types; in this way they are a kind of opposite to ultrapowers, which tend to omit few types. See Marker (2002, ch. 5).

6.4. Categoricity

Let κ be an infinite cardinal. We say that a theory T in a countable first-order language is *κ-categorical* if T has models with κ elements, but any two such models are isomorphic. We say T is *uncountably categorical* if it is κ-categorical for all uncountable cardinals κ. (Necessarily then T is complete, because otherwise it would have two models that are not elementarily equivalent, and by the Upward Löwenheim-Skolem Theorem these two models would have elementary extensions with the same number of elements.) Morley (1965) provided a powerful range of techniques for analyzing the models of an uncountably categorical theory. This paper and subsequent work by Marsh, Baldwin, Lachlan, and Zilber gave us a compelling picture of how these models are built up. At their core they have a geometric structure called a *strongly minimal set*, which is determined once we know its dimension; the rest of the model is closely tied to this set. Thanks to this work, model theorists became more interested in the models than in the theory, and they began to speak of the models themselves as "uncountably categorical structures." Buechler (1996) gives a modern account of Morley's results.

Morley's paper generated two programs of research. The first was due to Shelah. As Shelah noted, knowing that a theory is uncountably categorical gives a vast amount of information about its models, but knowing that it is not uncountably categorical tells us almost nothing. He urged that it would be better to find a wider class than that of uncountably categorical theories, in such a way that we get the maximum amount of information both from knowing that a theory is in the class, and from knowing that it is not. Typically a theory in the class—a "good" theory—would have its models so tidily arranged that we can catalog them with parameters, like the dimension of a vector space. A theory outside the class—a "bad" theory—would have a chaotic family of models; for example, we should be able to show that it has two models that are not isomorphic but are very hard to tell apart. For any countable first-order language, Shelah found such a division of the class of complete theories into a "good" class and a "bad" class, and he called it the *main gap*. The full definition of the main gap is complicated, but the heuristic was very successful in generating new techniques, both on the "good" side and on the "bad." He reported his main results in 1990. In more recent work, which continues to the present, he extended these ideas to logics in which the Compactness Theorem fails, and more generally to classes of structure that are defined by closure under certain general constructions rather than by the truth of sentences of any formal language.

The second program was mainly due to Zilber. His idea was not to broaden the class of uncountably categorical theories but to refine it. He introduced a classification of these theories into three types, according to the lattices of definable sets in their models: *disintegrated* (typical case an infinite set with no structure), *modular* (typical case a vector space), and *nonmodular* (typical case an algebraically closed field). This description gave new information, or at least a new viewpoint, on various classical structures arising in algebraic geometry and field theory. A striking success was Hrushovski's use of this machinery to complete the proof (previously given only in some special cases) of the Mordell-Lang conjecture on function fields; the book by Bouscaren (1998) surveys Hrushovski's argument. This area of research remains very active, but is too technical to explore further here. Wagner (2000) reports one important direction.

The terminology in this area has shifted with changing interests. After Morley's work, Rowbottom classified theories according to the cardinalities of the families of definable relations on their models; a theory with small families on models with κ elements was said to be *κ-stable*, and in consequence this branch of model theory became known as *stability theory*. But Shelah preferred the name *classification theory* for the general study of classifications of theories as good/bad. The purely first-order branch of stability theory that used Zilber's modular/nonmodular distinction became known as *geometric stability theory* (Pillay 1996). More general is *geometric model theory*, which includes the program begun by Van den Dries, Pillay, and Steinhorn, to study *o-minimal structures*. These are structures carrying a definable linear ordering, and their definable subsets are finite unions of singletons and open intervals in the ordering; the field of real numbers is the paradigm example. The theories of o-minimal structures are always unstable and hence on the bad side of Shelah's main gap; nevertheless, these structures turn out to be very tractable and there are important examples. See Van den Dries (1998).

7. Computability Theory I: The Notion of Computability

Set theory and model theory were created in response to developments within mathematics as a whole. Not so recursion theory: This theory sprang up quite suddenly in the 1930s, and provided answers to questions that nobody had seen any reason to ask.

In 1936, Alan Turing, then a student at Cambridge University, published a paper (Turing 1936) in which he characterized the class of functions whose domains and values consist of strings of symbols from a finite alphabet and which can be mechanically computed by a human being. According to his analysis, the human being can only survey a bounded part of the calculation at any one time, and (so far as the calculation is concerned) he can only be in one of a finite number of mental states at any one time; the number of states depends on the function being computed. What he does at any stage in the calculation depends on what he can see on the paper and the state of his mind at the time, and the only things he can do are to shift his attention to a different part of the calculation, add or erase a symbol, or move to a different mental state. Turing showed that the functions computable in this way are exactly those calculated by a certain type of machine, now known as a *Turing machine*. Turing also showed that for any alphabet A of symbols, there is a Turing machine (a *universal Turing machine*) that, given as input a second Turing machine for calculating with symbols in A and a possible input to the second machine, computes whatever value the second machine would compute when given that input. He showed that there is no possible Turing machine which, given a second Turing machine for calculating with A and a possible input, will always answer the question whether or not the second machine computing with that input will ever complete its calculation. (This is the *halting problem for Turing machines*.) A function computed by a Turing machine is said to be *Turing-computable*, or more briefly *computable*. See Sieg (1994) on Turing's analysis.

When the alphabet A is suitable for describing natural numbers (for example $\{0, 1\}$ for naming numbers in binary notation), the everywhere-defined Turing-computable functions are exactly the generalized recursive functions that Gödel (1934) had introduced in connection with his incompleteness theorem (though without anything like Turing's analysis of computability). Kleene (1943), starting out from Gödel's definition but dropping the assumption that the functions are defined everywhere, defined the class of *partial recursive functions*, which were soon seen to be exactly the Turing-computable functions. Later the word "partial" was dropped. Kleene, Péter, and others built up a mathematical theory of recursive functions, for which see Rogers (1967). Particularly important in this theory are Kleene's fixed point theorems, which he himself referred to as the "recursion theorems" (1952, 348, 352f.); they say that recursive functions can be constructed as fixed points of certain kinds of operation. The entire theory generalizes from sets of natural numbers to n-ary relations on natural numbers, for any fixed natural number n.

Turing's Thesis is the claim that independent of Turing we have an intuitive notion of an effectively computable function, and Turing's analysis exactly captures this class of functions. (On Turing's Thesis and the related Church's Thesis, see Davis 1982.) This kind of claim is impossible to verify. Work like Turing's has a power of *creating* intuitions. As soon as we read it, we lose our previous innocence. Certainly people have had vague but suggestive ideas of mechanical computation for many centuries. Already in the thirteenth century Robert Kilwardby (1976, 62) said, "The art of mathematical algorithms is part of arithmetic proper; it is a practical and mechanical speciality of arithmetic." But we will see that there are plenty of non–Turing-computable functions that with hindsight one might count as Kilwardby-computable. In the years immediately before Turing's paper, other researchers working on the mathematical analysis of algorithms made assumptions completely at odds with Turing's. For example Hermann (1926) assumed that we can compute with the elements of any field (in other words, with any object whatever, concrete or abstract), but she demanded that we can calculate in advance a finite upper bound on the number of steps required in each computation. Herbrand, in a letter to Gödel in 1931, required that we can show by an intuitionistic proof that each function that we use in a computation has a unique calculable value at each argument (Gödel 2003, 14f.). Turing's work doesn't show that these conditions were wrong; the moral is only that there was no single intuition to be formalized. In fact, a large part of computability theory since Turing has been devoted to bringing out other intuitions.

A natural line of research is to find other mathematical characterizations of the class of computable functions. Some characterizations used other kinds of machine. Teachers of computability theory tend to like the *limited register machines* of Shepherdson and Sturgis (1963). A radically different kind of machine is a quantum computer (Nielsen and Chuang 2000); these machines use quantum theory to improve the efficiency of calculations dramatically, but the functions calculated are exactly the computable functions.

It makes sense to ask whether we can reach a different class of functions by formalizing what can be "physically calculated," not necessarily in the digital style of a Turing machine. For example, one can write down differential operators whose integral eigenvalues are not listed by any computable function, and maybe some of these operators have physical significance. See Pour-El and Richards (1989) for more on this theme. Gandy (1980) and Davies (2001) examine what can be computed by machines in a Newtonian universe.

One can also look for other purely mathematical characterizations of Turing's functions. There are three notions, any one of which is definable in terms of either of the other two. The first is computable functions; the second is *computably enumerable sets*, which are sets X such that some Turing machine outputs 1 if and only if it is given an input in X; the third is *computable sets*, which are sets X such that some Turing machine outputs 1 if it is given an input in X, and 0 if it is given an input that uses the appropriate alphabet but is not in X. A set is computable if and only if both it and its complement (within the set of strings from the appropriate alphabet) are computably enumerable. Post (1944) created a beautiful mathematical theory of computably enumerable sets; part of the beauty is that we can handle them securely without having to write out yards of machine code, as one generally has to do when working with Turing machines.

Computably enumerable sets appear in many parts of mathematics. One striking example is the theorem of Matiyasevich (1993): A set X of natural numbers is computably enumerable if and only if there is a finite set $E(x, y_1, \dots, y_n)$ of polynomial equations with integer coefficients and indeterminates $x, y_1, \dots, y_n$ such that a natural number n is in X if and only if the set of equations $E(n, y_1, \dots, y_n)$ has some solution in integers. (A set of natural numbers meeting this condition is said to be *diophantine*.) Matiyasevich's theorem gives a negative solution to Hilbert's Tenth Problem (Browder 1976): Find an algorithm to determine whether a given diophantine equation has a solution, or show that there is no such algorithm.

By implication, Turing's paper characterized two other notions besides "computable function"; it also characterized "algorithm" (viz. Turing machine, given by its set of instructions) and "computation" (viz. run of a Turing machine on a given input). Unfortunately these further characterizations are very bad from the point of view of computer science. Turing machine instructions are written in machine code; they owe much more to the architecture of a Turing machine than they do to the idea behind the algorithm. This puts them at the wrong level of abstraction. Kolmogorov and Uspensky (1958) made the first attempt to remedy this fault. Work continues; it seems necessary to distinguish between serial, parallel, and distributed computation. Relevant papers are Blass and Gurevich (2004) and Moschovakis (2001).

To return to Hermann: It seemed to her that one can meaningfully talk about calculations with elements of an arbitrary field. And so it has seemed to many other people. For example, Euclid's algorithm for polynomials of one variable doesn't require us to be able to write down the coefficients of the polynomials. The same applies to various well-known matrix algorithms, for example, Gaussian elimination. Probably the most general notion of computability in this direction is one that describes computation on a structure, counting the primitive operations of the structure as computable. Moschovakis (1974) writes in such a setting; he defines computability in terms of inductive definability, an idea that builds on Kleene's first recursion theorem.

If the entities that one computes with have some related structure, then computations can use this. For example computation with reals can use the arithmetic of the reals. Blum, Cucker, Shub, and Smale (1998) describe a notion of computation on real numbers that is meant to capture the "real number algorithms" of classical mathematics. Other mathematical entities with associated structure are functions and ordinals. Kleene made the breakthrough to computation on higher-type objects (functions, sets of sets, etc.) in a series of papers in the 1950s. Sacks (1990) studies notions of computing on higher-type objects and ordinals; see also Normann (1980). In the 1960s, Kripke and Platek independently showed that large parts of classical recursion theory generalize from ω to ordinals that satisfy certain closure conditions; these ordinals are known as *admissible ordinals*. Barwise (1975) was an influential exposition of this theory.

Constructive mathematicians of various hues have reconstructed various parts of classical mathematics so as to use only methods that they find acceptable. To some extent, classical mathematicians can simulate this kind of reconstruction by restricting themselves to structures whose domains and operations are recursive, either literally or in some generalized sense. In settings of this kind, some classical theorems survive and others don't, and it is often hard to predict which. One example was the recursive model theory initiated by Mal'tsev (1961). The two volumes of Ershov et al. (1998a, 1998b) form an encyclopedia of recursive mathematics.

8. Computability Theory II: Hierarchies and Degrees

Borel (1898, 46) introduced a hierarchy of subsets of the set $\mathbb{R}$ of real numbers. At the bottom level are the F sets; these are the topologically closed sets. The F_σ sets are the countable unions of F sets. Then $F_{\sigma\delta}$ sets are the countable intersections of F_σ sets. And so on, alternating countable unions and countable intersections. Taking unions at limit ordinals, one can iterate this construction into the transfinite ordinals, and new sets of reals appear at every step before ω_1. The sets gathered up in this way are called the *Borel sets*. They have many pleasant properties; for example, they are all analytic (see the end of this section), and all analytic sets are Lebesgue measurable.

The Borel hierarchy had a huge influence in logic. Within computability theory, we can imitate it with subsets of the set ω of natural numbers, provided that we take "computable" unions and intersections rather than arbitrary countable ones. The trick is to make the definition for n-ary relations, simultaneously for all n; then we drop a dimension each time we take a union or an intersection. To be precise, let C be the set of all computable relations on ω. The Σ^0_1 relations are those of the form

$$\{\,(x_0,\ldots,x_{n-1}) : \text{for some } x_n,\ (x_0,\ldots,x_n) \in R\,\},$$

where R is in C. Then the Π^0_2 relations are those of the form

$$\{\,(x_0,\ldots,x_{n-1}) : \text{for every } x_n,\ (x_0,\ldots,x_n) \in R\,\},$$

where R is in Σ^0_1; and so on, alternating "some" and "every." If we start with "every" and then take "some" and so on, we get the hierarchy Π^0_1, Σ^0_2, and so on. We write Δ^0_n for $\Pi^0_n \cap \Sigma^0_n$. One easily shows that both Π^0_m and Σ^0_m are subsets of Δ^0_{m+1}, for all $m \geqslant 0$, so that we have an interlocking double hierarchy. The relations that appear at some finite stage in this hierarchy are said to be *arithmetical.*

Kleene (1943) showed that the computably enumerable sets are the subsets of ω that are in Σ^0_1, and the computable sets are those in Δ^0_1. So the hierarchy generalizes the notion of computability: As n increases, the sets in Σ^0_n are in some sense less and less computable. As Kleene demonstrated, a diagonalization argument (as in the proof of Gödel's incompleteness theorem) shows that new relations enter the hierarchy at each level as n increases. It was natural to extend the hierarchy into the transfinite, like Borel's hierarchy. A problem was that one needed computable ways of climbing up to limit ordinals. (Church and Kleene 1937 first formulated this problem; today it spills over into proof theory.) The hierarchy gives out at a countable ordinal ω^c_1, known as *the least noncomputable ordinal.* The sets of natural numbers gathered up in the hierarchy before ω^c_1 are known as the *hyperarithmetical* sets. These sets reappear in several guises in higher recursion theory.

We can guess that if X and Y are arithmetical sets, and Y is higher in the hierarchy than X, then information about Y might allow one to calculate what is in X, but not vice versa. There are many ways of making this picture more precise, and each of them corresponds to some notion of "degrees." For example, suppose we design a machine that is like a Turing machine, except that it is allowed to consult an "oracle for Y" that, given a number k, will answer 1 if k is in Y and 0 otherwise. If some machine of this description will calculate for us whether or not any given natural number is in X, then we say that X is *Turing reducible* to Y, in symbols $X \leqslant_T Y$. We say two sets of natural numbers are *Turing equivalent* if each of them is Turing reducible to the other. This is an equivalence relation on the set of sets of natural numbers, and its equivalence classes are known as *Turing degrees.* Turing reducibility induces a partial ordering, in fact an upper semilattice, on the set of Turing degrees. One can refine the semilattice by using more delicate reducibilities; for example, we say X is *many-one reducible* to Y ($X \leqslant_m Y$) if there is a computable function f such that for all n,

$$n \in X \quad \text{if and only if} \quad f(n) \in Y.$$

The corresponding degrees are called *many-one degrees*. The diagonalization that led one up the arithmetical hierarchy yields for each degree d a degree d' called the *jump* of d; d' is strictly higher than d in the partial ordering $\leqslant_T$. All of this first appeared in Post (1944) and Kleene and Post (1954), and it constitutes the subject matter of *degree theory*, or more elaborately the *theory of degrees of unsolvability*.

Post (1944) raised the question whether all computably enumerable but noncomputable sets belong to the same Turing degree. This was the famous *Post's problem*, answered negatively in 1956 by the American Friedberg and the Russian Muchnik, both under the age of 20. These two researchers independently found the same ingenious argument known as the *priority method*. Their discovery provoked a torrent of work on the structure of the degrees, and of the computably enumerable degrees in particular. Shoenfield, Sacks, Yates, and Lachlan led the field. Conjectures of Shoenfield stimulated further work, and soon Lerman (1983) and Soare (1987) were able to report detailed descriptions of the degrees.

Around 1970 people began to study some much stricter types of reducibility, mostly defined in terms of Turing machines with severe limits on the space and time available to them. These allowed a classification of problems according to their *computational complexity*; see Garey and Johnson (1979). But complexity theory is generally assigned to combinatorics and theoretical computer science rather than logic.

In a different direction, the definition of the arithmetical hierarchy was adapted to consider relations in $\mathcal{P}(\omega)^m \times \omega^n$ for any finite m and n. Identifying real numbers with subsets of ω, one could define a computable hierarchy on the real numbers. It is known as the *computable Borel hierarchy*; we recover the original Borel hierarchy by allowing real number parameters. The hierarchy of sets of reals Σ^1_n, Π^1_n is defined like Σ^0_n, Π^0_n but with set quantifiers in place of number quantifiers. This hierarchy is known as the *light-face projective hierarchy*. The *bold-face projective hierarchy* $\boldsymbol{\Sigma}^1_n$, $\boldsymbol{\Pi}^1_n$ is the same but with real number parameters allowed. The sets in $\boldsymbol{\Sigma}^1_1$ are those known to the early descriptive set theorists as the *analytic sets*, and the Borel sets are exactly those in $\boldsymbol{\Delta}^1_1 = \boldsymbol{\Pi}^1_1 \cap \boldsymbol{\Sigma}^1_1$. These definitions put the resources of computability theory at the disposal of descriptive set theory. Moschovakis (1980) is an important text in this area.

Cooper (2003) is a general introduction to computability theory.

References

Aczel, Peter. 1988. *Non-well-founded Sets*. Stanford, Calif.: CSLI.

Artin, Emil. 1957. *Geometric Algebra*. New York: Interscience Publishers.

Barwise, Jon. 1975. *Admissible Sets and Structures: An Approach to Definability Theory*. Berlin: Springer-Verlag.

Barwise, Jon and Solomon Feferman. 1985. *Model-Theoretic Logics*. New York: Springer-Verlag.

Bell, John L. 1998. *A Primer of Infinitesimal Analysis*. Cambridge: Cambridge University Press.

Bernays, Paul, and Abraham Fraenkel. 1958. *Axiomatic Set Theory*. Amsterdam: North-Holland.

Birkhoff, Garrett. 1935. On the structure of abstract algebras. *Proceedings of Cambridge Philosophical Society* 31: 433–454.

Blass, Andreas, and Yuri Gurevich. 2004. Algorithms: a quest for absolute definitions. In *Currect Trends in Theoretical Computer Science: The Challenge of the New Century*. Vol. 2: *Formal Models and Semantics*, ed. G. Păun et al., 283–312. Singapore: World Scientific.

Blum, Lenore, Felipe Cucker, Michael Shub, and Steve Smale. 1998. *Complexity and Real Computation*. New York: Springer-Verlag.

Borel, Émile. 1898. *Leçons sur la Théorie des Fonctions*. Paris: Gauthier-Villars.

Bouscaren, Elisabeth, ed. 1998. *Model Theory and Algebraic Geometry: An Introduction to E. Hrushovski's Proof of the Geometric Mordell-Lang Conjecture*. Lecture Notes in Mathematics 1696. Berlin: Springer-Verlag.

Browder, Felix. 1976. *Mathematical Developments Arising from Hilbert Problems*. Providence, R.I.: American Mathematical Society.

Buechler, Steven. 1996. *Essential Stability Theory*. Berlin: Springer-Verlag.

Burnside, W. 1897. *Theory of Groups of Finite Order*. Cambridge: Cambridge University Press.

Carnap, Rudolf. 1935. Formalwissenschaft und Realwissenschaft. *Erkenntnis* 5: 30–37. English translation in Herbert Feigl and May Brodbeck, eds., *Readings in the Philosophy of Science*, 123–128. New York: Appleton-Century-Crofts, 1953.

Carnap, Rudolf. 1942. *Introduction to Semantics*. Cambridge, Mass.: Harvard University Press.

Chang, C. C., and Jerome Keisler. 1973. *Model Theory*. Amsterdam: North-Holland.

Church, Alonzo, and Stephen C. Kleene. 1937. Formal definitions in the theory of ordinal numbers. *Fundamenta Mathematicae* 28: 11–21.

Cohen, Paul. 1963. The independence of the continuum hypothesis, I. *Proceedings of the National Academy of Sciences* 50: 1143–1148.

Cooper, S. Barry. 2003. *Computability Theory*. Boca Raton, Fla.: Chapman and Hall.

Davies, E. Brian. 2001. Building infinite machines. *British Journal for Philosophy of Science* 52: 671–682.

Davis, Martin, ed. 1965. *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*. New York: Raven Press.

Davis, Martin. 1982. Why Gödel didn't have Church's thesis. *Information and Control* 54: 3–24.

De Morgan, Augustus. 1858. On the syllogism: III, and on logic in general. *Transactions of the Cambridge Philosophical Society* 10: 173–230. Reprinted in Augustus De Morgan, *On the Syllogism and Other Logical Writings*, ed. Peter Heath, 74–146. London: Routledge and Kegan Paul, 1966.

Devlin, Keith J. 1984. *Constructibility*. Berlin: Springer-Verlag.

Dodd, A. J. 1982. *The Core Model*. Cambridge: Cambridge University Press.

Ebbinghaus, Heinz-Dieter, and Jürg Flum. 1999. *Finite Model Theory*, 2nd ed. Berlin: Springer-Verlag.

Eklof, Paul. 1980. *Set Theoretic Methods in Homological Algebra and Abelian Groups*. Séminaire de Mathématiques Supérieures. Montreal: Université de Montréal.

Eklof, Paul and Alan Mekler. 1990. *Almost Free Modules: Set-Theoretic Methods*. Amsterdam: North-Holland.

Ershov, Y. L., S. S. Goncharov, A. Nerode, and J. B. Remmel. 1998a. *Handbook of Recursive Mathematics*. Vol. 1: *Recursive Model Theory*. New York: Elsevier.

Ershov, Y. L., S. S. Goncharov, A. Nerode, and J. B. Remmel. 1998b. *Handbook of Recursive Mathematics*. Vol. 2: *Recursive Algebra, Analysis and Combinatorics*. New York: Elsevier.

Forster, T. E. 1995. *Set Theory with a Universal Set*. Oxford: Clarendon Press.

Gandy, R. O. 1980. Church's thesis and principles for mechanisms. In *The Kleene Symposium*, eds. J. Barwise, J. J. Keisler, and K. Kunen, 123–145. Amsterdam: North-Holland.

Garey, Michael R., and David S. Johnson. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. New York: W. H. Freeman.

Gödel, Kurt. 1934. On undecidable propositions of formal mathematical systems. Reprinted with revisions in Davis (1965), 39–74.

Gödel, Kurt. 1940. *The Consistency of the Axiom of Choice and of the Generalized Continuum Hypothesis with the Axioms of Set Theory*. Princeton, N.J.: Princeton University Press.

Gödel, Kurt. 1947. What is Cantor's continuum problem? *American Mathematical Monthly* 54: 515–525. Errata 55: 151.

Gödel, Kurt. 1995. *Collected Works III: Unpublished Essays and Lectures*, ed. Solomon Feferman et al. New York: Oxford University Press.

Gödel, Kurt. 2003. *Collected Works V: Correspondence H–Z*, ed. Solomon Feferman et al. New York: Oxford University Press.

Halmos, Paul. 1974. *Naive Set Theory*. New York: Springer-Verlag.

Hausdorff, Felix. 1927. *Mengenlehre*, 2nd ed. Berlin: De Gruyter.

Hermann, Grete. 1926. Die Frage der endlichen vielen Schritte in der Theorie der Polynomideale. *Mathematische Annalen* 95: 736–788.

Hodges, Wilfrid. 1985. *Building Models by Games*. Cambridge: Cambridge University Press.

Hodges, Wilfrid. 1997. *A Shorter Model Theory*. Cambridge: Cambridge University Press.

Hodges, Wilfrid. 2004. What languages have Tarski truth definitions? *Annals of Pure and Applied Logic* 126: 93–113.

Jech, Thomas. 1986. *Multiple Forcing*. Cambridge: Cambridge University Press.

Jech, Thomas. 1997. *Set Theory*, 2nd ed. Berlin: Springer-Verlag.

Jensen, Ronald B. 1972. The fine structure of the constructible hierarchy. *Annals of Mathematical Logic* 4: 229–308.

Johnstone, P. T. 1977. *Topos Theory*. London: Academic Press.

Kanamori, Akihiro. 1994. *The Higher Infinite*. Berlin: Springer-Verlag.

Kaplansky, Irving. 1954. *Infinite Abelian Groups*. Ann Arbor: University of Michigan Press.

Kaye, Richard. 1991. *Models of Peano Arithmetic*. Oxford: Clarendon Press.

Kechris, Alexander S. 1995. *Classical Descriptive Set Theory*. Berlin: Springer-Verlag.

Kilwardby, Robert. 1976. *De Ortu Scientiarum*, tr. Albert G. Judy. London: British Academy.

Kleene, Stephen Cole. 1943. Recursive predicates and quantifiers. *Transactions of the American Mathematical Society* 53: 41–73. Reprinted in Davis (1965), 254–287.

Kleene, Stephen Cole. 1952. *Introduction to Metamathematics.* Amsterdam: North-Holland.

Kleene, Stephen Cole, and Emil Post. 1954. The upper semi-lattice of degrees of recursive unsolvability. *Annals of Mathematics* 59: 379–407.

Kolmogorov, Andrei N., and Vladimir A. Uspensky. 1958. On the definition of algorithm (Russian). *Uspekhi Mat. Nauk* 13(4): 3–28. Translation in *AMS Translations* 29 (1963), 217–245.

Lawvere, F. W. 1971. Quantifiers and sheaves. In *Actes du Congrès International des Mathématiciens, 1/10 Septembre 1970/NICE/France*, ed. Comité d'Organisation du Congrès, vol. 1, 329–334. Paris: Gauthier-Villars.

Lerman, Manuel. 1983. *Degrees of Unsolvability.* Berlin: Springer-Verlag.

Lindström, Per. 1969. On extensions of elementary logic. *Theoria* 35: 1–11.

Mac Lane, Saunders. 1971. *Categories for the Working Mathematician.* New York: Springer-Verlag.

Macintyre, Angus. 1971. On ω_1-categorical theories of fields. *Fundamenta Mathematicae* 71: 1–25.

Mal'cev, Anatoliĭ Ivanovič. 1971. *The Metamathematics of Algebraic Systems: Collected Papers 1936–1967.* Amsterdam: North-Holland.

Mal'tsev, Anatoliĭ. 1941. A general method for obtaining local theorems in group theory (Russian). *Ucheniye Zapiski Ivanov. Ped. Inst.* (*Fiz-Mat. Fakul'tet*) 1(1): 3–9. Translation in Mal'cev (1971), 15–21.

Mal'tsev, Anatoliĭ I. 1936. Untersuchungen aus dem Gebiete der mathematischen Logic. *Matemat. Sbornik* 1(43): 323–336. Translated as "Investigations in the realm of mathematical logic," in Mal'cev (1971), 1–14.

Mal'tsev, Anatoliĭ I. 1961. Constructive algebras (Russian). *Uspekhi Mat. Nauk* 16(3): 3–60. Translated in Mal'cev (1971), 148–214.

Mancosu, Paolo, Richard Zach, and Calixto Badesa. 2009. The development of mathematical logic from Russell to Tarski, 1900–1935. Chapter 9 in this volume.

Mansfield, Richard, and Galen Weitkamp. 1985. *Recursive Aspects of Descriptive Set Theory.* New York: Oxford University Press.

Marker, David. 2002. *Model Theory: An Introduction.* New York: Springer-Verlag.

Martin, Donald A. 1975. Borel determinacy. *Annals of Mathematics* 102: 363–371.

Matiyasevich, Yuri V. 1993. *Hilbert's Tenth Problem.* Cambridge, Mass.: MIT Press.

Morley, Michael. 1965. Categoricity in power. *Transactions of American Mathematical Society* 114: 514–538.

Moschovakis, Yiannis N. 1974. *Elementary Induction on Abstract Structures.* Amsterdam: North-Holland.

Moschovakis, Yiannis N. 1980. *Descriptive Set Theory.* Amsterdam: North-Holland.

Moschovakis, Yiannis N. 2001. What is an algorithm? In *Mathematics Unlimited*, eds. Björn Engquist and Wilfried Schmid, 919–936. Berlin: Springer-Verlag.

Mostowski, Andrzej. 1948. *Logika Matematyczna.* Wrocław: Monografie Matematyczne.

Mycielski, Jan, and Hugo Steinhaus. 1962. A mathematical axiom contradicting the axiom of choice. *Bulletin de l'Académie Polonaise des Sciences, Série des sciences mathématiques, astronomiques et physiques* 10: 1–3.

Nielsen, Michael A., and Isaac L. Chuang. 2000. *Quantum Computation and Quantum Information.* Cambridge: Cambridge University Press.

Normann, Dag. 1980. *Recursion on Countable Functionals.* Lecture Notes in Mathematics 811. Heidelberg: Springer-Verlag.

Paris, J. B., and L. A. Harrington. 1977. A mathematical incompleteness in Peano Arithmetic. In *Handbook of Mathematical Logic*, ed. K. J. Barwise, 1133–1142. Amsterdam: North-Holland.

Peano, Giuseppe. 1891. Principii di logica matematica. *Rivista di Matematica* 1: 1–10.

Peano, Giuseppe. 1896–1897. Studii di logica matematica. *Atti della Accademia delle scienze di Torino, Classe di scienze fisiche, matematiche e naturali* 32: 565–583.

Pieri, Mario. 1898. I principii della geometria di posizione composti in sistema logico deduttivo. *Mem. della R. Accad. delle Sci. di Torino*, 1–61.

Pillay, Anand. 1996. *Geometric Stability Theory*. Oxford: Clarendon Press.

Post, Emil. 1944. Recursively enumerable sets of positive integers and their decision problems. *Bulletin of the American Mathematical Society* 50: 284–316. Reprinted in Davis (1965), 305–337.

Pour-El, Marian Boykan, and J. Ian Richards. 1989. *Computability in Analysis and Physics*. Berlin: Springer-Verlag.

Quine, Willard Van Orman. 1937. New foundations for mathematical logic. *American Mathematical Monthly* 44: 70–80.

Reinhardt, William N. 1974. Remarks on reflection principles, large cardinals, and elementary embeddings. In *Axiomatic Set Theory, II*, ed. Thomas Jech, 189–205. Providence, R.I.: American Mathematical Society.

Robinson, Abraham. 1950. On the application of symbolic logic to algebra. In *Proceedings of the International Congress of Mathematicians, Cambridge, Mass., 1950*, vol. I, 686–694. Providence, R.I.: American Mathematical Society. Reprinted in Robinson (1979), 3–11.

Robinson, Abraham. 1979. *Selected Papers of Abraham Robinson 1: Model Theory and Algebra*, ed. H. J. Keisler. Amsterdam: North-Holland.

Rogers, Hartley Jr. 1967. *Theory of Recursive Functions and Effective Computability*. New York: McGraw-Hill.

Rubin, H. and J. E. Rubin. 1985. *Equivalents of the Axiom of Choice*. Amsterdam: North-Holland.

Russell, Bertrand. 1903. *The Principles of Mathematics*. London: George Allen and Unwin.

Sacks, G. E. 1990. *Higher Recursion Theory*. Heidelberg: Springer-Verlag.

Scott, Dana. 1961. Measurable cardinals and constructible sets. *Bulletin de l'Académie Polonaise des Sciences, Série des sciences mathématiques, astronomiques et physiques* 9: 521–524.

Shelah, Saharon. 1990. *Classification Theory*, 2nd ed. Amsterdam: North-Holland.

Shelah, Saharon. 1994. *Cardinal Arithmetic*. Oxford: Clarendon Press.

Shelah, Saharon. 1998. *Proper and Improper Forcing*, 2nd ed. Berlin: Springer-Verlag.

Shepherdson, J. C., and H. E. Sturgis. 1963. Computability of recursive functions. *Journal of the ACM* 10: 217–255.

Shoenfield, Joseph. 1967. *Mathematical Logic*. Reading, Mass.: Addison-Wesley.

Sieg, Wilfried. 1994. Mechanical procedures and mathematical experience. In *Mathematics and Mind*, ed. Alexander George, 71–117. Oxford: Oxford University Press.

Sierpiński, Wacław. 1934. *Hypothèse du Continu*. Warsaw: Fundusz Kultury Narodowej.

Sinaceur, Hourya. 2000. Address at the Princeton University Bicentennial Conference on Problems of Mathematics (December 17–19, 1946), by Alfred Tarski. *Bulletin of Symbolic Logic* 6: 1–44.

Soare, Robert I. 1987. *Recursively Enumerable Sets and Degrees: A Study of Computable Functions and Computably Generated Sets*. Berlin: Springer-Verlag.

Soare, Robert I. 1996. Computability and recursion. *Bulletin of Symbolic Logic* 2: 284–321.

Specker, E. P. 1953. The axiom of choice in Quine's "New foundations for mathematical logic." *Proceedings of National Academy of Sciences of the USA* 39: 972–975.

Tarski, Alfred. 1936. *O logice matematycznej i metodzie dedukcyjnej*. Lwów: Biblioteczka Mat. 3–5, Książnica-Atlas. English revised translation (1994) *Introduction to Logic and to the Methodology of the Deductive Sciences*. New York: Oxford University Press.

Tarski, Alfred. 1954. Contributions to the theory of models I. *Indagationes Mathematicae* 16: 572–581.

Tarski, Alfred. 1983. *Logic, Semantics, Metamathematics*, revised ed. John Corcoran. Indianapolis, Ind.: Hackett.

Tarski, Alfred, Andrzej Mostowski, and Raphael Robinson. 1953. *Undecidable Theories*. Amsterdam: North-Holland.

Tarski, Alfred and Robert Vaught. 1957. Arithmetical extensions of relational systems. *Compositio Mathematica* 13: 81–102.

Turing, Alan M. 1936. On computable numbers, with an application to the Entscheidungsproblem. In *Proceedings of the London Mathematical Society* 42, 230–265. Reprinted in Davis (1965), 116–154.

Van den Dries, Lou. 1998. *Tame Topology and O-Minimal Structures*. Cambridge: Cambridge University Press.

Vaught, Robert. 1961. Denumerable models of complete theories. In *Infinitistic Methods, Proceedings of the Symposium on the Foundations of Mathematics, Warsaw 1959*. Warsaw: Państwowe Wydawnictwo Naukowe.

Vitali, Giuseppe. 1905. *Sul problema della misura dei gruppi di punti di una retta*. Bologna: Gamberini e Parmeggiani.

von Plato, Jan. 2009. Chapter 11 in this volume.

Wagner, Frank O. 2000. *Simple Theories*. Dordrecht: Kluwer.

Woleński, Jan. 1989. *Logic and Philosophy in the Lvov-Warsaw School*. Dordrecht: Kluwer.

Woodin, W. Hugh. 1999. *The Axiom of Determinacy, Forcing Axioms, and the Nonstationary Ideal*. Berlin: De Gruyter.

Woodin, W. Hugh. 2001. The continuum hypothesis. *Notices of the American Society* 48: 567–576 and 681–690.

Zeman, Martin. 2002. *Inner Models and Large Cardinals*. Berlin: De Gruyter.

11

Proof Theory of Classical and Intuitionistic Logic

Jan von Plato

In this chapter, we are mainly concerned with the development of Gerhard Gentzen's (1909–1945) structural proof theory and its connections with intuitionism. The latter is important in proof theory for several reasons. First, the methods of Hilbert's old proof theory were limited to the "finitistic" ones. These methods proved to be insufficient, and they were extended by infinitistic principles that were still intuitionistically meaningful. It is a general tendency in proof theory to try to use as weak principles as possible. A second reason for the importance of intuitionism for proof theory is that the proof-theoretical study of intuitionistic logic has become a prominent chapter of logic, with applications in computer science. For reasons of space and time, we do not discuss the development of ordinal proof theory, or intuitionism beyond its logic. (A brief historical outline of the former can be found in Pohlers 1986.)

1. Hilbert's Old Proof Theory

Hilbert's old proof theory was based on the formalization of logical inferences through the axiomatic method. Formal proofs start with instances of the axioms of logic, and the number of rules of inference is kept as low as possible. The axiomatic method in logic was developed mainly by Frege, Peano, and Russell. It is almost impossible to use it in the actual construction of logical proofs, but it served the purposes of Hilbert's metamathematics as far as pure logic was concerned. The big questions of the consistency and completeness of the basic logical systems, those of classical propositional and predicate logic, were settled by 1930. The third big question, the existence of a decision method that answers if a formula is a theorem of logic or not, was cleared in axiomatic studies by 1936.

Intuitionistic logic, too, was studied by the axiomatic method. What the axioms are, was established by Arend Heyting (1930). As a curiosity, it can be pointed out that there is a characterization of intuitionistic logic within the algebraic tradition of Ernst Schröder that antedates Heyting's work by 11 years, namely, Thoralf Skolem's paper (1919). It gives the axioms of an algebraic structure, today called Heyting algebra, that characterizes intuitionistic logic in the same way as Boolean algebra characterizes classical logic.

The main open goal of Hilbert's proof theory around 1930 was the consistency problem of the arithmetic of natural numbers. Gödel's second incompleteness theorem (1931) showed that this goal cannot be reached by Hilbert's finitary methods. The result made it necessary to rewrite the aims of the Hilbert school, even in quite a concrete sense: At the time, Paul Bernays had almost finished the comprehensive presentation *Grundlagen der Mathematik*, but had to redo the work that came out much later in two parts, in 1934 and 1939 (see Hilbert and Bernays 1934, preface). Hilbert himself did not take part in this project, except for stating in a preface his disbelief in the finality of Gödel's results, a feeling not shared by Bernays or others of the Hilbert school.

2. Interpretation of Classical Logic in Intuitionistic Logic

Soon after Gödel's results, it was realized that the consistency problem of arithmetic is within the reach of intuitionistic reasoning. Gödel (1933a) himself and Gentzen (1933) arrived at this insight, through the *double-negation* translation of classical predicate logic into intuitionistic predicate logic. The first such translation was invented by Andrei Kolmogorov (1925) in a paper that itself is a predecessor to Heyting's axiomatization of intuitionistic logic. Kolmogorov's paper, written in Russian, remained unknown well until the 1930s, as shown by Gentzen's letter to Bernays in 1938 (see Menzler-Trott 2001, 79). The paper had, however, an indirect influence through Valeri Glivenko (1929), and in the case of Gödel, through the manuscript version of Heyting (1934) that Gödel read in the summer of 1932 (see Gödel 1995, 75, note y). Glivenko proved a theorem now bearing his name, stating that if a negated formula is provable in classical propositional logic, it is already provable in intuitionistic propositional logic. Only propositional but not predicate logic has this property. The connection between Glivenko's theorem and Kolmogorov's double-negation translation is close: By the latter, to each formula A, there is a formula of the form $\sim\sim A^*$ that is classically equivalent to A, and if A is classically provable, $\sim\sim A^*$ is intuitionistically provable. For propositional logic, one can have A^* equal to A. Now, if A is a classically provable negation, say, $\sim B$, then by above $\sim\sim\sim B$ is intuitionistically provable. Because $\sim\sim\sim B \supset \sim B$ is a theorem of intuitionistic logic, $\sim B$ is also intuitionistically provable.

In classical logic, the laws of excluded middle $A \vee \sim A$ and double negation $\sim\sim A \supset A$ are admitted. In intuitionistic logic, instead, these laws are not assumed. In particular, because $\exists x A \vee \sim\exists x A$ is not available, existence has always to be proved directly. The 1920s saw a bitter fight between Hilbert and Brouwer, the former defending classical existence, the latter intuitionism. Hilbert felt that intuitionism would "mutilate" existing mathematics into an insignificant fragment.

Gödel and Gentzen showed that the intuitionistic restrictions on existence proofs do not have the effect Hilbert feared as far as theories expressible in first-order logic are concerned. In independent work, they defined translations that give, for any formula A, a formula A^G such that their equivalence is provable in classical predicate logic. Furthermore, if A is a theorem of classical logic, A^G is a theorem of intuitionistic logic. The elementary arithmetic of the natural numbers is expressible in the language of predicate logic. If a contradiction is provable in classical arithmetic (Peano arithmetic), it is already provable in arithmetic based on intuitionistic predicate logic. (This latter came to be known as Heyting arithmetic.) Therefore, the consistency of classical arithmetic reduces to that of intuitionistic arithmetic. Gentzen's paper "On the relation between intuitionistic and classical arithmetic" was already typeset in 1933 when he heard of Gödel's paper with substantially the same result. Gentzen's withdrawal of publication was unfortunate, for a clarification on the pages of the *Mathematische Annalen* of the relation of classical and constructive existence proofs in arithmetic would have cleared many misunderstandings. Gödel in an unpublished lecture of 1941 states that "nothing at all is lost by dropping the law of excluded middle, but only the interpretation of the theorems has to be changed" (see Gödel 1995, 189–190). The essential difference instead concerns *impredicativity*: whether it is permitted to quantify over domains that have not been generated by some inductive process.

3. The Beginnings of Structural Proof Theory

With the reduction of classical into intuitionistic arithmetic in mind, Gentzen started in early 1932 a program for the solution of what was at the time the main problem of proof theory and foundations of mathematics, the consistency of classical arithmetic (see Gentzen's letter to Hellmuth Kneser in Menzler-Trott 2001, 35). Its first part consisted in the structural proof theory of classical and intuitionistic predicate logic, finished in Gentzen's doctoral thesis "Untersuchungen über das logische Schliessen" (Investigations into logical deduction) presented to the Göttingen faculty in June 1933 and published in two parts (Gentzen 1934–1935). The work was followed by Bernays, but due to the Nazi intrusion of the university, examination of the thesis was handled by Hermann Weyl, who also seems to have written the report for the faculty. (This document together with an explanation of the circumstances is found in Menzler-Trott 2001, 39ff.). The thesis already contained the first steps in the proof analysis of arithmetic. The arithmetic part of Gentzen's program was finished in late 1934 (letter to Kneser in Menzler-Trott 2001, 49) and published, after the main proof was revised through criticisms by Bernays, in

Gentzen (1936). The original proof can be found in Gentzen's collected papers (1969), and an account is given in Bernays (1970). Gentzen's last and greatest problem, the consistency of analysis, remains open today.

Gentzen (1934–1935) recast the presentation of logical inferences into a system of natural deduction. It is natural in the simple sense of "being a formalism as close to actual reasoning as possible" (Gentzen 1934–1935, introduction). The structural analysis of proofs through natural deduction was successful for intuitionistic logic, but for classical logic, Gentzen had to devise a more general logical calculus, known as sequent calculus.

Gentzen's main observation about "actual reasoning" in mathematics was that it is *hypothetical*, based on the making of assumptions. As an example of a formula to prove, consider $A \vee B \supset C \,\&\, D$. According to Gentzen, the natural way to proceed is to *assume* $A \vee B$ to be the case, and then to consider what can be done under this assumption. There are two possibilities. (1) A is the case: The task is to prove C from assumption A and D from assumption A. (2) B is the case: The task is to prove C from assumption B and D from assumption B. If all of this succeeds, we have proved $A \vee B \supset C \,\&\, D$ with no assumptions left open. Gentzen suggested that to each form of proposition, $A \,\&\, B$, $A \vee B$, $A \supset B$, and so on, corresponds a principle of proof: namely, one that gives the *sufficient* conditions for concluding the proposition. The rules for the propositions are, with a line separating the *premisses* (one or two) and the *conclusion* of the rule,

$$\frac{A \quad B}{A \,\&\, B}\,\&I \qquad \frac{A}{A \vee B}\,\vee I_1 \qquad \frac{B}{A \vee B}\,\vee I_2 \qquad \frac{\begin{matrix}[A]\\ \vdots \\ B\end{matrix}}{A \supset B}\,\supset I\,.$$

There are signs next to the inference line to indicate what rule has been applied. Rule $\&I$, *conjunction introduction*, tells that to prove $A \,\&\, B$, it is sufficient to have proved A and proved B. Next, to prove $A \vee B$, it is sufficient to have proved one of A and B. To prove $A \supset B$, it is sufficient to have proved B from the temporary assumption A. The square brackets indicate that the conclusion $A \supset B$ does not depend on the temporary assumption A that has been *discharged*. There are similar rules for the universal and existential quantifiers.

The object of logic, in Gentzen's view, is to study the general structure of proofs. It is a complete break with the logicist tradition of Frege, Peano, and Russell that Hilbert and his school had been pursuing and in which the notion of logical truth is basic.

Proofs that follow a precise set of rules are called *derivations*, to distinguish them from most of the informal proofs found in mathematics. The combination of several steps of inference produces a tree-like figure. Given a derivation with conclusion C, those top formulas in the derivation tree that have not been discharged, are the *open assumptions*. If there are none, the conclusion is a *theorem*.

In addition to the introduction rules for each of the connectives, Gentzen gives *elimination rules* that are reverses of sorts to the introduction rules, where the *major premisses* are the formulas with the connective:

$$\frac{A \,\&\, B}{A}\,\&E_1 \qquad \frac{A \,\&\, B}{B}\,\&E_2 \qquad \frac{A \lor B \quad \begin{matrix}[A]\\ \vdots \\ C\end{matrix} \quad \begin{matrix}[B]\\ \vdots \\ C\end{matrix}}{C}\,\lor E \qquad \frac{A \supset B \quad A}{B}\,\supset E\,.$$

The rule for disjunction elimination is entirely natural, and we used it already informally, where we had an assumption of the form $A \lor B$. When $A \lor B$ appears as the major premiss of an elimination step, there are two *cases* in the proof, A and B, and if both lead to the same conclusion C, that conclusion obtains irrespective of whether it was A or B that was the case. This is what the discharge brackets indicate in the rule. The elimination rule for implication is the same as *modus ponens*.

Gentzen's main result about natural deduction is what is today called the *normal form theorem*. For reasons to be explained soon, he only indicates the result by one example in (1934–1935, II.5.1 3). By the theorem, derivations in natural deduction can be transformed into a certain transparent form in which no step of introduction is followed by a corresponding elimination. As an example, assume there is an instance of $\&I$ followed by $\&E_1$:

$$\cfrac{\cfrac{\begin{matrix}\vdots\\ A\end{matrix} \quad \begin{matrix}\vdots\\ B\end{matrix}}{A \,\&\, B}\,\&I}{A}\,\&E_1\,.$$

This can be *converted* into the simpler form

$$\begin{matrix}\vdots\\ A\end{matrix}\,,$$

and similarly in the other cases. The first derivation has a part, the derivation of B, that has disappeared in the second derivation. Inspecting the rules of natural deduction, the formulas above the inference lines of the I-rules are all parts of the conclusion. Similarly, disregarding rule $\lor E$, which we set aside to not complicate matters, the conclusions of the E-rules are all parts of premisses. It can be shown that derivations in normal form contain no extraneous parts, but all formulas are parts of the conclusion or of the open assumptions. This is Gentzen's famous *subformula* property of derivations in normal form, the central tool in the analysis of proofs.

So far we have not shown any rules for negation. One way to handle negation in natural deduction, suggested by Gentzen (1934–1935, II.5.2) and followed by Dag Prawitz (1965), is to assume there is a proposition that is always false, denoted $\bot$. Now $\sim A$ can be taken to be just $A \supset \bot$, and negation introduction is a special case of implication introduction, with $B = \bot$. The rule of *ex falso* *quodlibet* is added: It is $\frac{\perp}{C}$ (from falsity, anything follows). If a contradiction such as $A \,\&\, {\sim}A$ is provable, by conjunction elimination also A and ${\sim}A$ are provable. Applying implication elimination to these two, $\perp$ is provable. Then it is provable by a derivation in normal form, and by the subformula property, all formulas in the derivation are parts of the conclusion $\perp$. But $\perp$ has no parts and there cannot be any such derivation, so that the consistency of the system of rules of natural deduction can be concluded.

Before turning to Gentzen's sequent calculus, we note the independent development of systems of natural deduction by Stanislaw Jaskowski (1934). This work contains no profound results on the structure of derivations, in contrast to Gentzen. In Jaskowski's systems, the formulas are arranged in a linear numbered succession, a device that has been followed in many pedagogical presentations of natural deduction. Gentzen himself had some doubts about the tree-like arrangement of formulas in his natural deduction derivations, thinking that such derivations "deviate from actual inference in which there necessarily is a linear sequence of propositions, caused by the linearity of thinking" (1934–1935, II.2.2). In the end of the 1930s, Gentzen found Jaskowski's work and considered its linear arrangement an improvement on the tree form (see Mentzler-Trott 2001, 41, note 6).

4. Sequent Calculus

In *sequent calculus* Gentzen found a beautiful way of expressing the principles of proof of classical logic. (Use of the word "sequent" as a noun was begun by Stephen Kleene 1952a, 441. Gentzen's "Sequenz" means "sequence.") A sequent is an expression of the form $\Gamma \to \Delta$, in which the *antecedent* Γ and *succedent* Δ are *lists* of formulas. An account of the origins of sequent calculus, in the work of Paul Hertz in the 1920s and in Gentzen's first paper (1932), can be found in Schroeder-Heister (2002).

If Δ has just one formula C, we have a *single succedent* sequent $\Gamma \to C$. In this case, the "Gentzen arrow" $\to$ can be taken as a notation that replaces the vertical dots in the natural deduction rules for $\supset I$ and $\vee E$. The former becomes: If $A \to B$, then $\to A \supset B$. For the latter, we have that if $A \to C$ and $B \to C$, then $A \vee B \to C$. Thus, sequent calculus formalizes the relation of derivability of a formula from other formulas. It also generalizes derivability into derivability of a finite number of cases instead of just one formula. With arbitrary finite lists of formulas $\Gamma, \Delta, \ldots$ on both sides of the derivability symbol $\to$, Gentzen's logical rules for his classical sequent calculus LK are

$$\frac{\Gamma \to \Delta, A \quad \Gamma \to \Delta, B}{\Gamma \to \Delta, A \,\&\, B}{\scriptstyle R\&} \qquad \frac{A, \Gamma \to \Delta}{A \,\&\, B, \Gamma \to \Delta}{\scriptstyle L\&_1} \qquad \frac{B, \Gamma \to \Delta}{A \,\&\, B, \Gamma \to \Delta}{\scriptstyle L\&_2}$$

$$\frac{A, \Gamma \to \Delta \quad B, \Gamma \to \Delta}{A \vee B, \Gamma \to \Delta}{\scriptstyle L\vee} \qquad \frac{\Gamma \to \Delta, A}{\Gamma \to \Delta, A \vee B}{\scriptstyle R\vee_1} \qquad \frac{\Gamma \to \Delta, B}{\Gamma \to \Delta, A \vee B}{\scriptstyle R\vee_2}$$

$$\frac{A, \Gamma \to \Delta, B}{\Gamma \to \Delta, A \supset B}\, R\supset \qquad \frac{\Gamma \to \Theta, A \quad B, \Delta \to \Lambda}{A \supset B, \Gamma, \Delta \to \Theta, \Lambda}\, L\supset \ .$$

The assumption of a formula A in natural deduction corresponds to the *initial sequent* $A \to A$ by which derivations start. (Nowadays such sequents are often called "logical axioms.") The rule of ex falso quodlibet can be given by letting derivations start also with sequents of the form $\bot \to \Delta$. Gentzen himself did not do this, but used rules for negation:

$$\frac{\Gamma \to \Delta, A}{\sim A, \Gamma \to \Delta}\, L\sim \qquad \frac{A, \Gamma \to \Delta}{\Gamma \to \Delta, \sim A}\, R\sim \ .$$

In Gentzen's thesis, the reading of sequents was not clear yet. Gentzen suggested a reading of $\Gamma \to \Delta$ as: The conjunction of formulas in Γ implies the disjunction of formulas in Δ. The reading in terms of derivability of a number of cases under open assumptions is suggested in Gentzen's second paper on the proof theory of arithmetic (1938).

In his 1936 paper on the consistency of arithmetic, Gentzen uses what is today known as "natural deduction in sequent calculus style," with sequents of the form $A_1, \ldots, A_m \to B$. The calculus has single succedent sequents, but the rules are those of natural deduction: Instead of the left rules of sequent calculus one has, with Γ and Δ lists of formulas, two rules of conjunction elimination concluding $\Gamma \to A$ and $\Gamma \to B$ from the premiss $\Gamma \to A \,\&\, B$, and a rule of implication elimination concluding $\Gamma, \Delta \to B$ from the two premisses $\Gamma \to A \supset B$ and $\Delta \to A$. In this calculus, says Gentzen, the sequents are "a formal expression of the meaning of *a proposition in a proof* in its dependency on some assumptions" (1936, 5.21). The passage points at the possibility of a proof-theoretical reading of single-succedent sequents, used for "indicating fully the meaning of a proposition as it occurs in a proof" (ibid., 5.1).

In Gentzen's second paper on the consistency of arithmetic, of 1938, multisuccedent sequents are used (1938, 1.2):

> In the previous work I had introduced the concept of a sequent, with just one succedent formula, in its immediate connection to the natural representation of mathematical proofs (1936, §5). It is possible to arrive at the new, symmetric concept of a sequent also from that same point of view, namely, by striving at a particularly natural representation of the *division into cases* (see §4 of the previous work, and in particular 5.26). Namely, a $\vee$-elimination can now be represented simply as: From $\to A \vee B$ one concludes $\to A, B$, to be read as: "Both possibilities A and B obtain."

Gentzen's suggestion is that a sequent $\Gamma \to \Delta$ gives a listing of the *open cases* Δ under the *open assumptions* Γ. Logical rules change and combine open assumptions and cases: For example, Gentzen's left conjunction rules replace the open assumption A or B by the open assumption $A \,\&\, B$, and his right disjunction rules change the open case A or B into the open case $A \vee B$, and so on. There can also be an *empty case* representing impossibility, with nothing on the right of the sequent arrow. Some care is needed in the above reading, for the open cases are to be understood classically: It need not be decidable which formula of Δ is the case.

Arnold Schmidt's review of Gentzen's thesis sheds some light on the interpretation of sequents. Schmidt writes (1935, 145) that a sequent $A_1, \ldots, A_m \rightarrow B_1, \ldots, B_n$ means that "B_1 *or* ... B_n depends on the assumptions A_1 *and* ... A_m." The B_j are referred to as "claims." The interpretation in terms of the truth of the implication $A_1 \,\&\, \cdots \,\&\, A_m \supset B_1 \vee \cdots \vee B_n$ is not explicitly given, but it is referred to as "the trivial interpretation" (ibid., 146). Schmidt was Gentzen's contemporary in Göttingen.

The rules of natural deduction follow the standard pattern of introductions and eliminations. There has been little development in these since Gentzen: Some generalizations of elimination rules have been proposed, as in Schroeder-Heister (1984), as well as different ways of handling the discharge of open assumptions, as in Prawitz (1965) and Leivant (1979). The introduction and elimination rules give intuitionistic logic, but the rule of double negation elimination that leads to classical logic is of a different character as emphasized by Gentzen (1934–1935, II.5.3). Prawitz (1965) was able to extend the normal form theorem from intuitionistic logic to that part of classical logic that does not use disjunction or existence. Normal derivations start with assumptions, followed by E-rules, then the classical rule of double negation or indirect proof, and last there follow I-rules.

On the side of sequent calculus, no comparable stability regarding the rules has been achieved. Gentzen's choice of his particular set of rules is dictated by the requirements set by the proof of his cut elimination theorem, the *Hauptsatz* (1934–1935, III.1.1). These rules have specific properties: In rule $L\supset$, the lists (*contexts* in today's terminology) Γ, Θ, Δ, Λ in the two premisses are *independent* and are added up in the conclusion. In the two other two-premiss rules $R\&$ and $L\vee$ instead, contexts are *shared*, the same in both premisses.

Gentzen designed the proof of the *Hauptsatz* so that its proof for his intuitionistic sequent calculus, denoted LJ, is directly a special case of the proof for the classical calculus LK. The calculus LJ is obtained from LK by putting on the rules the requirement that each succedent of a sequent consist of not more than one formula. It follows that there must be two right disjunction rules and that the left implication rule must have independent contexts, for otherwise there would be no single succedent instances of these rules. Gentzen mentions as a second design principle that sequent calculus must display the duality of & and $\vee$ (1934–1935, III.2.4). Because there must be two right disjunction rules, there must be dually two left conjunction rules. In fact, all the rules except $L\supset$ and $R\supset$ display the symmetry. Gentzen arranged the rules in his table (1934–1935, III.1) in two columns that are dual mirror images of each other. The structural rule of cut is self-symmetric. The implication rules are an exception to symmetry and they are given last, separated from the other rules. (Unfortunately the layout of the rules is not reproduced in the

English translation of Gentzen's papers.) Gentzen was very much struck by the left-right symmetries of the classical sequent calculus *LK*. The emergence of the symmetry in the case of the classical negation rules $L{\sim}$ and $R{\sim}$ made him exclaim (1938, 1.6): "The special position of negation, which makes for an annoying exception in the natural calculus, is lifted away as if by magic."

It is certain that Gentzen considered various forms of rules, as that would be the only way of arriving at the ones he had. For example, in his 1943 work on transfinite induction, written around 1940, Gentzen removed rule $L \supset$ by allowing instead also initial sequents of the form $A \supset B, A \rightarrow B$. Incidentally, this shows that the possiblity of invertible rules (see the following for this notion) had not occurred to Gentzen.

Next we consider the *structural* rules of Gentzen's sequent calculi. These rules are those of *weakening*, *contraction*, *exchange*, and *cut*. Weakening is the addition of superfluous assumptions: If $\Gamma \rightarrow \Delta$ is derivable, then $A, \Gamma \rightarrow \Delta$ is. Second, by the rule of contraction, one concludes from $A, A, \Gamma \rightarrow \Delta$ into $A, \Gamma \rightarrow \Delta$. Mirror image rules hold for the succedent parts. The rule of exchange permits us to change the order of formulas in the antecedent and succedent parts of a sequent. The most important structural rule is cut that can be informally motivated as follows.

Suppose that we have found a derivation of a result C from assumptions A, Δ, so that $A, \Delta \rightarrow C$. If next we find a derivation of A from assumptions Γ, these two are put together through the *rule of cut*:

$$\frac{\Gamma \rightarrow A \quad A, \Delta \rightarrow C}{\Gamma, \Delta \rightarrow C}\,{\scriptstyle Cut}\,.$$

Gentzen's *Hauptsatz* is the formulation of the normal form theorem within sequent calculus. It is proved by giving an algorithm for eliminating all applications of the cut rule in derivations. An inspection of the rules of sequent calculus shows that *all formulas* that appear in a cut-free derivation of a sequent $\Gamma \rightarrow \Delta$, are subformulas of Γ or Δ. Gentzen was able to prove the normal form theorem only for intuitionistic natural deduction, which is why he developed the multisuccedent sequent calculus and proved cut elimination for both intuitionistic and classical logic (see 1934–1935, introduction).

Gentzen's thesis gives the rules of sequent calculus in two groups, the structural rules of weakening, contraction, exchange, and cut as the first group, and the logical rules as the second. Two years later he calls the rules of weakening, contraction, exchange, and change of bound variable "Strukturänderungen," structural modifications (1936, 5.22). All of these latter except weakening "do not change the meaning of a sequent, ... all these possibilities of modification are of a purely *formal nature*. It is only because of special features of the formalism that these rules must be expressly given" (ibid., 5.244). Gentzen's view of the purely formal nature of structural rules should be based on a comparison between the situation in sequent calculus and in natural deduction. The latter has no explicit structural rules. He describes the left and right rules of sequent calculus as corresponding to the elimination and introduction rules of natural deduction, respectively, even if this correspondence is not quite perfect (1934–1935, III.1.1; see also Schmidt 1935, 145). The reason for this discrepancy lies in the form of rules $\&E$ and $\supset E$.

Gentzen's thesis contained some of the first results on intuitionistic logic. If a formula of the form $A \vee B$ is derivable, A or B already is derivable. This follows from the cut elimination theorem, because the last rule has to be $R\vee_1$ or $R\vee_2$. This *disjunction property* of intuitionistic logic was also mentioned in passing by Gödel (1932), without proof or even indication of how he knew the result to hold true. A related result of Gentzen is a proof of the underivability, by purely syntactic means of proof analysis, of the law of excluded middle.

The *Hauptsatz* also has as a corollary a positive solution to the decision problem of intuitionistic propositional logic. The proof is not immediate, but proceeds through a lemma showing that it is sufficient to apply the rule of contraction on any given formula not more than twice in Gentzen's intuitionistic calculus *LJ*. (Incidentally, Dosen 1987 shows that this limit is optimal.) With this restriction on contraction, it follows that the set of possible cut-free derivation trees of a given sequent is bounded.

5. Later Developments in Structural Proof Theory

The first contributions to structural proof theory by others than Gentzen come from the late 1930s. In 1944, Oiva Ketonen improved the rules of sequent calculus by replacing some of Gentzen's original rules for classical propositional logic so that all rules became *invertible*, meaning that if a sequent of the form of the conclusion is derivable, the sequent, or sequents, of the form of the premiss is also derivable. The changed rules are

$$\frac{A, B, \Gamma \rightarrow \Delta}{A \,\&\, B, \Gamma \rightarrow \Delta}\,L\& \qquad \frac{\Gamma \rightarrow \Delta, A, B}{\Gamma \rightarrow \Delta, A \vee B}\,R\vee \qquad \frac{\Gamma \rightarrow \Delta, A \quad B, \Gamma \rightarrow \Delta}{A \supset B, \Gamma \rightarrow \Delta}\,L\supset .$$

Each connective has just one left and right rule. Furthermore, all two-premiss rules have shared contexts, and Ketonen obtains a fully invertible classical propositional sequent calculus in which derivations are found by decomposing the endsequent in any order whatsoever: Given a sequent $\Gamma \rightarrow \Delta$ to be proved, choose from Γ or Δ any formula with a connective. The corresponding rule determines uniquely the premisses. Repeating this "root-first proof search," formulas are decomposed into parts until there is nothing to decompose. At this stage, it can be determined if the original sequent $\Gamma \rightarrow \Delta$ is provable or not, by controlling if all topsequents are initial sequents. The process *terminates* in the case of classical propositional logic in a bounded number of steps as determined by the number of connectives in the given sequent.

Ketonen's proof of invertibility of the rules used the structural rule of cut. Later Kurt Schütte (1950) and Haskell B. Curry (1963) gave direct proofs of invertibility, the latter with the explicit result that the inversions are *height preserving*: If a given sequent is derivable in at most n steps, the premisses in a rule that can conclude that sequent also have a derivation in at most n steps.

As noted, Gentzen (1934–1935) had proved the decidability of intuitionistic propositional logic. A direct terminating method of proof search for intuitionistic propositional logic, similar to Ketonen's method for the classical calculus, was found as late as around 1990, independently by Jörg Hudelmaier (1992) and Roy Dyckhoff (1992). These discoveries follow a line of research from Kleene (1952, 480ff., and 1952) who found a way of avoiding also Gentzen's structural rules of weakening and contraction in classical and intuitionistic sequent calculus. If the rule of cut were indispensable, one could try to derive a sequent $\Gamma \rightarrow \Delta$ from two premisses with an arbitrary cut formula. The rule of contraction has a similar effect: With it, root-first proof search could go on forever, by the multiplication of formulas in sequents. A minor modification concerns Gentzen's exchange rules that permit the change of order of formulas in sequents. These rules disappear through the use of "lists without order," or *multisets* as one says.

Research in sequent calculi from Gentzen on has led to the remarkable sequent calculi known as *G3*-calculi that have no structural rules. The logical rules of the classical calculus *G3c* are, in their propositional part, the same as Ketonen's. For intuitionistic logic, there are both single succedent and multisuccedent calculi *G3i* and *G3im*. The most remarkable property of these calculi is the height-preserving admissibility of contraction, meaning that if a sequent with a duplication of a formula is derivable, the sequent without the duplication is derivable and the derivation of the latter is not a bigger derivation than that of the former. This property and the exact form of the calculi *G3c* and *G3im* are due to Albert Dragalin (1988, Russian original 1979), and in the case of the intuitionistic single succedent calculus *G3i*, to Anne Troelstra in Troelstra and Schwichtenberg (1996). The *G3*-family of logical calculi offers the strongest known methods for the structural analysis of proofs.

Gentzen had shown the disjunction property of intuitionistic propositional logic. The related *existence property* follows easily through proof analysis in intuitionistic sequent calculus: If $\exists x A$ is derivable, there is some individual a such that the instance $A(a/x)$ is derivable. Thus, intuitionistic logic corresponds on a formal level to the constructive notion of existence. Both properties were generalized to hold also under assumptions if these do not contain, overtly or hidden, any disjunctive or existential assumptions, by Ron Harrop (1960).

6. The Computational Semantics of Intuitionistic Logic

In the time of the beginning of intuitionistic logic, around 1930, Heyting and Kolmogorov suggested an explanation of the principles of intuitionistic logic in terms of the notion of proof. Also Gentzen makes in his doctoral thesis the suggestion that the introduction rules give the meanings of the various forms of propositions in terms of provability. These rules make more precise Heyting's

(1931) discussion. In Kolmogorov (1932), it is suggested that intuitionistic logic is a logic of "problem solving": The atomic formulas express the primitive problems that have no logical structure. $A \,\&\, B$ expresses a problem that is solved by solving A and B separately, $A \vee B$ is solved by solving at least one of A and B, and $A \supset B$ is solved by *reducing* the solution of B to one of A. Falsity $\perp$ is an impossible problem that has no solution. In Kolmogorov's interpretation, the notion of a problem comes before the notion of a theorem: A theorem can be considered that special case of a problem in which the task is to find a proof. (A note in passing: The priority of problems versus theorems was a much debated question in the times of Pappus already. Kolmogorov's interpretation dissolves that ancient dispute.) In Heyting (1931, 1934), a very general explanation is given, in terms of Edmund Husserl's theory of intentional acts: The propositions of logic express expectations and proofs are acts that fulfill these expectations.

The idea of intuitionistic logic as a "logic of provability" is explicit in a one-page paper of Gödel's (1933b) in which he adds a modal provability operator to propositional logic and is able to interpret intuitionistic logic from a classical point of view.

The crucial point of the "BHK-interpretation" (for Brouwer, Heyting, and Kolmogorov) is the explanation of an implication. In Heyting's (1931) terms, $A \supset B$ is proved by devising a method that reduces a proof of B to one of A. In other words, given any proof of A, the method must give some proof of B. There were doubts about this explanation, because it purports to explain what a proof of an implication is by reference to an arbitrary proof, thus, possibly a proof of an implication. The publication of Gödel's collected papers has brought to light some of his lectures from the late 1930s, and a lecture "In what sense is intuitionistic logic constructive?" of 1941. Gödel develops as an alternative to the seemingly circular BHK-interpretation what is known as the functional or *Dialectica* interpretation of intuitionistic logic and arithmetic. (The latter name refers to the journal in which Gödel [1958] finally published his interpretation.)

If a formal system of proof is defined, such as that of natural deduction for intuitionistic logic, the foregoing problem about implication can be solved. There is a class of inductively defined finite formal derivations. The explanation of what an arbitrary derivation is consists of two parts. Following Michael Dummett (1975), derivations of a formula with an introduction rule can be called *canonical*, and other derivations *noncanonical*. This fixes the (outermost) form of canonical derivations. It is crucial that every noncanonical derivation reduce to a canonical one; this is accomplished by a normal form theorem.

The search for normal form theorems for various calculi has led to a hierarchy of notions of growing strength: The first notion is the *existence* of a normal form. In terms of sequent calculus, it corresponds to the *closure* of a system of rules under the rule of cut: If a sequent is derivable, there exists a derivation without the rule of cut. A result of this type can be proved by showing that the system is already complete without the rule of cut, so that its addition will not make more sequents derivable. Gentzen's (1934–1935) proof of the equivalence of *LK* without cut and a Hilbert-style axiomatic classical calculus proves closure under cut. Typically, later proofs of such results have used semantical methods. In the notion of *normalization* (resp. *cut elimination*), it is required that an algorithm be given for the transformation of a nonnormal derivation (derivation with cuts) into a normal (cut-free) one. As noted, Gentzen was able to prove normalization for intuitionistic natural deduction, even if he only illustrated the result by an example. Normalization proofs were first published by Prawitz (1965) and, in a somewhat more schematic form, by Andres Raggio (1965). *Strong normalization* requires that the conversion into normal (cut-free) form must terminate irrespective of the order in which the various nonnormalities (cuts) are eliminated. Typically, this property holds for calculi of natural deduction but fails for sequent calculi. The first proofs of strong normalization were given by Jean-Yves Girard (1971), Per Martin-Löf (1971), and Prawitz (1971). The relatively late appearance of these results has been explained by those involved by the comment that "before those times, nobody was interested in strong normalization" (Martin-Löf to the present author in 2001). A final notion in this order is *uniqueness* of normal form: For a given derivation, all conversion sequences terminate with the same normal derivation. We shall soon see where the interest in strong normalization and confluence came from, namely, the proof-theoretical notion of computation.

In 1969, William Howard made precise some of the ideas behind the BHK-interpretation, in his paper "The formulae-as-types notion of construction." The paper circulated as a manuscript and was finally published as Howard (1980). It established what came to be called the "Curry–Howard isomorphism" or "Curry–Howard correspondence." Curry's role was that he suggested the idea for implication in Curry and Feys (1958). Kleene's (1945) notion of *realizability* for intuitionistic arithmetic also anticipated the development (see also Kleene 1952a, §82). The basic idea of Curry and Howard is that a formula corresponds to the set of its proofs. More precisely, to each formula A there is the set of proofs of A in the sense of formal derivation. The notation $a : A$ stands for "a is a proof of A." In terms of sets, the reading is "a is an element of the set A." An introduction rule shows how to construct a proof from proofs of components: If $a : A$ and $b : B$, then $(a, b) : A \,\&\, B$. The operation of forming the *pair* (a, b) is the construction that gives a proof of the conjunction $A \,\&\, B$. If $a : A$, then $i(a) : A \vee B$, if $b : B$, then $j(b) : A \vee B$. The two operations indicated by i and j carry the information from which component of the disjunction the proof is constructed, proof of A or proof of B. Implication is more difficult: Assume an *arbitrary* proof x of A, so symbolically $x : A$. If you succeed in constructing from x a proof $b(x)$ of B, then the proof of $A \supset B$ is written as $(\lambda x)b$. This is the *lambda-abstract* of the expression $b(x)$ depending on the variable x, as invented by Alonzo Church (1932) (see also Church 1941 and Barendregt 1997). Set-theoretically, $A \,\&\, B$ is the Cartesian product of the sets A, B, and $A \vee B$ their (disjoint) union, and $A \supset B$ the set of functions from A to B.

As to the elimination rules, they show how to pass from an arbitrary proof of a formula to its components: If $x : A \,\&\, B$, then $p(x) : A$ and $q(x) : B$ are the projection constructions that do this. For disjunction, the rule is too complicated to be given here. For implication, using a suggestive symbol for a member of $A \supset B$, if $f : A \supset B$ and $x : A$, then $f(x) : B$. In terms of sets, a proof f of $A \supset B$ is a function f that transforms any proof x of A into some proof $f(x)$ of B. Thus, rule modus ponens is the same as the *application* of a function. Implication introduction, then, is *functional abstraction* as invented by Church. The rules of natural deduction become the rules of *typed lambda-calculus* under the Curry–Howard correspondence.

Truth of a formula A is established by a proof $a : A$. Thus, A *is true* corresponds to A being, when considered as a set, *nonempty*. Typed lambda-calculus shows the rules of intuitionistic natural deduction to be *sound* under the semantics given by the Curry–Howard correspondence. If the premisses of a rule are assumed true, each of them has an element, and the rules show how to construct an element of the conclusion that thereby also must be true.

In the natural representation of mathematical proofs, their characteristic form is that a claim B follows under some conditions A, and this is expressed concisely as the implication $A \supset B$ obtained by the rule of implication introduction. If at some stage the conditions A obtain, B follows by implication elimination. The Curry–Howard correspondence gives this latter step as a functional application: An argument $a : A$ is fed into the function $f : A \supset B$, and a value $f(a) : B$ is obtained. Reasoning constructively, without use of the classical law of excluded middle, the function $f : A \supset B$ is an *algorithm* or *computable function*. Gentzen's idea of normalization, which is basically the same thing as cut elimination, has the following specific meaning: Given the function $f : A \supset B$ and the argument $a : A$, normalization consists in the computation of the value of $f(a)$. The computation of the value of $f(a)$ is *the same* as the conversion of the nonnormal derivation into normal form, which makes apparent the importance of strong normalization and uniqueness. Formal proofs in intuitionistic natural deduction are computable functions. Constructivity, which used to be the philosophical principle behind intuitionistic logic and mathematics, now has the role of guaranteeing that computations do not go on indefinitely, but terminate after some bounded number of steps.

Formal proofs in the sense of the Curry–Howard correspondence are often called *proof-objects*. The idea of a functional hierarchy of such proof-objects can also be found in Nicolaas de Bruijn (1970). He also introduces the notion of *dependent types*, a notion found independently by Martin-Löf in 1970 (see his 1975). The Curry–Howard correspondence, or "propositions-as-sets" principle and the construction of dependent types are at the basis of Martin-Löf's *constructive type theory*. Dependent types are families of sets indexed by a set: Given a set A, $B(x)$ is a set for each $x : A$. Quantifiers in type theory use dependent types: They can be written as $(\forall x : A)B(x)$ and $(\exists x : A)B(x)$ ("for all x in A, $B(x)$; there is an x in A such that $B(x)$"). These quantifiers are *bounded* to a set of values A that can vary within one formula. Implication turns out to be a special case of bounded universal quantification and conjunction a special case of bounded existential quantification. These cases obtain when B is a constant set over A. The language with dependent types goes beyond first order logic in expressive power.

In recent decades, structural proof theory has found important applications in computer science. Sequent calculus is at the basis of logic programming (Prolog). Type theory, in turn, stems from natural deduction. In type theory, we can read a formal proof as a *program*. Then $a : A$ can be read as: Program a executes the task expressed by the proposition A. In particular, with $f : A \supset B$ we have a program f, and given an *input* a, the program computes an output $f(a)$ through normalization of $f(a)$ (see Martin-Löf 1982). The *correctness* of a computer program can now be controlled in exactly the same way in which the correctness of a formal mathematical proof can be controlled.

References

Barendregt, H. 1997. The impact of the lambda calculus in logic and in computer science. *Bulletin of Symbolic Logic* 2: 181–214.

de Bruijn, N. 1970. The mathematical language AUTOMATH, its usage and some of its extensions. As reprinted in R. Nederpelt et al., eds., *Selected Papers on Automath*, 73–100. Amsterdam: North-Holland.

Church, A. 1932. A set of postulates for the foundation of logic. *Annals of Mathematics* 33: 346–366.

Church, A. 1941. *The Calculi of Lambda-Conversion.* Princeton, N.J.: Princeton University Press.

Curry, H. 1963. *Foundations of Mathematical Logic*, as republished by Dover, New York, 1977.

Curry, H., and R. Feys. 1958. *Combinatory Logic*, vol. 1. Amsterdam: North-Holland.

Dosen, K. 1987. A note on Gentzen's decision procedure for intuitionistic propositional logic. *Zeitschrift für Mathematische Logik und Grundlagenforschung* 33: 453–456.

Dragalin, A. 1988. *Mathematical Intuitionism: Introduction to Proof Theory.* Providence, R.I.: American Mathematical Society. Russian original 1979.

Dummett, M. [1975] 1978. The philosophical basis of intuitionistic logic. As reprinted in Dummett, *Truth and Other Enigmas*, 215–247. London: Duckworth.

Dyckhoff, R. 1992. Contraction-free sequent calculi for intuitionistic logic. *Journal of Symbolic Logic* 57: 795–807.

Fenstad, J., ed. 1971. *Proceedings of the Second Scandinavian Logic Symposium.* Amsterdam: North-Holland.

Gentzen, G. 1932. Ueber die Existenz unabhängiger Axiomensysteme zu unendlichen Satzsysteme. *Mathematische Annalen* 107: 329–250.

Gentzen, G. 1933. On the relation between intuitionist and classical arithmetic. (Translation of unpublished German original.) In Gentzen (1969), 53–67.

Gentzen, G. 1934–1935. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift* 39: 176–210 and 405–431.

Gentzen, G. 1936. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen* 112: 493–565.

Gentzen, G. 1938. Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie. *Forschungen zur Logik und zur Grundlegung der exakten Wissenschaften* 4: 19–44.

Gentzen, G. 1943. Beweisbarkeit und Unbeweisbarkeit der Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie. *Mathematische Annalen* 120: 140–161.

Gentzen, G. 1969. *The Collected Papers of Gerhard Gentzen*, ed. M. Szabo. Amsterdam: North-Holland.

Girard, J.-Y. 1971. Une extension de l'interprétation de Gödel à l'analyse, et son application à l'elimination des coupures dans l'analyse et la théorie des types. In Fenstad (1971), 63–92.

Glivenko, V. 1929. Sur quelques points de la logique de M. Brouwer. *Académie Royale de Belgique, Bulletins de la Classe des Sciences* 15: 183–188.

Gödel, K. 1931. On formally undecidable propositions of *Principia mathematica* and related systems I. (Translation of German original.) In van Heijenoort (1967), 596–617.

Gödel. K. 1932. Zum intuitionistischen Aussagenkalkül. As reprinted in Gödel (1986), 222–225.

Gödel, K. 1933a. Zur intuitionistischen Arithmetik und Zahlentheorie. As reprinted in Gödel (1986), 286–295.

Gödel, K. 1933b. Eine Interpretation des intuitionistischen Aussagenkalküls. As reprinted in Gödel (1986), 300–303.

Gödel, K. 1941. In what sense is intuitionistic logic constructive? A lecture first published in Gödel (1995), 189–200.

Gödel, K. 1958. Ueber eine noch nicht benutzte Erweiterung des finiten Standpunktes. *Dialectica* 12: 280–287. Also in Gödel (1990), 240–251.

Gödel, K. 1986, 1990, 1995. *Collected Works*, vols. 1–3. New York: Oxford University Press.

Harrop, R. 1960. Concerning formulas of the type $A \to B \vee C$, $A \to (Ex)B(x)$ in intuitionistic formal systems. *Journal of Symbolic Logic* 25: 27–32.

van Heijenoort, J., ed. 1967. *From Frege to Gödel, A Source Book in Mathematical Logic, 1879–1931*. Cambridge, Mass.: Harvard University Press.

Heyting, A. 1930. Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der Preussischen Akademie von Wissenschaften, Physikalisch-mathematische Klasse*, 42–56.

Heyting, A. 1931. Die intuitionistische Grundlegung der Mathematik. *Erkenntnis* 2: 106–115.

Heyting, A. 1934. *Mathematische Grundlagenforschung, Intuitionismus, Beweistheorie*. Berlin: Springer.

Hilbert, D., and P. Bernays. 1934, 1939. *Grundlagen der Mathematik*, 2 vols. Berlin: Springer.

Howard, W. 1969. The formulae-as-types notion of construction. Published in 1980 in J. Seldin and J. Hindley, eds, *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, 480–490. New York: Academic Press.

Hudelmaier, J. 1992. Bounds for cut elimination in intuitionistic propositional logic. *Archive for Mathematical Logic* 31: 331–354.

Jaskowski, S. 1934. On the rules of supposition in formal logic. In S. McCall, ed, *Polish Logic 1920–1939*, 232–258. Oxford: Oxford University Press, 1967.

Ketonen, O. 1944. *Untersuchungen zum Prädikatenkalkül.* Annales Acad. Sci. Fenn, Ser. A.I. 23. Helsinki.

Kleene, S. 1945. On the interpretation of intuitionistic number theory. *Journal of Symbolic Logic* 10: 109–124.

Kleene, S. 1952a. *Introduction to Metamathematics.* Amsterdam: North-Holland.

Kleene, S. 1952b. Permutability of inferences in Gentzen's calculi LK and LJ. *Memoirs of the American Mathematical Society*, 10: 1–26.

Kolmogorov, A. 1925. On the principle of excluded middle. (Translation of Russian original.) In van Heijenoort (1967), 416–437.

Kolmogorov, A. 1932. Zur Deutung der intuitionistischen Logik. *Mathematische Zeitschrift* 35: 58–65.

Leivant, D. 1979. Assumption classes in natural deduction. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 25: 1–4.

Martin-Löf, P. 1971. Hauptsatz for the theory of species. In Fenstad (1971), 217–233.

Martin-Löf, P. 1975. An intuitionistic theory of types: predicative part. In H. Rose and J. Shepherson, eds., *Logic Colloquium '73*, 73–118. Amsterdam: North-Holland.

Martin-Löf, P. 1982. Constructive mathematics and computer programming. In L. Cohen et al., eds, *Logic, Methodology and Philosophy of Science IV*, 153–175. Amsterdam: North-Holland.

Menzler-Trott, E. 2001. *Gentzens Problem: mathematische Logik im nationalsozialistischen Deutschland.* Basel: Birkhäuser.

Pohlers, W. 1986. *Proof Theory.* Springer.

Prawitz, D. 1965. *Natural Deduction: A Proof-Theoretical Study.* Stockholm: Almqvist & Wicksell.

Prawitz, D. 1971. Ideas and results in proof theory. In Fenstad (1971), 235–308.

Raggio, A. 1965. Gentzen's Hauptsatz for the systems NI and NK. *Logique et analyse* 8: 91–100.

Schmidt, A. 1935. Review of Gentzen (1934–1935). *Zentralblatt für Mathematik und ihre Grenzgebiete* 10: 145–146.

Schroeder-Heister, P. 1984. A natural extension of natural deduction. *Journal of Symbolic Logic* 49: 1284–1300.

Schroeder-Heister, P. 2002. Resolution and the origins of structural reasoning: Early proof-theoretic ideas of Hertz and Gentzen. *Bulletin of Symbolic Logic* 8: 246–265.

Schütte, K. 1950. Schlussweisen-Kalküle der Prädikatenlogik. *Mathematische Annalen* 122: 47–65.

Skolem, T. 1919. Untersuchungen über die Axiome des Klassenkalküls und über Produktations- und Summationsprobleme, welche gewisse Klassen von Aussagen betreffen. As reprinted in Skolem (1970), 67–101.

Skolem, T. 1970. *Selected Works in Logic*, ed. J. E. Fenstad. Oslo: Universitetsforlaget.

Troelstra, A., and H. Schwichtenberg. 1996. *Basic Proof Theory.* Cambridge: Cambridge University Press.

12

Modal Logic from Kant to Possible Worlds Semantics

Tapio Korte, Ari Maunu, and Tuomo Aho

1. Kant's Theory of Judgment-Forms

Although in the *Critique of Pure Reason* Kant's definition of logic as "the science of the rules of the understanding in general" (A52=B76) is so wide that according to it logic comprises much of epistemology, he also has a narrower conception, which comes closer to our understanding of what logic is. Kant uses two criteria to divide logic into subclasses. He calls logic *general* (*die allgemeine Logik*) if it pays no attention to differences in objects of thinking; that is, if it contains only "the absolutely necessary rules of thought without which there can be no employment whatsoever of the understanding" (ibid.) and is therefore concerned only with forms of judgments and leaves their contents unattended. On the other hand, he calls logic *pure* (*die reine Logik*) if it does not contain empirical rules of thinking, namely, rules that follow from the psychological and physical conditions of thinking, but is concerned only with the a priori principles of the understanding (A52–53=B76–77). According to Kant, only logic that is general and pure can be considered as a proper science. *Nongeneral* logics are *Organa*, instruments or aids, of special sciences (A52=B76). Of *nonpure* logic, or applied logic (*die angewandte Logik*) as his term goes, Kant says that it can work only as a *kathartikon* for a corrupted understanding (A53=B78).

In agreement with the tradition of logic, Kant divides *pure general* logic[1] into theories of *concept* (*Begriff*), *judgment* (*Urtheil*), and *inference* (*Schluss*).

This chapter results from the collaboration of its writers in the following way: Korte is responsibe for sections 1–5, Maunu for sections 6–9 and 11–12, and Aho for sections 10 and 13 and some material in sections 7 and 9.

Kant argues that analogically the faculty of knowledge, or understanding in a broad sense of the word, can be divided into three parts, namely, to the faculties of *understanding*, *judgment*, and *reason* (A130–131=B169).[2]

From the standpoint of the whole project of Kant's *Critique*, the second of these faculties, the faculty of judgment, is the central one, since Kant holds that all acts of the understanding—in a broad sense of the word—are reducible to judgments and therefore, the understanding can be represented as a faculty of judgment (A69=B94). Consequently, although Kant does not hold that logic could be represented simply as a theory of judgment, it certainly forms the most important and original part of his logic. Furthermore, as we shall see in the sequel, it is Kant's theory of judgment that most affected the development of the nineteenth-century logic in Germany.

The main goal of Kant's theory of judgment seems to be to reveal the formal elements of judgments. According to this theory, every judgment consists of three parts, namely, two representations and a copula which connects them. The two representations Kant calls the *matter* (*Materie*) of a judgment. The copula represents its *form* (*Form*), namely, the way in which the representations are connected in a judgment (e.g., A266=B322; J, §18). The form of a judgment is analyzed into four titles (*Titeln*), namely, "Quantity," "Quality," "Relation," and "Modality," so that each title can have one of three possible values or moments (*Momente*). This gives the possibility of $3^4 = 81$ judgment-forms.[3] The following table represents the judgment-forms with appropriate examples.[4]

Quantity	universal	"Every A is B"
	particular	"Some A is B"
	singular	"The A is B"
Quality	affirmative	"A is B"
	negative	"A is not B"
	infinite	"A is not-B"
Relation[5]	categorical	"A is B"
	hypothetical	"If A then B"
	disjunctive	"A or B"
Modality	problematic	"A may be B"
	assertoric	"(It is true that) A is B"
	apodeictic	"A must be B"

One must be careful not to mix judgment-forms with grammatical forms, because no one-to-one relationship is supposed to obtain between them. For example, although the universal judgment-form is usually expressed with the grammatical form "every S is P," this is not always the case: "A bird is an animal" is an example of a sentence that can be used to express a universal judgment, although it is not of the mentioned form. Similarly, "every figure can be surrounded with twelve equilateral pentagons" expresses usually an apodeictic judgment, although its grammatical form suggests that it is a problematic one; "the soul must be immortal," on the other hand, can be taken to express a problematic judgment, although it looks like an apodeictic one.[6]

Explanations that Kant gives for Quantity and Quality require knowledge of some basic notions of traditional logic. He says that both in a singular and in a universal judgment the predicate can be asserted of the subject without exception, and this is because the subject of a singular judgment has no extension whatsoever (A71=B96). Kant's comment reflects his definition of the concept of extension (*Umfang*, *Sphäre*), according to which the extension of a concept consists, not of things that fall under it, as we are used to thinking, but of concepts which are subordinate to it. Later this became the standard definition of *Umfang* in nineteenth-century German logic and was accepted among others by such prominent mainstream logicians as Drobish (1887, §25), Ueberweg (1882, §53), and Sigwart (1872, §42).[7] Of Quality, Kant says that in an affirmative judgment the extension of the subject is thought "under the sphere of a predicate," in a negative one the subject "is put outside its sphere," and in an infinite judgment the subject is in the same way put to the complement of the sphere of the predicate (J, §22).[8]

Kant explains that Relation concerns the relationship between material elements of a judgment. The matter of a categorical judgment consists of two concepts, which are related to each other as a predicate to a subject. In a disjunctive judgment, two or more judgments, which constitute its matter, are in "a relation of logical opposition to each other" and together they "occupy the sphere of the proper cognition involved" (A73=B99). By "logical opposition," Kant means that the constituents of a disjunctive judgment exclude each other, and by "the sphere of the proper cognition" he seems to refer to the content of the disjunctive judgment as a whole. This seems to amount to nothing else than that a disjunctive judgment is an exclusive disjunction of two or more judgments. On the other hand, a disjunctive judgment was traditionally taken to be a disjunction of judgments that have either a common subject or, in rare cases, a common predicate. If this is also what Kant is after, his use of the word "sphere" in this connection becomes more understandable, because both "A is B or A is C" and "A is C or B is C" seem to have in some sense a disjunctively divided concept as the predicate or as the subject, respectively. Commonly used abbreviations like "A is B or C" and "A or B is C" reveal this more explicitly.

Kant's view on the nature of a hypothetical judgment is more obscure. He says that in a hypothetical judgment neither the antecedent nor the consequent is assumed to be true, but only the relation of consequence (*Consequenz*) between them (A73=B98–99). Kant is not very specific about the nature of this relation, but it is certain that the question is not about material implication. He says that a hypothetical judgment expresses the relation of a ground (*Grund*) to its consequence (*Folge*) (A73=B98). The relationship between a ground and a consequence is a multifaceted relation in Kant's philosophy and it is difficult to say what he means by it in this connection. However, the general feature of any kind of ground seems to be that the ground in some way determines (*bestimmt*) its consequence (see B112). It is not clear whether in a hypothetical judgment the consequent is determined logically, causally, or by some other means by the antecedent, but it is certain that the relation of consequence is thought by Kant to be stronger than material implication.

According to Kant, Modality differs from Quality, Quantity, and Relation in that moments of the former contribute "nothing to the content of the judgement (for, besides Quantity, Quality and Relation, there is nothing that constitutes the content of a judgement)" (A74=B100). Kant's claim is odd since according to him we get the judgment-forms by abstracting "from all content of judgement" (A70=B95), which seems to imply that none of the judgment-forms contribute to the content of the judgment. This suggests that Kant uses the term "content" equivocally. In the claim that judgment-forms are abstracted from all content, Kant seems to mean with "content" what he elsewhere calls "the matter" of a judgment. Then again, when he says that modal judgment-forms add nothing to the content of a judgment he cannot mean the matter of a judgment but rather that which he according to *Jäsche Logik* calls "*the thing* about which we judge" (J, §30n1), namely, that which is judged true in a judgment. Accordingly, when he now says that modal judgment-forms contribute nothing to the content of a judgment he must mean that modal judgment-forms are not ways in which the subject and predicate of a judgment are connected, but that "modality concerns only the value that the copula has in reference to thought as such" (A74=B100). This passage sounds as if modal concepts were, according to Kant, epistemic concepts,[9] so that modal judgment-forms represent ways in which judgments are related to the faculty of judgment in general. Closer scrutiny reveals, however, that this is not the case.

Kant gives the following definitions of modal judgment-forms in the *Critique* (A74–75=B100):

> *Problematic judgments* are those in which "affirmation or negation is taken as merely possible (optional)."
>
> *Assertoric judgments* are those in which "affirmation or negation is viewed as actual (true)."
>
> *Apodeictic judgments* are those in which "affirmation or negation is viewed ... as necessary."[10]

Kant specifies that by "possibility" he means in this connection "logical (which is not objective) possibility," with "truth" logical truth and with "necessity" logical necessity (A75–76=B101).

By calling modal notions "logical," Kant separates them from the concepts of *real* modality, which he defines in the Postulates of Empirical Thought (A218=B265–266):

1. That which agrees with formal conditions of experience, that is, with the conditions of intuition and concepts, is *possible*.

2. That which is bound up with the material conditions of experience, that is, with sensation, is *actual*.
3. That which in its connection with actual is determined in accordance with universal conditions of experience, is (that is, exists as) *necessary*.

Kant says that these postulates are needed, if the notions of possibility, actuality, and necessity "are not to have only logical significance, analytically expressing the form of thought, but are to refer to the actuality, or necessity of *things*" (A219=B267). A few pages later, he specifies that a concept is logically possible if it does not contradict itself and that only "the simple-minded" mix this with the concept of transcendental possibility of things (A244=B302). Accordingly, we have the following definitions of logical possibility, impossibility, and necessity:

A judgment is *logically possible*, iff it does not contain a contradiction;

A judgment is *logically impossible*, iff it contains a contradiction;

A judgment is *logically necessary*, iff its negation contains a contradiction.

By supplementing the definitions of modal judgment-forms with these explications of logical possibility and necessity, we get the following definitions of *problematic*, *assertoric*, and *apodeictic* judgments:

A judgment is *problematic*, iff it is viewed as containing no contradiction;

A judgment is *assertoric*, iff it is viewed as true;

A judgment is *apodeictic*, iff its denial is viewed as containing a contradiction.

It is noteworthy that although logical possibility, truth, and logical necessity might be objective properties of judgments, the corresponding modal forms are not. They represent an attitude that the judging subject takes toward a judgment. Accordingly, a judgment is problematic, assertoric, or apodeictic, if it is, as Kant puts it, *taken to be* or *viewed as being* logically possible, true, or logically necessary independently of what its logical nature in this respect really is.

On the basis of Kant's definitions, one might feel justified to think that problematic, assertoric, and apodeictic judgments are nothing but attributions of possibility, truth, and necessity to a judgment and that one might find such logical relations between them as are familiar to us in modern modal logic. This is not, however, the case. According to Kant, both the antecedent and the consequent of a hypothetical judgment are problematic judgments, since in a hypothetical judgment the truth of neither is asserted (A75=B100). Consequently, when somebody makes a problematic judgment, he does not only judge that the judgment in question does not contain a contradiction but also asserts that he does not take any other alethic modal attitude toward it; that is, in a problematic judgment the judging subject asserts that he or she withdraws from attributing truth, falsity, or necessity to the judgment.

Therefore it is not true in Kant's logic that assertoric or apodeictic judgments imply problematic ones, in the manner in which necessity and truth imply possibility in even the weakest systems of modern modal logic.

2. Theories of Judgment-Form after Kant

The basic idea of Kant's theory of judgment—the idea that judgment-forms can be analyzed into moments—maintained much of its charm until the latter part of the nineteenth century,[11] and only a few logicians before Frege were willing to abandon the schema altogether.[12] Still, one of the most conspicuous trends in nineteenth-century logic was the dissatisfaction with the details of Kant's theory of judgment.

Not all judgment-forms were taken to be equally important. Kant himself had already noticed that from the point of view of pure general logic there is no difference between singular and universal judgment-forms and between affirmative and infinite ones (A71–72=B96–97). In accordance with this, they were not taken to be separate judgment-forms in the traditional formal logic of the nineteenth century. The remaining four forms of Quality and Quantity, namely, affirmative, negative, universal, and particular, had, in contrast, quite a strong position. This was mainly due to the traditional theory of Aristotelian syllogistic, which recognized only these judgment-forms.

From the standpoint of the theory of inference, modal judgment-forms were perhaps among the least important. In spite of this, the theory of modal judgment-forms was a subject that every respected theory of logic had to deal with. Kant's definition, according to which modal judgment-forms concern the attitude we take to the logical status of contents of judgments, was quite generally taken to mean that modality has something to do with certainty or reliability. This conception is hinted at already in *Jäsche Logik*. According to it, Kant held that *opinion* is constituted of problematic judgments, *belief* of assertoric, and *knowledge* of apodeictic ones (J, IX). Friedrich Ueberweg, one of the most influential traditional logicians in nineteenth-century Germany, defined Modality along these lines. According to him, by using problematic judgment-form, we express our uncertainty about whether the content of the judgment matches reality, by assertoric judgments we express immediate certainty based on perception, and by apodeictic ones that the judgment can be proved to be true (Ueberweg 1882, §69). Perhaps not completely in line with these definitions, Ueberweg holds that apodeictic judgments entail both assertoric and problematic ones, and that assertoric judgments entail problematic ones (Ueberweg 1882, §98). He does not, however, put forward any special modal syllogisms, although he examines in detail how the modal form of a premise affects the modal form of a conclusion in immediate inferences (i.e., in one-premise inferences) (Ueberweg 1882, §82–98).

Although a psychological definition of Modality, like the one offered by Ueberweg, was perhaps considered to be the standard definition in the first half of the nineteenth century (see Lange 1877, 33), at the end of the century it was held mainly by such psychological logicians as Wundt (1880), Erdmann (1892), and Lipps (1893). Most mainstream logicians either tried to define modal notions in a more objective way or rejected the relevance of Modality to logic altogether. A Herbartian logician, Moritz Drobisch, is an example of the former attitude. He tied modal judgment-forms directly to notions of logical necessity and possibility so that according to him a judgment is apodeictic, if its negation contains a contradiction, and is problematic, if it isn't apodeictic and does not contain a contradiction (Drobisch 1887, §61–62). He also studied how in some inferences the modal form of the conclusion depends on modal forms of the premises. He did not, however, suggest any special modal syllogisms.

That modal judgment-forms can be reduced to other forms was first suggested by Adolf Trendelenburg in his widely celebrated *Logische Untersuchungen* (1862). According to him, apodeictic, problematic, and assertoric judgments express the same as corresponding universal, particular, and singular ones, respectively (Trendelenburg 1862, 258–259).[13] The idea was adopted by Friedrich Lange and Christoph Sigwart. Lange argues that the concepts of necessity and possibility are relevant in logic only under the framework of Aristotelian metaphysics and that outside it Modality coincides with Quantity (Lange 1877, 40). Sigwart argues that Kant's definition of a problematic judgment as a judgment which is taken to be only possible is contradictory, since an essential ingredient of any judgment is the acknowledgment of its truth (Sigwart 1873, §31n3). Moreover, he holds that we are not free to choose whether to make a judgment or not, and in this sense all judgments are necessary. Therefore, he argues, there is no real difference between assertoric and apodeictic judgments (Sigwart 1873, §31n4). Although he thereby excludes modal judgment-forms from logic, he does not reject alethic modal notions from logic altogether but suggests, presumably under Trendelenburg's influence,[14] that necessity in logic is commonly expressed with universal judgments and possibility with particular ones (Sigwart 1873, §33–34). We shall return to Sigwart's view shortly in connection with his suggestion that categorical judgments can be reduced to hypothetical ones.

3. The Relationship between Categorical and Hypothetical Judgments from Wolff to Herbart

Traditionally the theory of inference was split into two separate fields, of which one, Aristotelian syllogistic, dealt exclusively with categorical judgments, and the other, Stoic hypothetical syllogistic, dealt only with hypothetical judgments. The division is still visible in *Jäsche Logik* according to which Kant held that inferences are divided into categorical, hypothetical, and disjunctive ones (J, §60). Because of this important role of Relation, the question whether categorical and hypothetical judgment-forms really are separate or whether they can be reduced to each other or to some third form was an interesting question. The reduction would have made possible a more unified theory of inference by decreasing the number of fundamental inference schemas.

The suggestion that the division of judgments into hypothetical and categorical ones is not fundamental was probably familiar to every logician of the early nineteenth century who was sufficiently acquainted with the history of his own discipline. This is because Christian Wolff argues in the eighteenth century that from the point of view of logic there is no difference between hypothetical and categorical judgments (Wolff 1754, 3. cap. §6–7).[15] He explains that every statement (*Satz*) contains a predication and the ground or reason (*Grund*) for the predication. In universal judgments the ground belongs to the nature of the subject and in particular judgments it does not. For example, the ground for the predication of giving off heat is the warmth of a stone—a contingent property of a stone. This predication is explicitly, and its ground implicitly, contained in the particular statement

some stones give off heat.

On the other hand, because weight belongs to the nature of a stone, the predication of heaviness of stones is a universal statement:

every stone is heavy.

Wolff holds further that every particular statement can be transformed into a universal one by expressing the ground explicitly. The statement

every warm stone gives off heat

is the result of the universalization of the particular statement above. Finally, the transformation of categorical statements into hypothetical ones happens by putting the ground for the predication to the *antecedent*, and the predication itself to the *consequent* of the statement. Accordingly, the statements above can be transformed into

if a stone is warm then it gives off heat.

and

if a thing is made of stone then it is heavy.[16]

Kant was familiar with Wolff's proposal, because according to both *Jäsche Logik* and *Wiener Logik* Kant had explicitly commented on the question whether "it is easy to transform a hypothetical proposition [*Satz*] into a categorical one" (J, §25n2), that is, as *Wiener Logik* puts it, whether it is "the same if I say, All men are mortal, or, If something is a man, then it is mortal" (W, 934).

Kant, naturally, rejected the proposal. According to both *Jäsche Logik* and *Wiener Logik*, he sought to justify the difference between hypothetical and categorical judgments by referring to the observation that what is asserted in a hypothetical judgment is the truth, neither of the antecedent nor of the consequent, but only of their connection, the *consequential.* In a categorical judgment, Kant argues, not only the categorical connection between its subject and predicate but also the concepts themselves are asserted, that is, thought of under the category of Reality (J, §25; W, 936). In other words, Kant denies the transformation of hypothetical judgments into categorical ones or vice versa because he holds that universal categorical judgments entail the existence of their subject, that is, that they have an existential import. He argues that the judgment "every A is B" cannot be transformed into "if something is A then it is B" because only the former entails the judgment that A exists.

The reason Kant holds that universal categorical judgments have existential import is probably that it seems to follow from the principle *dictum de omni et nullo*—the cornerstone of the theory of syllogism. According to this principle, whatever holds generally of every member of some group holds separately of each of them, and whatever is true of no member of some group is not true of any of them. The dictum warranted, among other things, the inference *ad subalternatum*, namely, the inference from a universal categorical judgment "every A is B" to its particular counterpart "some A is B." To deny that universal categorical judgments have existential import seems to be to deny, in the same breath, the validity of this inference and therefore the principle *dictum de omni et nullo* as well. It is no wonder that the existential import of universal categorical judgments seemed to be beyond doubt. Bolzano, for example, states explicitly that the reason he does not accept transformations between hypothetical and categorical judgments is that the truth of a universal judgment, like "every griffin is a bird," presupposes that there are griffins, since, as Bolzano writes, "how else could logicians teach that from every universal proposition the particular one can be deduced, if not because the proposition 'every griffin is a bird' already includes the proposition 'some birds are griffins'?" (Bolzano 1837, §225n).[17]

The existential import of universal categorical judgments does not, however, follow directly from the inference of *ad subalternatum* and therefore not directly from *dictum de omni et nullo* either. It does so only on the assumption that particular categorical judgments have existential import, that is, on the assumption that particular categorical judgments are judgments of existence. Consequently, one could keep the principle *dictum de omni et nullo* and deny that universal categorical judgments have existential import by denying that particular judgments are existential judgments. This possibility was considered, for example, by Leibniz.[18]

In the nineteenth century, the celebrated philosopher Johann Friedrich Herbart used this strategy to argue that from the logical point of view there is no fundamental difference between hypothetical and categorical judgments. According to Herbart, both categorical and hypothetical judgments are combinations of two concepts (Herbart 1813, §52). The difference between them is that in a hypothetical judgment the constituents of the two connected concepts are expressed explicitly by using categorical judgments (Herbart 1813, §60). That is, sentences

if A is B, then C is D,

and

AB is CD,

are two ways to express one and the same judgment. This deprives universal categorical judgments of existential import, as Herbart explicitly notes (Herbart 1813, §53) and, through the inference *ad subalternatum*, from particular categorical judgments as well. For Herbart this is not a problem because for him the particularity of a categorical judgment means only that the extension (*Umfang*) of its subject is indeterminately limited (§56). Since the extension of a concept consists, according to Herbart, of concepts that are subordinate to it (§43), neither a particular categorical judgment nor its universal counterpart contains an assertion that some object falls under its subject term.

The obvious motivation for Herbart to hold that there is no logical difference between categorical and hypothetical judgments is that it makes possible to merge the theories of categorical and hypothetical syllogisms and to get a unified theory of inference. According to Herbart, the whole theory of syllogism is based on two hypothetico-categorical syllogisms he calls inferences *modo ponente*, namely,

A is B,
C is A,
thus C is B,

or, in a hypothetical form,

if D is E, then F is G,
if M is N, then D is E,
thus if M is N then F is G,

and *modo tollente*

A is B,
C is not B,
thus C is not A,

which in a hypothetical form is

if D is E, then F is G,
if M is N, then F is not G,
thus if M is N, then D is not E. (Herbart 1813, §65)

Although Herbart's general logical insight inspired some logicians to the extent that we can talk of a special Herbartian school of logic,[19] his view on the relationship between hypothetical and categorical judgments and on the unified theory of inference was not generally accepted. This is true even of Herbart's closest followers. Drobisch, for example, who was perhaps the best known and the most respected of the Herbartian logicians, did not accept Herbart's view on hypothetical and categorical judgments (Drobisch 1887, §41). By rejecting it, Drobisch gave up the possibility to continue Herbart's work on a unified theory of inference, although he otherwise had ambitions to systematize the theory of inference (see Drobisch 1887, §84–113).

4. Toward the Theory of Quantification: Sigwart on the Reduction of Categorical Judgments to Hypothetical Ones

It took some 60 years from the publication of the first edition of Herbart's *Lehrbuch* for the question of the relationship between hypothetical and categorical judgments to be taken seriously under study again. This happened in Christoph Sigwart's *Beiträge zur Lehre vom hypothetischen Urtheile* (1871). The main objective of Sigwart's paper is to explore through the history of the theories of a hypothetical judgment and to find a compromise between Kant's and Herbart's extreme views on the relationship between categorical and hypothetical judgments (Sigwart 1873, 248). According to Sigwart, there is, pace Herbart, a logical difference between categorical and hypothetical judgments, although the difference does not lie in the connection between the constituents of a judgment as Kant had held (Sigwart 1871, 60). The difference is more fundamental, since hypothetical judgments consist of categorical ones; in a categorical judgment two concepts or ideas are connected, whereas in a hypothetical judgment, two categorical judgments are connected. This is why we can never get rid of the categorical judgment-form by reducing it to the hypothetical one. In other words, hypothetical and categorical judgments are not only different but are not even comparable, since they are not, so to say, at the same level (ibid.).[20]

Although Sigwart partially agrees with Kant when he holds that there is a logical difference between categorical and hypothetical judgments, he agrees to some extent also with Herbart, since he holds that in some cases categorical judgments can be transformed into hypothetical ones. Sigwart divides categorical judgments into two classes. First, there are judgments in which the predicate is predicated of the determinate subject or subjects, as is the case when the judgment is singular or when the subject term is used as a common proper name of several objects (Sigwart 1871, 62). Sigwart's example of the latter kind of judgment is:

all the planets circle the sun from west to east.

In this, Sigwart argues, "planet" behaves like a common proper name of all the objects of the solar system that we have discovered and decided to call "planets." The expression of the judgment is in fact an abbreviation of the expressions "a_1 circles the sun," "a_2 circles the sun," and so on (ibid.). In this sense of "planet" there are no unknown planets.[21] Second, there are categorical judgments in which the predicate is predicated of every object falling under the subject concept. Sigwart's example of this type of a judgment is:

> planets are solid objects, which revolve around central bodies along constant orbits determined by gravity.

According to Sigwart's explanation, because in this case being a planet is the ground for the attribution of the predicate, "planet" has the role of a concept word and the predicate is attributed of every object which falls under it (Sigwart 1871, 63). In his *Logik*, two years after the publication of *Beiträge*, Sigwart calls the former type of universal judgments "empirically universal" (*empirisch allgemeine*) and the latter "unconditionally universal" (*unbedingt allgemeine*) (1873, §27n6).

Sigwart argues that unconditionally universal judgments can be transformed into hypothetical ones. This happens by dividing the original categorical predication in

$$\text{every } A \text{ is } B$$

or

$$\text{no } A \text{ is } B$$

into two predications, which are then connected by the relation of necessary consequence (Sigwart 1871, 42, 52, 62–63):

$$\text{if something is } A\text{, then it is } B\text{,}$$

or

$$\text{if something is } A\text{, then it is not } B\text{.}$$

With the word "something" we express, according to Sigwart, that the predications, which constitute the hypothetical judgment, have "a single undefined something" as a common subject (Sigwart 1871, 42). In his *Logik* Sigwart specifies the nature of this "indefinite something" and identifies it with variables as they are used in "equations of geometry" and in "formulas of algebra" (Sigwart 1873, §36n10).

By means of this innovation, Sigwart is able to rationalize the theory of inference. He anticipates the division of logic into propositional and predicate logic in that he separates inferences in which categorical judgments are not analyzed into parts (Sigwart 1873, §49), and those in which the parts of judgments are relevant (ibid., §53). He introduces his theory of inference starting from what we might call the theory of propositional inference and argues that it has only three fundamental rules, namely, *modus ponens*, *modus tollens*, and the rule

If A holds then M holds
If M holds then X holds
Therefore, if A holds then X holds.

To be able to reduce the rules of categorical inference to these hypothetical ones, he takes advantage of his idea on the hypothetical nature of universal judgments and adds the following rule of inference to the three purely hypothetical ones:

If something is A then it is B
C is A
Therefore, C is B.

Sigwart goes through several other rules of inference, such as medieval conversions, and although it is not clear whether he takes the rules to be derivable from the four mentioned ones or whether he takes them to be independent, it is obvious that he takes the mentioned rules as somehow fundamental. All the same, Sigwart argues that the traditional theory of categorical syllogisms, on which the theory of inference of mainstream German logic was still based in the beginning of the 1870s, represents unnecessary specialization, *überflüssig Specialisierung*, and can be reduced to these few hypothetical rules of inference (Sigwart 1873, §54).

Despite appearance, Sigwart's analysis of universal categorical judgments is not a theory of quantification in the proper sense of the word. What the analysis lacks is the universal quantifier; there is no overt expression of generality in Sigwart's analysis. Still, Sigwart holds that judgments of the form "if something is A then it is B" express generality. This is possible because, according to him, the relationship between the antecedent and consequent of a hypothetical judgment is that of a necessary consequence, that is, "if something is A then it is B" says that nothing can be A without being B, which is just another way of saying that every A is B (Sigwart 1871, 62–63; 1873, §36n8). In this way Sigwart puts into practise Trendelenburg's claim that there is a close connection between necessity and generality (Sigwart 1873, §33n9).

5. Frege's *Begriffsschrift*

Frege's *Begriffsschrift* came out in 1879. It contains a description of an axiomatic logical system, a logical language as Frege wants to see it (Frege 1879, IV), which consists of the signs for negation, material implication, identity, functions, variables, and universal quantification. For propositional logic, Frege introduces the following six axioms:

(A1) $p \rightarrow (q \rightarrow p)$[22]

(A2) $(p \rightarrow (q \rightarrow r)) \rightarrow ((p \rightarrow q) \rightarrow (p \rightarrow r))$

(A3) $(p \rightarrow (q \rightarrow r)) \rightarrow (q \rightarrow (p \rightarrow r))$

(A4) $(p \rightarrow q) \rightarrow (\sim q \rightarrow \sim p)$

(A5) $\sim\sim p \rightarrow p$

(A6) $p \rightarrow \sim\sim p$

and as the rule of inference *modus ponens*

(R1) $p \rightarrow q, p \vdash q$

For quantified predicate logic with identity, Frege adds three further axioms:

(A7) $\forall x Px \rightarrow Pa$

(A8) $a = a$

(A9) $a = b \rightarrow (Pa \rightarrow Pb)$

and the rule

(R2) $p \rightarrow Pa \vdash p \rightarrow \forall x Px$

Kantian judgment-forms have no place in the *Begriffsschrift*. Frege argues, first, that the division of judgments into universal and particular ones or into affirmative and negative ones is misleading, since it is not judgments that are universal, particular, affirmative, or negative but their contents. Second, Frege argues, the whole difference between categorical, hypothetical, and disjunctive judgments is merely grammatical and that in the *Begriffsschrift* no distinction between problematic, assertoric, and apodeictic judgments is made (Frege 1879, 4–5).

Frege's conception of Modality is in line with that of Trendelenburg, Sigwart, and Lange. He argues that an apodeictic judgment has the same conceptual content[23] as the corresponding assertoric one because the former suggests, but only suggests, that the judgment can be derived from some universal judgment. Because only the conceptual content of a judgment is relevant from the point of view of the *Begriffsschrift* there is no need to make a distinction between assertoric and apodeictic judgments (Frege 1879, 4). Of problematic judgments Frege says that they express either that the negation of a judgment cannot be derived from a general law or that the negation of a judgment is generally false (ibid.) and holds, presumably, that problematic judgments are like apodeictic ones in that they add nothing to the conceptual content of the judgment. Because every universal judgment can trivially be derived from some universal judgment, namely, from itself, it follows that according to Frege apodeictic and problematic judgments coincide with universal and particular ones, respectively.

The obvious reason for Frege's other claim—that the distinction between categorical and hypothetical judgments is only a grammatical one—is that he has in mind only universal and particular judgments. It is true that in Frege's logic there are no universal or particular categorical judgments, because he shows that universal and particular categorical judgments are in fact hypothetical ones. According to him, the sentence

$$\forall x(Px \rightarrow Qx)$$

expresses what the following natural language sentences do:

> If something has the property P it also has the property Q,
> Every P is Q,
> All P's are Q's. (Frege 1879, 23)

Correspondingly,

$$\sim\forall x(Px \rightarrow \sim Qx)$$

expresses the same as

Some P's are Q's

or

It is possible that P is Q. (Frege 1879, 24)

As explained, the connection between universal and hypothetical judgments was suggested by Sigwart eight years before Frege.[24] However, Sigwart's and Frege's analyses differ in two important respects. First, in Frege's analysis the connection between the antecedent and the consequent of a hypothetical judgment is a material implication, whereas Sigwart takes it to be a strict conditional. Second, Frege expresses generality explicitly with a quantifier, while Sigwart has nothing like that. These differences are interconnected. To recall, in Sigwart's analysis it is unnecessary to express generality explicitly, since due to the relation of necessary consequence between its antecedent and consequent the statement "if something is P then it is Q" says that from something being P it follows necessarily that it is Q as well; that is, it says that all P's are Q's without exception. This possibility is barred from Frege, because in his analysis the relation between the antecedent and the consequent of a hypothetical judgment is a material implication. He needs a quantifier to express generality.

Interestingly, although Frege uses material implication, he seems to think that when the antecedent and the consequent are tied up with a universal quantifier the sentence expresses not only generality but necessity as well. This is a direct consequence of Frege's view that apodeictic judgments coincide with universal ones. Accordingly, the difference between Sigwart's and Frege's analyses is that Sigwart expresses generality with necessary consequence, whereas Frege does the opposite and expresses necessary consequence with generality. This is possible because in the *Begriffsschrift* the domain of discourse is the universal class: "$\forall x(Px \rightarrow Qx)$" says that whatever is P is Q, which means that there cannot be a Q which is not also P (see, e.g., van Heijenoort 1967, 325).

The use of a quantifier to express necessary consequence is not limited to analyses of universal judgments. In §11 of *Begriffsschrift*, Frege explains that there are two different uses of letters in the *Begriffsschrift*: A letter can be used as a variable, which is explicitly tied with a quantifier or it can be used as a variable, that is implicitly tied with a quantifier that has the whole sentence in its scope. Consequently, there are no schematic letters or free variables in the *Begriffsschrift*. For example, letters in the first axiom of the *Begriffsschrift*,

$$a \rightarrow (b \rightarrow a)$$

are not schematic letters, but are used to abbreviate the following sentence:

$$\forall xy(y \rightarrow (x \rightarrow y))$$

Since Frege allowed quantification over functions and truth values as well as over objects, the same holds good for any occurrence of a letter, which is not explicitly tied with a quantifier, as is the case in the theorem (54),

$$a = a$$

or in the theorem (52):

$$a = b \rightarrow (Fa \rightarrow Fb)$$

In all these cases, italic letters, which look like schematic letters, are in fact variables implicitly tied with quantifiers, because they are abbreviations for

$$\forall x(x = x)$$

and

$$\forall Fxy(x = y \rightarrow (Fx \rightarrow Fy)$$

respectively.

One consequence is that sentential expressions of the *Begriffsschrift* are not schemas, unlike sentential expressions in every other pre-Fregean logical system, but full-blown sentences capable of being true or false. In this sense the *Begriffsschrift* is, as Frege is eager to point out, a *characteristica universalis*, a logical language, and not only a *calculus ratiocinator*, a logical calculus (Frege 1879, V).

Another consequence of this use of a quantifier is that in sentences of the *Begriffsschrift* the sign for material implication can be used to express logical consequence. For example, the theorem (52) says in post-Fregean terms that "$Fa \rightarrow Fb$" expresses the truth in every interpretation in which "$a = b$" expresses the truth, which means that "$a = b \rightarrow (Fa \rightarrow Fb)$" expresses the truth no matter how "a," "b," and "Fx" are interpreted. This means that despite the use of material implication, the relation of necessary consequence, in the sense of a logical consequence, has an expression in the *Begriffsschrift*.

6. C. I. Lewis and the Beginnings of Modern Modal Logic

Even though there are some anticipations in C. S. Peirce's "Gamma Systems" and also in the work of Hugh MacColl (in the series of papers between 1880 and 1906, and in MacColl 1906), modal logic, as contemporarily conceived, is generally taken to begin with C. I. Lewis's dissatisfaction with the notion of implication in the "algebra of logic," or the "calculus of propositions" (in Whitehead and Russell 1910, in particular). Lewis (1912, 1918) holds, quite plausibly, that the material implication ($\rightarrow$) is unfitting if we want to "represent the logical nexus of proof and demonstration" (Lewis 1918, 328); he therefore introduces his strict implication ($\Rightarrow$) better to reflect the "ordinary meaning of implication" (Lewis 1912, 359). Impressed by the formal calculus, he intends to devise a logical system in which "[the] meaning of implication is precisely that of ordinary inference and proof" (Lewis 1912, 359). Lewis notices that we use sentences such as "Matilda loves me implies that I am beloved," intending these as some kind of inferences; logical calculus would be "more useful," he says, if it attended to these as well (Lewis 1912, 358).

In view of what was brought up in the previous section, Lewis's claim that the new development in logic misguidedly neglects inferential aspects is unjustified, at least as far as Frege is concerned. This is because Lewis's sentence involving Matilda, for example, is formalizable in Frege's *Begriffsschrift* language (in effect) as "$\forall Lm(Lm \rightarrow \exists xLx)$," which appears to capture precisely what Lewis is after (for in our contemporary terms this amounts to saying that "$Lm \rightarrow \exists xLx$" is a theorem, or that "$\exists xLx$" may be inferred from "Lm"). Thus, had Frege's work been better known, or, as far as it was known, better understood, the development of modal logic might have begun right from Frege's *Begriffsschrift*, that is, several decades before Lewis's contribution.

In terms of the distinction that was to become standard later, Lewis (as well as Frege) seems to smuggle something belonging to the *metalanguage*, viz., the proof-theoretical consequence ($\vdash$), into the *object language*. That is, Lewis evidently interprets (what we are accustomed to write as) "$p \vdash q$" (which belongs, from our present viewpoint, to the metalanguage) as "$p \Rightarrow q$" (which belongs to his object language). He calls it *strict implication*: It is impossible that p is true and q false. Writing, as usual, the necessity operator as "$\Box$," the latter is "$\Box(p \rightarrow q)$," which we today, conscious of the distinction, accept as belonging to the (modal) object language. We have just introduced "$\vdash$," to be used in meta-statements such as "$\Box p \vdash \Box(p \rightarrow q)$," which Lewis would write rather as "$\Box p \Rightarrow (p \Rightarrow q)$."

Although Lewis's "Russellian" reading of "$p \rightarrow q$" as "p implies q" is liable to censure, the criticism that he (and Frege) fallaciously conflates the two levels of language seems somewhat anachronistic and thus unfair. It may even be argued that the Frege-Lewis attitude is more natural in that the object language/metalanguage distinction is artificial from the viewpoint of ordinary speakers. To separate "Matilda loves me" and "I am beloved," which are sentences in the "natural object language," from the perfectly colloquial "It follows from Matilda's loving me that I am beloved," which according to the standard contemporary approach belongs to a higher level of language—or, to put it frankly, to a *different* language—is to make a distinction which ordinary speakers certainly do not make.

Lewis's treatment, confused or not, opens up the important possibility of introducing "intra-sentential" modal operators. That is, even if we write "$p \vdash q$" as "$\Box(p \rightarrow q)$," we have only prefix formulas. But once we have the latter at hand we are naturally led into considering formulas such as "$p \rightarrow \Box(p \rightarrow q)$," which cannot be extracted directly from presentations of the form "$p \vdash q$."

In Lewis (1912, 1918) can be found many ideas that were to be important, even crucial, in the later development of modal logic, such as the notion of "possible situation" or "circumstance" (Lewis 1918, 333–336), some rudiments of the distinction between actualist and possibilist quantifiers, with considerations on nonexistents (Lewis 1918, 328–331), and some reflections on counterfactual conditionals (Lewis 1912, 358).

7. Proof-Theoretic Approach to Modal Logic

The early modal logicians got some semantic support from the many-valued logics that were developed by Jan Łukasiewicz, but mainly their systems were purely proof-theoretic in character.

The final result of Lewis's work with strict implication was Lewis and Langford (1932, especially appendix II), where the famous "Lewis systems" S1–S5 were axiomatized. It is noteworthy that the Lewisian logics do not use material implication at all; they are formulated purely intensionally. In these systems, the rules of inference are substitution of strict equivalents, *modus ponens* for strict implication, and adjunction (i.e., $A \,\&\, B$ may be inferred from A, B), and the basic axioms are as follows:

(B1) $p \,\&\, q \Rightarrow q \,\&\, p$

(B2) $p \,\&\, q \Rightarrow p$

(B3) $p \Rightarrow p \,\&\, p$

(B4) $(p \,\&\, q) \,\&\, r \Rightarrow p \,\&\, (q \,\&\, r)$

(B5) $p \Rightarrow {\sim}({\sim}p)$

(B6) $(p \Rightarrow q) \,\&\, (q \Rightarrow r) \Rightarrow (p \Rightarrow r)$

(B7) $p \,\&\, (p \Rightarrow q) \Rightarrow q$

These define S1. Lewis's own favorite was the comparatively weak system S2, got from S1 by adding just the axiom

(B8) $\Diamond(p \,\&\, q) \Rightarrow \Diamond p$

The first axioms (B1)–(B7) allow infinitely many modalities, that is, non-equivalent combinations of modal operators. In 1930, Oskar Becker suggested reduction rules that would lead to an efficient elimination of such iterations. It was natural that this idea found much support. Becker considered the reduction axioms

$$\Box p \Rightarrow \Box\Box p \tag{C10}$$

and

$$\Diamond p \Rightarrow \Box\Diamond p \tag{C11}$$

and Lewis discussed these as characteristic for his systems S4 and S5, respectively. In other words, S4 and S5 are generated by adding (C10) or (C11) to (B1)–(B7). However, he was not willing to accept them, preferring an unlimited multitude of modalities. (S4 contains only 12 modalities and S5 four—necessity, possibility, and their negations.) S5 is simplest, and it has remained by far the most popular alethic system of modal logic even today.

The presentations of propositional modal logics that are standard today follow Gödel's (1933) formulation of S4, which, unlike Lewis's "systems of strict implication," is built as an *extension* of the nonmodal propositional logic.[25] Such a turn makes the systems much simpler and more transparent. S5, for example, is defined as the smallest set of modal formulas which

(CC) contains every tautological formula of classical (propositional) logic;

(CMP) is closed under *modus ponens*: if p and $p \to q$ are in the set, so is q;

(CS) is closed under substitution: if p is in the set, so is $p[r/q]$ (i.e., a formula resulting from a uniform substitution of a formula r for a subformula q in p);

(CN) is closed under *necessitation*: if p is in the set, so is $\Box p$;

(AX) contains the following axioms (axiom schemas):

$$\Box p \,\&\, \Box(p \to q) \to \Box q \tag{K}$$

$$\Box p \to p \tag{T}$$

$$\Diamond p \to \Box\Diamond p \tag{5}$$

A formula p contained in a logic X is called a *theorem* of that logic—in symbols, $\vdash p$. If in a logic X a formula p can be obtained from a set of formulas S (i.e., by using formulas in S as further "axioms"), p is said to be *derivable* from (or a *proof-theoretic consequence* of) the set S in X—in symbols, $S \vdash p$.

Dropping from S5 the axiom (5)—the *characteristic axiom* of S5—results in the "minimal" modal logic, usually called the logic T, and by excluding (T) as well we have the logic K (which is the basis for all so-called *normal* logics). Other useful modal systems may be obtained by adding to K (or to T) other characteristic axioms, such as

(4)	$\Box p \rightarrow \Box\Box p$
(D)	$\Box p \rightarrow \Diamond p$
(B)	$p \rightarrow \Box\Diamond p$
(H)	$(\Diamond p \,\&\, \Diamond q) \rightarrow \Diamond(p \,\&\, q) \vee \Diamond(p \,\&\, \Diamond q) \vee \Diamond(\Diamond p \,\&\, q)$

Important modal systems include T, S4 (= T + (4)), S5 (= T + (5)), D (= K + (D)), and S4.3 (= T + (4) + (H)).[26]

8. Possible Worlds

It seems natural to most of us today to hold that things could be (could have been, could be in the future) otherwise than they actually are (were, will be). Even determinists who say that things could not really be otherwise (because of, say, earlier events and necessitating laws) should understand what is meant by "things could be otherwise," at least in the sense of there being a difference between a contradictory idea (like that of a married bachelor) and an idea that is as such consistent even if it were never to find an actualization (like, say, that of a 10-foot-tall bachelor).

The notion of possible worlds, as it is used in the prevailing modal logics, may be seen as arising from such pretheoretic intuition that things could be otherwise. It seems equally natural to speculate that there might be things that are necessary in the sense that they really could not be otherwise—that they would be as they are no matter how other things were. The last characterization is of course a pretheoretic counterpart to "holds in every possible world." It is often claimed that this depiction of necessity as truth in all possible worlds is Leibniz's idea. However, possible worlds are utilized in the explanation of modalities well before Leibniz, for example, by Duns Scotus and Luis de Molina. It must be said, however, that Leibniz deserves credit for employing the notion of alternative possible universes more clearly and systematically than his predecessors.

It is also customary to credit Rudolf Carnap with the introduction of this possible worlds account of necessity into modern modal logic (see, e.g., Bull and Segerberg 1984, 13). However, Lewis seems to have the ingredients of this possible worlds delineation at hand in the following passage (Lewis 1918, 333–336; writing Lewis's strict implication still as "⇒"):

> Any set of mutually consistent propositions may be said to define a "possible situation" or "case" or "state of affairs." And a proposition may be "true" of more than one such possible "situation"—may belong to more than one such set. ... In these terms, we can translate $p \Rightarrow q$ by "Any situation in which p should be true and q false is impossible." ... It is in the nature of an absurd proposition

> that it is not logically conceivable that it should be true under any possible circumstances.

However, Lewis does not take systematic advantage of his "situations" or "circumstances." Carnap does this (with his "state-descriptions"), and for this reason the attribution of a pioneering role to Carnap is not seriously misleading either.

Possible worlds are nowadays most commonly conceived as (conjunctions of) propositions (or states of affairs, or properties), rather than (sets of) sentences (as in Carnap). According to Alvin Plantinga (1974, 45), for instance, a state of affairs w is a possible world if it is maximal or complete in the sense that for every state of affairs s, whenever w obtains, s either obtains or does not obtain (but not both).

9. Possible Worlds Semantics

After the achievements in nonmodal model theory, it was natural to ask if something similar could be done in modal logic as well. As indicated, the model-theoretic approach to modal logic may be said to begin with Carnap (1942, 1946, 1947). According to Carnap, we may represent (what Leibniz called) possible worlds by means of sets of atomic sentences, or *state-descriptions*, as he calls such sets. Then, a sentence "p" is according Carnap "logically necessary" just in case it is true in all (relevant) state-descriptions. More precisely, according to *state-description semantics*, given a set S of state-descriptions, the *truth* of a sentence "p" with respect to a state-description D (contained in S) is defined as follows:

> Atomic "p" is true with respect to D iff "p" is contained in D.
>
> "Not-p" is true with respect to D iff "p" is not true with respect to D.
>
> "If p, then q" is true with respect to D iff "p" is not true with respect to D or "q" is true with respect to D.
>
> "Necessarily p" is true with respect to D iff for every state-description E in S, "p" is true with respect to E.

Carnap did not yet have the notion of *relations* between states of affairs. Prior was led to this idea in his work on temporal logic (1957). Soon after that, the full apparatus of present-day possible world semantics in the strict sense was developed. As the important pioneers of modal model theory, we should mention Kanger (1957), Hintikka (1957a, 1957b, 1963), and Kripke (1959, 1963a, 1963b). They took the step from the state-description semantics to explicit *possible worlds semantics*. In the latter, a *model* $M = \langle W, R, V \rangle$ is an ordered triple consisting of a nonempty set W (of *possible worlds*), a dyadic relation R (of *accessibility*, or *alternativity*, or *relative possibility*) on W, and a function V (a *valuation* or *interpretation* function), which assigns to each propositional letter (i.e., atom) a subset of W. Now, the *truth* of a formula p with respect to a possible world w (contained in W) in a model M, or in symbols $\langle M, w\rangle \vDash p$, is defined as follows:

$\langle M, w\rangle \vDash p$ iff w belongs to $V(p)$, if p is an atom,

$\langle M, w\rangle \vDash {\sim}p$ iff not $\langle M, w\rangle \vDash p$,

$\langle M, w\rangle \vDash p \rightarrow q$ iff $\langle M, w\rangle \vDash q$ or not $\langle M, w\rangle \vDash p$,

$\langle M, w\rangle \vDash \Box p$ iff for every u in W such that Rwu, $\langle M, u\rangle \vDash p$.

A formula p is said to be (i) true in a model M, or *M-true*, if it is true with respect to every possible world in M; (ii) *valid* (in symbols, $\vDash p$), if it is M-true for every model M; (iii) a *logical* (or *model-theoretic*, or *semantical*) *consequence* of a set of formulas S (in symbols, $S \vDash p$), if for every model M and world w (in M), if $\langle M, w\rangle \vDash q$ for every formula q in S, then $\langle M, w\rangle \vDash p$.

By conditioning the accessibility relation R appropriately, various modal systems can be obtained very conveniently from this "skeleton" (which in fact is the basic normal logic K in the sense that K consists precisely in the formulas that are valid according to the definitions above). For example, because the axiom (T) results from reflexivity ($\forall x Rxx$), (5) from Euclidity ($\forall xyz(Rxy \,\&\, Rxz \rightarrow Ryz)$), (4) from transitivity ($\forall xyz(Rxy \,\&\, Ryz \rightarrow Rxz)$), (D) from seriality ($\forall x \exists y Rxy$), (B) from symmetricity ($\forall xy(Rxy \rightarrow Ryx)$), and (H) from connectedness ($\forall xy(Rxy \vee Ryx)$), the system T is obtained from the skeleton above by requiring reflexivity of R, S5 by requiring reflexivity and Euclidity, S4 by requiring reflexivity and transitivity, and so on. In fact, S5 can be defined more simply by discarding the accessibility relation altogether and replacing the last truth clause above with

$$\langle M, w\rangle \vDash \Box p \text{ iff for every } u \text{ in } W,\ \langle M, u\rangle \vDash p.$$

Let us say that a formula p is *X-valid* ($\vDash_X p$), if it is valid for a system X according to the above definitions. The *soundness theorem* for a system X says that every formula contained in the X-logic is X-valid (or, more generally, if $S \vdash_X p$, then $S \vDash_X p$), and the *completeness theorem*, conversely, that every X-valid formula is contained in the X-logic (or, more generally, if $S \vDash_X p$, then $S \vdash_X p$).[27] The holding of both of these metatheorems means that the $\vDash$-relation (model-theoretic logical consequence) coincides perfectly with the $\vdash$-relation (proof-theoretic derivability), or that validity is in fact the same as theoremhood.[28]

10. Correspondence Theory

Possible worlds semantics grew quickly common in the 1960s. During its early years, completeness results were proved for numerous familiar systems. Later, this began to be seen in a mathematically generalized perspective: Those modal systems were, in fact, *axiomatizations* for the theory of certain types of accessibility relations (see Segerberg 1971). However, in 1974 it was shown that there exist even incomplete normal modal logics.

The search for semantic characterizations of modal systems in completeness results generated a whole new branch of modal logic. As van Benthem writes,

> it took modal logicians some time to realize that there are also direct semantic equivalences involved here, having nothing to do with deduction in modal logics. Indeed the whole present *Correspondence Theory* arose out of simple observations such as the following, made in the early seventies: The T-axiom $\Box p \rightarrow p$ is true in a Kripke frame $\langle W, R\rangle$ if and only if R is reflexive. Here, "true in a frame" means true in all worlds, under all assignments to the proposition letters. (van Benthem 1984, 168)

As indicated, the *modal* axiom T, $\Box A \rightarrow A$, is valid in the frame $\langle W, R\rangle$ if and only if the *nonmodal* axiom $\forall x Rxx$ concerning the reflexivity of R is true in W,[29] the axiom (5) holds if the relation is Euclidean, that is, $\forall xyz(Rxy \mathbin{\&} Rxz \rightarrow Ryz)$, and so on. Thus there is a *correspondence* between the modal and nonmodal axioms.

Correspondence theory was rapidly developed by logicians like Johan van Benthem and R. I. Goldblatt. It was shown that numerous modal systems, even very complex ones, can be equivalently defined in a semantical way by characterizing the accessibility relation R with sentences of first-order predicate logic. In this way, large sections of modal propositional logic were translated into the language of nonmodal logic. This achievement is by no means trivial, since a *complete* correspondence is not possible: It was shown (a) that for some modal axioms there is *no* corresponding condition on R, and (b) that, conversely, for some quite simple conditions on R there is *no* modal counterpart. Profound results could then be achieved concerning what classes of modal sentences do have nonmodal representations—already Sahlqvist (1975) proved this for a large class of sentences. Modal logics with first-order equivalents can then profit from the strong metalogical results of predicate logic.

Let us mention in brief two examples of the subtler methods used in technical matters, just to illustrate the abstract nature that the notion of a possible world has obtained. First, van Benthem introduced the concept of "bisimulation" between (frames and) models. Bisimilar worlds have the same truths, and this offers a useful test for when a first-order formula has no modal equivalent: possibly there are bisimilar worlds in two models, one fulfilling the formula, the other not. Second, proofs often become easier if there are only finitely many worlds, and in the 1970s it was proved that many modal systems could be characterized in terms of their soundness and completeness with respect to certain classes of *finite* frames. (However, this does not hold universally: The first example of modal logics requiring infinite frames was given in Makinson 1969.)

11. Modality and Quantification

Quantified modal logic (QML) began in the forties with the work of Ruth Barcan (1946) and Carnap (1946, 1947), and with Willard V. Quine's (1943) sort of preemptive doubts about the intelligibility of the combination of modality and quantification. Even though propositional modal logic (PML) is quite straightforward, systems of QML, indeed, involve disputable philosophical issues related to notions such as existence.[30] The simplest way to devise QML is just to add (classical) quantifiers directly on top of PML. One system resulting from this strategy, the *simplest quantified modal logic* (SQML), is presented in this section.[31] Some philosophical problems inherent in modal logics, and in QML in particular, are addressed in the next section.

An SQML model M is an ordered quadruple $\langle W, D, R, V\rangle$, where W (worlds) and R (accessibility) are as before, D is a nonempty set (a domain of *individuals*), and V is a valuation function, which assigns to each individual constant an individual contained in D and to each pair consisting in a world (in W) and an n-adic predicate constant an n-adic relation on D (i.e., a set of n-tuples of members of D). Calling a function f from variables to the domain D of M an M-*assignment*, the relation $\langle M, w, f\rangle \vDash p$, or "$p$ is true in M with respect to w (in W) under an assignment f" is defined as follows:

> $\langle M, w, f\rangle \vDash Pt_1 \ldots t_n$ iff $\langle d(t_1), \ldots, d(t_n)\rangle$ is contained in $V(w, P)$, where each t_I is a term (i.e., an individual constant or variable), and $d(t)$ is $V(t)$ if t is a constant, and is $f(t)$ if t is a variable;
>
> $\langle M, w, f\rangle \vDash t = t'$ iff $d(t) = d(t')$;
>
> $\langle M, w, f\rangle \vDash {\sim}p$ iff not $\langle M, w, f\rangle \vDash p$;
>
> $\langle M, w, f\rangle \vDash p \rightarrow q$ iff $\langle M, w, f\rangle \vDash q$ or not $\langle M, w, f\rangle \vDash p$;
>
> $\langle M, w, f\rangle \vDash \forall xp$ iff for every assignment g different from f at most in that $g(x) \neq f(x)$, $\langle M, w, g\rangle \vDash p$;
>
> $\langle M, w, f\rangle \vDash \Box p$ iff for every world u in W such that Rwu, $\langle M, u, f\rangle \vDash p$.

A formula p is said to be (i) M-true with respect to a world w, if $\langle M, w, f\rangle \vDash p$ for all assignments f; (ii) M-true (simply), if it is M-true with respect to every world; (iii) *valid* ($\vDash p$), if it is M-true for every model M (i.e., if $\langle M, w, f\rangle \vDash p$ for every M, f and w); and (iv) a *logical* (model-theoretic, semantical) *consequence* of a set of formulas S ($S \vDash p$), if for every model M, world w (in M) and assignment f, if $\langle M, w, f\rangle \vDash q$ for every formula q in S, then $\langle M, w, f\rangle \vDash p$.

As before, the important S5 system is obtained by dropping the accessibility relation R altogether and replacing the clause for $\Box p$ by

> $\langle M, w, f\rangle \vDash \Box p$ iff for every world u in W, $\langle M, u, f\rangle \vDash p$.

A proof-theory for S5 may be given by adding to the propositional S5 logic (provided above) the following closure principle CG, quantifier axioms UG and UI, and identity axioms Id and InI:

(CG) closure under *generalization*: if p is in the logic, so is $\forall xp$;

(UG) $\forall x(p \to q) \to (p \to \forall xq)$, for any variable x not free in p;

(UI) $\forall xp \to p[t/x]$, for any term t different from a variable that is bound in $p[t/x]$;

(Id) $x = x$, for any variable x;

(InI) $x = y \,\&\, p \to p\{y/x\}$, where $p\{y/x\}$ results from p by substituting y for some x's in p so that no such substituted y becomes bound in $p\{y/x\}$.

12. Philosophical Issues

Possible worlds semantics involves many controversial matters, most of which arise in QML. However, some Quinean general objections to modal logic and some ontological issues related to possible worlds themselves will be considered first.

The distinction between *de dicto* and *de re* readings of modal statements offers a convenient entry to Quine's objections to modal logic. Though several meanings have been attached to these terms, one quite classical and clear-cut distinction deserving to be called a *de dicto*/*de re* distinction may be introduced by means of definite descriptions in the following way. On Russell's (1905) widely accepted "contextual" analysis of definite descriptions, "The number of planets is greater than seven," for example, is to be analyzed (roughly) as "There is a unique number of planets, and this number is greater than seven," or, more precisely,

$$\exists x(Nx \,\&\, \forall y(Ny \to y = x) \,\&\, x > 7),$$

where "N" represents "(is) a number of planets." Then, the modal statement "The number of planets is necessarily greater than seven" has both the *de dicto* reading,

(DD) $$\Box\exists x(Nx \,\&\, \forall y(Ny \to y = x) \,\&\, x > 7),$$

which seems false, and the *de re* reading,

(DR) $$\exists x(Nx \,\&\, \forall y(Ny \to y = x) \,\&\, \Box(x > 7)),$$

which seems true. Quine's doubts concern especially *de re* statements. For one thing, he holds that *de re* statements such as DR are "meaningless" since "9" in "$\Box(9 > 7)$," for instance, is not "purely designative": "the context 'necessarily…' … is similar to the context of single quotes," such as "'9 > 7' is analytic" and "'Cicero' contains six letters" (Quine 1943, 123–124; see also, e.g., Quine 1953a). That is, the problem Quine sees in DR lies in the last conjunct which according to him is to be read along the lines "The sentence '$x > 7$' is necessarily true." This indeed does not make much sense. However, though it might be admitted that it is possible to construe modalities as predicates of sentences, this is by no means compelled. It seems that the issue comes down to some "basic intuitions" about how to understand modalities, and, especially, whether they are entirely of linguistic nature (or based entirely on relations between sentences), or whether there are necessities and possibilities that are not so strictly dependent on language.

This brings us to Quine's second, closely related objection to *de re* statements such as "$\exists x \Box (x > 7)$." This is that (even if they made sense) taking them as true commits one to "Aristotelian essentialism," or "the doctrine that some of the attributes of a thing (quite independently of the language in which the thing is referred to, if at all) may be essential to the thing, and others accidental" (Quine 1953a, 175–176). Quine finds this unacceptable on the basis that "necessity resides in the way in which we say things, and not in the things we talk about" (Quine 1953a, 176). He seeks to justify this position by presenting the following dilemma: "For, would 9, that is, the number of planets, be one of the numbers necessarily greater than 7? But such an affirmation would be at once true in the form ['9 is greater than 7'] and false in the form ['The number of planets is greater than 7']" (Quine 1943, 123–124). His conclusion is that "to be necessarily greater than 7 is not a trait of a number but depends on the manner of referring to the number" (Quine 1953b, 148). However, a believer in modal logic may at this point reply (or insist) that the *modalized predicate* "being necessarily greater that 7" ("$\Box(x > 7)$") is true of 9, no matter how this number is referred to. Due to Kripke (1971, 1972, 1980) and others, some forms of essentialism Quine found objectionable are now widely accepted as being quite innocuous.

Positions taken on the ontological status of possible worlds may be classified in many ways, one being a sort of imitation of the traditional division to nominalism, realism, and conceptualism (Haack 1978, 191). According to such division, possible worlds may be construed either as sets of sentences ("nominalism"; Carnap 1947; Hintikka 1962, 1969, 1975), or as real mind- and language-independent entities ("realism"; D. Lewis 1973, 1986), or (roughly) as ways in which the world could be conceived to be otherwise ("conceptualism"; Leibniz; Kripke 1971, 1972, 1980; Plantinga 1974; Stalnaker 1976).[32]

Turning then to issues pertinent to quantified modal logics in particular, another important ontological question concerns the status of individuals. According to *actualism* everything there is is actual (there are no merely possible things), whereas according to *possibilism* it may be said, in some sense or other, that there are also things which are not actual (things that only might exist). Actualists are partial to possible worlds semantics in which each world has its own domain (thought of as containing just the existents of that world), and quantifications are, accordingly, world-relative (so that "$\forall x p$," for instance, is true with respect to a world w just in case "p" holds for all individuals existing in w). In contrast, a possibilist semantics is typically one with only a single (fixed) domain (of all "possible individuals"), and with (at least apparent) quantification over possibilia. So, with its fixed domain SQML appears as a possibilist system.

The basic problem with the varying domains approach is that because "$\exists x(a = x)$" is a theorem, so should be, by the necessitation principle CN, "$\Box\exists x(a = x)$" as well. This will not do for an actualist. One way to resolve this difficulty is to drop (or weaken) the CN principle. This move, however, is implausible since it surely seems that "necessarily p" is a logical truth whenever "p" is. Another, much more popular way to deal with this problem is to deny theoremhood of "$\exists x(a = x)$," and thereby adopt *free logic*, or a logic without the classical assumption that all individual constants are referring. This can be accomplished, basically, by introducing the *existence predicate* "E" (defined by "$Ex \leftrightarrow \exists y(x = y)$" in some free systems, but primitive in others) and adding an existence proviso to the quantifier axiom UI so that the revised form is $\forall xp \,\&\, Et \rightarrow p[t/x]$.

Objections to possibilist systems have been concentrated on qualms about the posited existence of individuals that are only possible (for this involves quantifications over nonexistent or nonactual possibilia, which contravenes Quine's widely accepted thesis that quantifiers reflect ontological commitment), and, relatedly, on some specific formulas that are valid in these systems. Regarding the latter issue, all of the following are valid in SQML:

(NE) $\forall x\Box\exists y(x = y)$,

(NNE) $\Box\forall x\Box\exists y(x = y)$,

(BF) $\forall x\Box p \rightarrow \Box\forall xp$,

(CBF) $\Box\forall xp \rightarrow \forall x\Box p$.

Actualists regard these as offending because, for them, NE attributes necessary existence to everything, NNE says, what is even worse, that this is necessary, the *Barcan formula* BF, in turn, states, for example, that if it is so much as possible for something to have the perfections that have been said to belong to God, then there actually is something that possibly have these perfections, and, finally, CBF would allow us to derive NE from the unproblematic $\Box\forall x\exists y(x = y)$ ("Necessarily, for every individual there is an individual that is it"). However, these are just actualist readings of the formulas in question. On possibilist readings these formulas are, as such, not as objectionable as actualists suggest: On such a reading NE, for instance, states only the platitude that every possible individual is necessarily such that there is a possible individual that is it. It may be said that these disputed formulas mean something different for actualists and possibilists. Accordingly, actualists should perhaps focus their criticism directly on the possibilist tendency of positing possibilia.

These pressures have brought many of those who identify themselves as possibilists quite near to the positions of some of those who see themselves rather as actualists (and vice versa). On a compromising view there are really no possibilia, but they are, so to speak, represented in the domain of a world by existing objects of appropriate sorts. Thus Plantinga (1974, 1976), who appears as an actualist, posits abstract *individual essences* as proxies for individuals proper, and an individual not existing in a world is accounted for by its individual essence's not being actualized or exemplified there (while these essences themselves exist, as abstract objects, in every world). In a somewhat analogous manner, Linsky and Zalta (1994), who call themselves possibilists, postulate existing *contingently nonconcrete* abstract objects as surrogates for so-called nonexistents: A concrete object that only might have existed with respect to the actual world is not a merely possible individual but an existing contingently nonconcrete one (with respect to the actual world). Making this parallel between these sorts of actualists and possibilists is not to deny that there is much disagreement between these camps (which is testified by the fact that Linsky and Zalta [1994] argue against Plantinga, among others), but at a general level their basic antidote against both extreme actualism and extreme possibilism seems pretty much alike.

Due to the work of (Barcan) Marcus (1961) and Kripke (1971, 1972, 1980), and others, the widely (though not universally) accepted view is that proper names in natural languages are *rigid designators*. Rigid designation is often said to mean "reference to the same individual in every possible world in which that individual exists" but is perhaps more properly characterized by reference to an individual *independently* of possible worlds, as David Kaplan (1989, 493–497) in particular emphasizes. This latter conception is reflected in the valuations of SQML's individual constants: They are not functions from worlds to individuals but provide directly individuals. This means that (interpreted) individual constants of SQML are referring with respect to all worlds. Hardline actualists who accept such rigid designation must accordingly accept reference to nonexistents; "Kofi Annan," for instance, refers to Kofi Annan even with respect to "Annanless" worlds (to put the point in terms of an example from natural discourse). Assuming that rigidity in the sense of nonfunctionality is accepted, this speaks in favor of fixed domains approach and against modal model theory with genuinely varying domains.

It is held at some quarters that modal logic is plagued, in addition to all sorts of difficulties and disagreements already discussed, with the *problem of transworld identity*, or that of individuation of individuals through possible worlds. That is, modal logicians are urged to provide criteria for the sameness of individuals in different possible worlds. Answers to this challenge have ranged from denying (numerical) identity of individuals in different worlds altogether to ascribing to individuals an inexplicable *haecceity* ("thisness") that is said to secure their modal identity. As brought up by Kripke (1980), it appears that there are two perspectives to the interplay between worlds and individuals. On one, *world-centered* picture we first posit possible worlds, as it were, and then start looking for the "same individuals" in these worlds. According to the other, *individual-centered* perspective, in modal reflection we, having already fixed onto an individual, ask how it would be were some things different from what they actually are; for instance, if there were no United

Nations at all, then, obviously, Kofi Annan—that fixed individual—would not be the Secretary General of the UN either. Arguably, it is the adoption of the first perspective that generates doubts about transworld identity (and, subsequently, doubts about rigid designation).

13. Applications

Finally, we can list some applications of modern modal logic to philosophically interesting subjects. Gödel had hinted at the possibility of understanding the modal operator as deductive *provability*, and this thought was rediscovered in the 1960s when logicians started to examine provability by means of formal calculi. There has recently been much progress in these untypical logics (see Boolos 1993).

Another step of generalization was taken by von Wright (1951a), who suggested that the modality could be seen as a purely formal operator open to many different interpretations in natural language. One such interpretation is *deontic* logic, the most successful philosophical application of modern modal logic. It was anticipated by Mally in 1926 but, in its present form, it starts from von Wright (1951b). It discusses the logical features of obligation; the operator $\Box$ is written as O, Ought. In semantics, the accessible worlds are deontic alternatives, those which are in accordance with *norms*. Most axioms of alethic modalities do not hold for O; especially T, $\mathrm{O}p \rightarrow p$, is not correct, because all norms are not fulfilled. It has been replaced by D, $\mathrm{O}p \rightarrow {\sim}\mathrm{O}{\sim}p$. On the other hand, the axiom $\mathrm{O}(\mathrm{O}p \rightarrow p)$ has often been defended. Probably the most discussed issue in deontic logic concerns the apparent paradoxes this seemingly plausible axiom brings about. Many of them have been solved, but the so-called conditional obligations seem to require some complication to the simple possible worlds machinery.

Hintikka's *Knowledge and Belief* (1962) was the cornerstone of *epistemic* and *doxastic* logic, with operators K_a = "a knows," B_a = "a believes." The program was later generalized to other propositional attitudes. In semantics, the epistemic alternatives are all the worlds compatible with a's knowledge—those states of affairs that a's knowledge does not exclude. Modal logic is thus applicable. Hintikka first thought that propositional epistemic logic was simply S4 and doxastic logic K4. But this immediately leads to the problem of "logical omniscience": The agent must know all the consequences of his knowledge, and all logical truths. This problem has been hotly debated and various partial solutions have been proposed, but the situation still remains unclear. The root of the difficulties lies in the fact that an attitude sentence is not simply a modal clause but also a factual claim about the person. In any case, epistemic logic again became very popular because of the interest of computer scientists in the 1980s. A philosophically more interesting development than formal calculi may be in the speculation concerning objective and subject-bound quantification, starting from Hintikka (1975).

Notes

1. From now on, "logic" refers to pure general logic unless otherwise indicated.

2. The word "understanding" has two meanings in the *Critique of Pure Reason*. On the one hand Kant defines it very broadly as the "non-sensible faculty of knowledge" (A67=B92). In this sense, every cognitive faculty, save intuition, belongs to the understanding. On the other hand, he uses "understanding" also in a more restricted sense to mean the faculty of grasping nonintuitive representations, that is, concepts. Kant recognizes the equivocation, since he says that "general logic is constructed upon a ground plan which exactly coincides with the division of the higher faculties of knowledge. These are: understanding, judgment, and reason. In accordance with the Functions and order of these mental powers, which in current speech are comprehended under the general title of understanding, logic in its analytic deals with concepts, judgments, and inferences" (A130–131=B169).

3. Textual evidence does not reveal whether Kant accepted the possibility of all 81 judgment-forms. This might well be doubted, since it is not obvious, for example, whether hypothetical judgments can be of different Quantity. Some logicians of the nineteenth century held that every judgment has some moment from every four titles so that there are, for example, universal as well as particular hypothetical judgments (see, e.g., Drobisch 1887, §51).

4. Each judgment in the table has the moment of which it is supposed to be an example. Because every judgment has some moment from every title, the same judgments could have been used as examples of several moments. "Every A is B," for example, is not only a universal judgment but also an affirmative, categorical, and assertoric one.

5. Hypothetical and disjunctive judgment-forms do not seem to be in accordance with Kant's general characterization of the form of judgment as the relation between the two representations in a judgment, since both hypothetical and disjunctive judgments assert a relation between judgments or propositions and not between representations or concepts. From Kant's point of view, there is no such problem since according to him judgments are representations; they are representations of representations (A68–69=B93–94)

6. The last two examples are from Bolzano (1837, §191).

7. Sigwart separates the logical extension of a concept from the empirical extension of a name. He argues that the concept of man and that of bipedal animal without feathers are different concepts with different extensions, although as names "man" and "bipedal animal without feathers" can be used to refer exactly to the same objects (Sigwart 1873, §42n5).

8. According to Kant, from the standpoint of pure general logic singular judgments are universal ones and infinite judgments are negative ones. In the *Critique* Kant says that he separates them because there is a difference between them from the standpoint of transcendental logic (A71–73=B96–98). However, both singular and infinite judgment-forms occur also in *Jäsche Logik*, although it deals exclusively with pure general logic.

9. About epistemic and psychological theories of modal notions in the nineteenth century, see Haaparanta (1988).

10. The view presented in *Jäsche Logik* is essentially the same: "Judgements are either *problematic* or *assertoric* or *apodeictic*. The problematic ones are accompanied with the consciousness of the mere possibility of the judging, the assertoric ones with

the consciousness of its actuality, the apodeictic ones, finally, with the consciousness of its necessity" (J, §30).

11. As late as 1884, Wilhem Windelband wrote that "the revolution that nowadays is happening in logic is nowhere as visible as in the system of judgment-forms. Although the old scheme, which due to the support it got from Kant's authority dominated formal logic a century ago, is still traditionally followed, it is for every new presentation of the science of logic almost nothing else than an object of critique" (Windelband 1884, 167).

12. One of the few was Bernard Bolzano. He rejected Kant's theory of judgment-forms as early as 1837 (esp. §188–191). He had, however, little or no effect on the development of logic in the nineteenth century, since he was relatively unknown to logicians at that time.

13. The connection between universality and necessity is also hinted at by Kant. He writes in the *Critique* (B4) that "If ... a judgment is thought with strict universality, that is, in such manner that no exception is allowed as possible, it is not derived from experience, but is valid absolutely *a priori*."

14. Sigwart openly acknowledges his debt to Trendelenburg's *Logische Untersuchungen* (Sigwart 1873, VI).

15. Lambert's view in *Neues Organon* is essentially the same (Lambert 1764, §137).

16. Wolff's own example is "every triangle has three angles" and "if a space is delineated with three lines, then it has three angles" (Wolff 1754, 3. cap. §7).

17. Bolzano admitted, however, that in some cases a proposition that seems categorical, like "a golden mountain is bald," should be understood as a hypothetical judgment, "if a mountain were made of gold, it would be bald" (Bolzano 1837, §196).

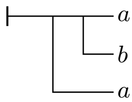
18. "An accepted conversion seems sometimes to lead to what is false. I have in mind the conversion *per accidens* of a universal affirmative proposition in a case such as 'Every laugher is a man, therefore some man is a laugher'. For the former is true even if no man laughs, whereas the latter is not true unless some man actually laughs; the former speaks of possibles, the latter of actuals. However, a difficulty of this kind does not occur if you remain within the limits of possibles: e.g. 'Every man is an animal, therefore some animal is a man'. It must therefore be said that the conclusion, 'Some man is a laugher', is true in the region of ideas, i.e., if you take 'laugher' for some species of possible entity, just as 'soldier' is a species of man; or, just as man is a species of animal, so some man is a laugher; the proposition will be true, even if no laugher exists" (Leibniz, G VII, 211).

19. The most famous Herbartian logicians were perhaps Moritz Drobisch (1802–1896), Franz Lott (1807–1874), and Robert Zimmermann (1824–1898).

20. In fact, Sigwart's view is more complicated than this. According to his definition, the assertion of truth or falsity is an essential ingredient of any judgment. Because in a hypothetical judgment neither its antecedent nor its consequent is asserted, Sigwart cannot hold that hypothetical judgments consist of categorical judgments (Sigwart 1871, 37). Instead, he says that it consists of two "categorical predications" (Sigwart 1871, 59–60).

21. In his *Logik* (1873) Sigwart argues that in this sense of "planet" there were only six planets in 1781 (§27n6).

22. In the cumbersome notation used in *Begriffsschrift*, the first axiom is



In the sequel only canonical notation is used.

23. Two judgments have the same conceptual content (*begriffliche Inhalt*) just in case exactly the same conclusions can be derived from them using the same additional premises (Frege 1879, 3).

24. It is known that Sigwart's logic was familiar to Frege. In Heinrich Scholtz's list of Frege's literary remains, there is mentioned a 19-page notebook titled "Logik von Dr. Christoph Sigwart." Unfortunately, it was lost in World War II (Veraart 1976, 103).

25. The most significant other writings from the 1930s to the 1950s include Parry (1939), McKinsey (1941), McKinsey and Tarski (1948), and Jónsson and Tarski (1951). For a more detailed account of this period, see Goldblatt (2003, sec. 3). The original idea of pure logic of entailment survived in numerous complicated systems that were later developed especially by A. R. Anderson.

26. For a more comprehensive catalog of modal systems (as well as an account of their strength with respect to each other), see, for instance, Garson (2003).

27. Of course, these theorems were already taken for granted by calling, for instance, T-valid formulas the logic T.

28. Sometimes the combination of completeness (as just defined) and soundness is called completeness.

29. A formula is said to be *valid* on a frame $\langle W, R\rangle$ iff it is true in every model $\langle W, R, V\rangle$.

30. See Garson (1984) for a useful overview of various systems of QML.

31. The primary source of the formulations to be given is Linsky and Zalta (1994).

32. The terminology in this area is not wholly fixed, for the last view is often called "moderate realism," in contradistinction to D. Lewis's "extreme realism."

References

Barcan, Ruth. 1946. A functional calculus of first order based on strict implication. *Journal of Symbolic Logic* 11(1): 1–16.

van Benthem, Johan. 1984. Correspondence theory. In *Handbook of Philosophical Logic*. Vol. II: *Extensions of Classical Logic*, eds. Dov M. Gabbay and Franz Guenthner, 167–247. Dordrecht: Reidel.

Bolzano, Bernard. [1837] 1929. *Wissenschaftslehre*. 4 vols. 2nd ed. Ed. Wolfgang Schultz. Leipzig: Verlag von Felix Meiner.

Boolos, George. 1993. *The Logic of Provability*. Cambridge: Cambridge University Press.

Bull, Robert, and Krister Segerberg. 1984. Basic modal logic. In *Handbook of Philosophical Logic*. Vol. II: *Extensions of Classical Logic*, eds. Dov M. Gabbay and Franz Guenthner, 1–88. Dordrecht: Reidel.

Carnap, Rudolf. [1942] 1948. *Introduction to Semantics*. Cambridge, Mass.: Harvard University Press.

Carnap, Rudolf. 1946. Modalities and quantification. *Journal of Symbolic Logic* 11(2): 33–64.

Carnap, Rudolf. 1947. *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.

Drobisch, Moritz Wilhelm. [1836] 1887. *Neue Darstellung der Logik nach ihren einfachsten Verhältnissen mit Rücksicht Mathematik un Naturwissenschaft*, 5th ed. Hamburg: L. Voss.

Erdmann, Benno. 1892. *Logik*. Halle: Max Niemeyer.

Frege, Gottlob. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Verlag von Louis Nebert.

Garson, James W. 1984. Quantification in modal logic. In *Handbook of Philosophical Logic*. Vol. II: *Extensions of Classical Logic*, eds. Dov M. Gabbay and Franz Guenthner, 249–307. Dordrecht: Reidel.

Garson, James W. 2003. Modal logic. In *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. At http://Plato.stanford.edu/archives/win2003/entries/logic-modal.

Gödel, Kurt. [1933] 1986. Eine Interpretation des intuitionistichen Aussagenkalküls. [*Ergebnisse eines mathematischen Kolloquiums* 4: 39–40.] Reprint, with an English translation (by John Dawson), in Kurt Gödel 1986, *Collected Works*, Vol. I, ed. Samuel Feferman et al., 300–303. Oxford: Oxford University Press.

Goldblatt, Robert. 2003. Mathematical modal logic: A view of its evolution. *Journal of Applied Logic* 1(5–6): 309–392.

Haack, Susan. 1978. *Philosophy of Logics*. Cambridge: Cambridge University Press.

Haaparanta, Leila. 1988. Frege and his German contemporaries on alethic modalities. In *Modern Modalities: Studies of the History of Modal Theories from Medieval Nominalism to Logical Positivism*, ed. S. Knuuttila, 239–274. Dordrecht: Kluwer.

van Heijenoort, Jean. 1967. Logic as calculus and logic as language. *Synthese* 17: 324–330.

Herbart, Johann Friedrich. [1813] 1912. *Lehrbuch zur Einleitung in die Philosophie*, 4th ed. Leipzig: Verlag von Felix Meiner.

Hintikka, Jaakko. 1957a. Quantifiers in deontic logic. *Societas Scientiarum Fennica, Commentationes Humanarum Literarum* 23(4): 3–23.

Hintikka, Jaakko. 1957b. Modality as referential multiplicity. *Ajatus* 20: 49–64.

Hintikka, Jaakko. 1962. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, N.Y.: Cornell University Press.

Hintikka, Jaakko. 1963. The modes of modality. *Acta Philosophica Fennica* 16: 65–82.

Hintikka, Jaakko. 1969. *Models for Modalities: Selected Essays*. Dordrecht: Reidel.

Hintikka, Jaakko. 1975. *The Intentions of Intentionality and Other New Models for Modalities*. Dordrecht: Reidel.

Jónsson, Bjarni, and Alfred Tarski. 1951. Boolean algebras with operators. Part I. *American Journal of Mathematics* 73(4): 891–939.

Kanger, Stig. 1957. Provability in logic. Ph.D. diss., University of Stockholm.

Kant, Immanuel (A/B). *Kritik der reinen Vernunft*. Riga 1781 (A), 1787 (B). In English, *Immanuel Kant's Critique of Pure Reason*. Translated by Norman Kemp Smith. Macmillan, 1989. (Originally published in 1929, second impression with corrections in 1931.)

Kant, Immanuel (J). 1923. *Jäsche Logik*. In *Kants gesammelte Schriften* (Preußischen Akademie der Wissenschaften) 9: 3–150. Berlin.

Kant, Immanuel (W). 1966. *Wiener Logik*. In *Kants gesammelte Schriften* (Deutschen Akademie der Wissenschaft) 24: 790–952. Berlin.

Kaplan, David. 1989. Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In *Themes from Kaplan*, eds. Joseph Almog, John Perry, and Howard K. Wettstein, 481–563. Oxford: Oxford University Press.

Kripke, Saul. 1959. A completeness theorem in modal logic. *Journal of Symbolic Logic* 24(1): 1–14.

Kripke, Saul. 1963a. Semantical considerations on modal logic. *Acta Philosophica Fennica* 16: 83–94.

Kripke, Saul. 1963b. Semantical analysis of modal logic I. Normal modal propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 9: 67–96.

Kripke, Saul. 1971. Identity and necessity. In *Identity and Individuation*, ed. Milton K. Munitz, 135–164. New York: New York University Press.

Kripke, Saul. 1972. Naming and necessity. In *Semantics of Natural Language*, eds. Donald Davidson and Gilbert Harman, 253–355, 763–769. Dordrecht: Reidel.

Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, Mass.: Harvard University Press. [A reprint, with a substantial introduction (pp. 1–21) and some minor revisions, of Kripke 1972.]

Lambert, Johann Heinrich. [1764] 1990. *Neues Organon oder Gedanken über die Erforschung und Bezeichnung des Wahren und dessen Unterscheidung vom Irrtum und Schein*, 3 vols. Berlin: Akademie-Verlag.

Lange, Friedrich Albert. 1877. *Logische Studien: Ein Beitrag zur Neubegründung der formalen Logik und der Erkenntnisstheorie*. Iserlohn: Verlag von J. Baedeker.

Leibniz, Wilhelm (G). 1875–1890. *Die philosophischen Schriften von Gottfried Wilhelm Leibniz*, 7 vols., ed. C. I. Gerhardt. Berlin.

Lewis, Clarence I. [1912] 1970. Implication and the algebra of logic. [*Mind* 21(84): 522–531.] Reprint, in *Collected Papers of Clarence Irving Lewis*, eds. John D. Goheen and John L. Mothershead, 351–359. Stanford: Stanford University Press.

Lewis, Clarence I. 1918. *A Survey of Symbolic Logic*. Berkeley: University of California Press.

Lewis, Clarence I., and Cooper H. Langford. [1932] 1959. *Symbolic Logic*. New York: Dover.

Lewis, David. 1973. *Counterfactuals*. Cambridge: Harvard University Press.

Lewis, David. 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

Linsky, Bernard, and Edward N. Zalta. 1994. In defense of the simplest quantified modal logic. *Philosophical Perspectives 8, Logic and Language*: 431–458.

Lipps, Theodor. 1893. *Grundzüge der Logik*. Hamburg & Leipzig.

MacColl, Hugh. 1906. *Symbolic Logic and Its Applications*. London: Longmans, Green.

Makinson, David. 1969. A normal modal calculus between T and S4 without the finite model property. *Journal of Symbolic Logic* 34: 38.

Marcus, Ruth Barcan. [1961] 1993. Modalities and intensional languages. [*Synthese* 13(4): 303–322.] Reprint, with some changes, in Ruth Barcan Marcus, *Modalities: Philosophical Essays*, 5–23. Oxford: Oxford University Press.

McKinsey, J. C. C. 1941. A solution of the decision problem for the Lewis systems S2 and S4, with an application to topology. *Journal of Symbolic Logic* 6(4): 117–134.

McKinsey, J. C. C., and Alfred Tarski. 1948. Some theorems about the sentential calculi of Lewis and Heyting. *Journal of Symbolic Logic* 13(1): 1–15.

Parry, William T. 1939. Modalities in the *Survey* system of strict implication. *Journal of Symbolic Logic* 4(4): 137–154.

Plantinga, Alvin. 1974. *The Nature of Necessity*. Oxford: Oxford University Press.
Plantinga, Alvin. [1976] 1979. Actualism and possible worlds. [*Theoria* 42: 139–160.] Reprint, in *The Possible and the Actual: Readings in the Metaphysics of Modality*, ed. Michael J. Loux, 253–273. Ithaca, N.Y.: Cornell University Press.
Prior, Arthur N. 1957. *Time and Modality*. Oxford: Oxford University Press.
Quine, Willard V. 1943. Notes on existence and necessity. *Journal of Philosophy* 40(5): 113–127.
Quine, Willard V. [1953a] 1976. Three grades of modal involvement. [*Proceedings of the XI International Congress of Philosophy* 14: 65–81.] Reprint, in Willard V. Quine, *The Ways of Paradox and Other Essays* (revised and enlarged edition), 158–176. Cambridge, Mass.: Harvard University Press.
Quine, Willard V. [1953b] 1963. Reference and modality. Reprint, in Willard V. Quine, *From a Logical Point of View* (second edition), 139–159. New York: Harper.
Russell, Bertrand. 1905. On denoting. *Mind* 14(56): 479–493.
Sahlqvist, Henrik. 1975. Completeness and correspondence in the first- and second-order semantics for modal logic. In *Proceedings of the third Scandinavian logic symposium*, ed. Stig Kanger. Amsterdam: North-Holland.
Segerberg, Krister. 1971. An essay in classical modal logic. *Filosofiska studier* 13, Department of Philosophy, University of Uppsala.
Sigwart, Christoph. 1871. *Beiträge zur Lehre vom hypothetischen Urtheile*. Tübingen.
Sigwart, Christoph. 1873. *Logik*. Erster Band. Tübingen: Verlag der H. Laupp'schen Buchhandlung.
Stalnaker, Robert. 1976. Possible worlds. *Noûs* 10(1): 65–75.
Trendelenburg, Adolf. 1862. *Logische Untersuchungen I–II*. Leipzig.
Ueberweg, Friedrich. [1857] 1882. *System der Logik und Geschichte der Logischen Lehren*, 5th ed. Bonn: Adolph Marcus.
Veraart, Albert. 1976. Geschichte des wissenschaftlichen Nachlasses Gottlob Freges und seiner Edition. Mit einem Katalog des ursprünglichen Bestands der nachgelassenen Schriften Freges. In *Studien zu Frege I: Logik und Philosophie der Mathematik*, ed. Matthias Schirn, 49–108. Stuttgart and Bad Cannstatt: Frommann-Holzboog.
Whitehead, Alfred N., and Bertrand Russell. 1910. *Principia Mathematica, Vol. 1*. Cambridge: Cambridge University Press.
Windelband, Wilhelm. [1884] 1921. Beiträge zur Lehre vom negativen Urtheil. In *Sonder-Abdruck aus Strassburger Abhandlungen zur Philosophie Eduard Zeller zu seinem siebzigsten Geburtstage 1884*. Tübingen: J. C. B. Mohr.
Wolff, Christian [1754] 1965. *Vernünftige Gedanken von den Kräften des menschlichen Verstandes und ihrem Richtigen Gebrauche in Erkenntnis der Wahrheit*. 14th ed. Hildesheim: Georg Olms.
von Wright, Georg H. 1951a. *An Essay in Modal Logic*. Amsterdam: North-Holland.
von Wright, Georg H. 1951b. Deontic logic. *Mind* 60(237): 1–15.
Wundt, Wilhelm. 1880. *Logik: Eine Untersuchung der Principien der Erkenntnis und der Methoden Wissenschaftlicher Forschung, Erster Band: Erkenntnislehre*. Stuttgart.

Appendix to Chapter 12

Conditionals and Possible Worlds: On C. S. Peirce's Conception of Conditionals and Modalities

Risto Hilpinen

1.

C. S. Peirce took conditionals or *hypothetical propositions* (as he usually called conditionals) to be implicitly modal propositions. For example, in a manuscript written in 1895 (MS 787) he observes:

(H1) The quantified subject of a hypothetical proposition is a *possibility*, or *possible case*, or *possible state of things.* (CP 2.347)

Peirce accepted the view that any proposition can be regarded as having a subject-predicate form, but he interpreted the subject-predicate analysis in a new way and applied it to complex as well as simple sentences. For example, according to Peirce, the subject of a quantified sentence consists of the quantifier (or the quantifiers, in the case of multiply quantified sentences) (Peirce 1903 [1997], 180–181). In general, the function of the subject or subjects of a proposition is to indicate its object or objects (what the proposition is about), and the predicate (when attached to the subjects) states something about the objects (see CP 2.357–358; Hilpinen 1992, 472–478; 1998, 155–159). According to the observation just quoted, the objects of a hypothetical

I follow here the usual practice of citing from the *Collected Papers of Charles Sanders Peirce* (Peirce 1931–1935, 1958) by volume number and paragraph number, preceded by CP. The chronological edition of Peirce's writings, Peirce (1982–2000), will be abbreviated WCSP, followed by volume number and page numbers. References to the microfilm edition of Peirce's manuscripts (Widener Library, Harvard University), cataloged in Robin (1967), will be indicated by MS, followed by the manuscript number. The page numbers used in the references to the manuscripts are those used by Peirce (and contained in the microfilm edition of the manuscripts).

proposition include possibilities or possible states of things: Hypothetical propositions are statements about possibilities.

In his paper "On the Algebra of Logic: A Contribution to the Philosophy of Notation" (1885), Peirce observes:

(H2) A hypothetical proposition, generally, is not confined to stating what actually happens, but states what is invariably true throughout a universe of possibility. (CP 3.366)

The peculiarity of the hypothetical proposition is that it goes beyond the actual state of things and declares what *would* happen were things other than they are or may be. (CP 3.374)

Peirce accepted the Aristotelian-scholastic view that modal propositions are *quantified* propositions of a special kind. According to this view,

(M1) the necessary (or impossible) proposition is a sort of universal proposition; the possible (or contingent, in the sense of not necessary) proposition, a sort of particular proposition. (CP 2.382)

At various stages of his philosophical development Peirce presented several different variants of this conception. One of the earliest formulations can be found in his 1865 Harvard Lectures "On the Logic of Science":

(M2) When we say that the straight line is necessarily the shortest distance between two points we mean that it is so not merely in this or that state of things but in every state of things. It has always been so, it is so; it will always be so. (WCSP 1, 200–201)

The second statement in (M2) suggests a temporal or "statistical" conception of modalities, according to which modality involves quantification over time: A necessary proposition (or statement) is always true, a possible proposition holds sometimes, and an impossible one, never. This account of modalities goes back to Aristotle, and it was accepted by many scholastic philosophers (Hintikka 1973, 102–103; Knuuttila 1982, 345–346). It is clear that this view can be applied in a significant way only to temporally indeterminate statements. The first sentence of (M2) looks more interesting: It expresses the meaning of a modal proposition (statement) by means of the concept of *state of things* also used in (H1) and (H2). In (M2), the expression "state of things" appears to refer to something which makes a proposition or a statement true or false, or something *relative to which* a proposition may be true or false, for example, a possible (state of the) world. In CP 5.549 Peirce explains the concept of a state of things as follows: "A *state of things* is an abstract constituent part of reality, of such a nature that a proposition is needed to represent it." In contemporary philosophy, the expression "state of affairs" is sometimes used in this sense. States of things or states of affairs are thought to be the objective correlates of propositions or statements. If states of things are understood in this way, the first sentence of (M2) is not an acceptable characterization of the concept of necessity. On the other hand, (M2) is a plausible characterization of necessity if the states of things referred to are taken to be complete (or maximal) states, that is, "hypothetical states of the universe, each absolutely determinate in every respect" (CP 2.382). Complete states are often called *possible worlds*. The expression "universe of possibility" in (H2) should be taken to mean a collection of complete states or possible worlds. This interpretation of (M2) makes it equivalent to the possible worlds analysis of modal concepts, according to which necessity means truth in all possible worlds, and possibility means truth in some possible world. One of the first proponents of this account of modalities was John Duns Scotus (see Knuuttila 1982, 353–355), who influenced the basic tenets of Peirce's pragmatism, especially his scholastic realism, that is, the view that there are "real generals" (CP 5.453; see Moore 1964, 401).

2.

In his later writings, Peirce often analyzes modal statements by means of the concept of *state of information, state of knowledge*, or *state of ignorance*. In "The Essence of Reasoning" (1893, MS 409), Peirce defines the concept of informational possibility as follows:

> By "informationally possible," I mean possible so far as we, or the persons considered, know. Then, the *informationally possible* is that which in a given [state of] information is not perfectly known not to be true. The *informationally necessary* is that which is perfectly known to be true. The *informationally contingent*, which in the given [state of] information remains uncertain, that is, at once possible and unnecessary.
>
> The information considered may be our actual information. In that case, we may speak of what is possible, necessary, or contingent, *for the present*. Or it may be some hypothetical state of knowledge. (CP 4.65–66)

By "perfect knowledge," Peirce means here an opinion that is completely settled so that further inquiry, no matter how far pushed, would not change it (CP 4.62).

According to Peirce, different senses of necessity and possibility can be defined by varying the hypothetical state of knowledge on which modal assertions are based:

> Imagining ourselves to be thoroughly acquainted with all the laws of nature and their consequences, but to be ignorant of all particular facts, what we should then not know to be true is said to be *physically possible*; and the phrase *physically necessary* has an analogous meaning. If we imagine ourselves to know what the resources of men are, but not what their dispositions and desires are, what we do not know will not be done is said to be *practically*

> *possible*, and the phrase *practically necessary* bears an analogous signification. (CP, 4.66)

Thus "physical possibility" means compatibility with the laws of nature, and practical possibility means compatibility with the resources, dispositions, and desires of people. Logical or essential possibility can be defined in an analogous way:

> That is *essentially* or *logically possible* which a person who knows no facts, though perfectly *au fait* at reasoning and well-acquainted with the words involved, is unable to pronounce untrue. The *essentially* or *logically necessary* is that which such a person knows is true. . . .
>
> On the other hand, the substantially possible refers to the information of a person who knows everything now existing, whether particular law or fact, together with all their consequences. . . . In this sense, everything in the present which is possible is also necessary, and there is no present contingent. But we may suppose there are "future contingents." (CP 4.67)

It is clear that Peirce is using the word "know" in an idealized sense of knowability (or virtual knowledge), for example, he assumes that the epistemic subject in question is logically omniscient. He uses the concept of state of information (or state of ignorance) as convenient theoretical fiction for distinguishing different senses of possibility or necessity. The connection between this analysis and the characterization of modalities in terms of possible worlds is obvious. In his article "Modality" in J. M. Baldwin's *Dictionary of Philosophy and Psychology* (1901), Peirce observes that the state of ignorance of a subject

> will consist in its subject being unable to reject certain hypothetical states of the universe, each absolutely determinate in every respect, but all of which are, in fact, false. The aggregate of these unrejected falsities constitute the "range of possibility," or better, "of ignorance." Where there is no ignorance, this aggregate would be reduced to zero. (CP 2.382)

These "absolutely determinate" (hypothetical) states of the universe are also called possible worlds.

Peirce emphasizes that this way of characterizing modalities does not commit him to any ontological doctrine about the nature of possibility and necessity, for example, to the view that possibility and necessity are essentially epistemic concepts:

> The conclude from the above definitions that there is nothing analogous to possibility and necessity in the real world, but that these modes appertain only to the particular limited information which we possess, would be even less defensible than to draw

> precisely the opposite conclusion from the same premisses. It is a style of reasoning most absurd. Unfortunately, it is so common, that the moment a writer sets down these definitions nine out of ten critics will set him down as a nominalist. (MS 409; CP 4.68)

By "nominalism" Peirce means here the view that there are no "real possibilities" (or "real generals"), and that modalities are essentially epistemic concepts. The doctrine that there are real (unactualized) possibilities was an important element of Peirce's mature version of pragmatism (Peirce 1905a, 1905b), which he called "pragmaticism" to distinguish it from the "nominalistic" views of William James, F. C. S. Schiller, and others (CP 5.414).

3.

In his paper "Issues of Pragmaticism" (1905b), Peirce makes a distinction between *subjective* and *objective* modality and characterizes the former as follows:

> In the simplest case, the most subjective meaning, if a person does not know that a proposition is false, he calls it *possible*. . . . In this most subjective kind of Modality, that which is known by direct recollection is in the mode of *Actuality*, the determinate mode. But when knowledge is indeterminate among alternatives, either there is one state of things which alone accords with them all, when this is the mode of *Necessity*, or there is more than one state of things that no knowledge excludes, when each of these is in the mode of *Possibility*. (CP 5.454)

Here the expression "state of things" is used in the sense of a state of affairs (an objective correlate of a proposition), and the "alternatives" that a given state of things can accord or fail to accord are possible worlds (in this case, doxastic or epistemic possibilities). Concerning objective modality, Peirce notes:

> There are other cases, however, in which, justifiably or not, we certainly think of Modality as objective. A man says, "I can go to the seashore if I like." Here is implied, to be sure, his ignorance of how he will decide to act. But this is not the point of the assertion. It is that the complete determination of conduct in the *act* not yet having taken place, the further determination of it belongs to the subject of the action regardless of external circumstances. If he had said, "I must go where my employers may send me," it would imply that the function of such further determination lay elsewhere. (CP 5.455)

According to Peirce, objective modal propositions represent reality as analogous to the indecision of a person:

> for example, a superstitious cashier, impressed by a bad dream, may say to himself on a Monday morning, "*May be*, the bank has been robbed." No doubt, he recognizes his total ignorance in the matter. But besides that, he has in mind the absence of any particular cause which should protect his bank more than others that are robbed from time to time. He thinks of the variety in the universe as vaguely analogous to the indecision of a person, and borrows from that analogy the garb of his thought. (CP 5.455)

4.

Peirce was one of the principal architects of modern propositional logic and quantification theory. He and Gottlob Frege were the first philosophers who construed quantifiers as variable binding operators (Frege 1879; Peirce 1883, 1885). Peirce began to work on the algebra of logic in the 1860s, and his research in this field culminated in his paper "On the Algebra of Logic: A Contribution to the Philosophy of Notation" (1885), where he presented a truth-value analysis of propositional connectives, an axiom system and decision procedure for propositional logic (called "non-relative logic," CP 3.365–390), and a system of first-order quantification theory ("first-intentional logic of relatives," CP 3.392–397) and second-order logic ("second-intentional logic of relatives," CP 3.398–403). However, he did not regard the new logic as a universal instrument of reasoning and formalization, and (unlike many of his followers) he was aware of the limitations of first-order (extensional) logic. For example, he did not think that the truth-conditions of conditional propositions can be expressed in extensional first-order logic. As was observed (see H2), he regarded hypothetical propositions as statements about possibilities: A hypothetical proposition "if P, then Q" states that the "range of possibility" does not contain a state in which P is true and Q is false. The range of possibility under consideration "is in one case wider, in another narrower" (CP 3.374). The truth-functional ("Philonian") conditionals of Peirce's nonrelative logic (1885) constitute a special case in which the range of possibility is limited to "one individual state of things, the Actual" (CP 3.366, 3.375).

In the late 1890s Peirce developed a system of logical diagrams called "Existential Graphs," which he regarded as his most significant contribution to logic (Roberts 1973, 11–12). The system is divided into three parts, Alpha, Beta, and Gamma. Alpha is a systematization of propositional logic (what Peirce called "non-relative logic"), and the Beta graphs constitute a complete system of first-order logic with identity. However, Peirce was dissatisfied with the Beta graphs because they were restricted to extensional logic and could not represent reasoning involving modal and hypothetical propositions in a satisfactory way. One purpose of the Gamma graphs was to represent modal reasoning, and in his work on Gamma graphs Peirce anticipated some later developments in the semantics of modal logic, for example, he observed that modal graphs do not represent a single universe of "existent individuals," but several interconnected universes of discourse (that is, possible worlds) (see CP 4.512; Zeman 1997, 410–411; for the Gamma graphs, see CP 4.513–529, 4.573–584; Zeman 1964, ch. III; Roberts 1973, 64–71). Peirce's Gamma system remained a sketch, but it has recently been used as a basis of interesting systems and applications of modal logic (van den Berg 1993; Øhrstrøm and Hasle 1995, 320–343; Øhrstrøm 1996).

One of Peirce's most interesting arguments for the non–truth-functional and nonextensional character of conditional propositions can be found in his article "Prolegomena to an Apology for Pragmaticism" (Peirce 1906; CP 4.530–572), where he discusses, among other things, the logic of existential graphs, the truth-conditions of conditionals, and the objectivity and reality of possibilities. Peirce wants to show that actuality (what exists) does not exhaust reality:

> Let us . . . try whether we may not assume that there is but one kind of subjects which are either existing things or else quite fictitious. Let it be asserted that there is some married woman who will commit suicide in case her husband fails in business. Surely this is a very different proposition from the assertion that some married woman will commit suicide if all married men fail in business. (CP 4.546)

It is clear that the two propositions are not logically equivalent. The former proposition entails the latter, but not conversely: A woman who commits suicide if all married men go bankrupt need not commit suicide if only her husband goes bankrupt. However, if the truth-conditions of the statements in question, that is,

(1) Some married woman commits suicide if her husband fails in business.

and

(2) Some married woman commits suicide if all married men fail in business.

are expressed in the language of extensional logic (quantification theory), they turn out to be logically equivalent. Peirce shows this by pointing out that if (1) and (2) are understood in this way, they are false under exactly the same circumstances. According to the extensional reading, (1) can be expressed in the form

(3) Some married woman commits suicide or her husband does not fail business,

and (3) is false if and only if (i) there is no married couple, or (ii) no married woman commits suicide while every husband goes bankrupt. (Every married woman who commits suicide and every woman whose husband does not fail in business satisfies the disjunctive predicate in (3).) Such circumstances are also the circumstances under which (2) is false. According to Peirce, "the equivalence of these two propositions is the absurd result of admitting no reality but existence" (CP 5.546). This absurdity is avoided if (1) and (2) are regarded as modal conditionals, in other words, as propositions that refer to unactualized possibilities and not only to the actual course of events (the actual world). According to Peirce (CP 4.546), (1) should be taken to mean the same as

(4) There is some *one* married woman who under all possible conditions would commit suicide or else her husband would not have failed,

whereas the meaning of (2) can be expressed by

(5) There is some married woman who under all possible conditions would commit suicide or else not all married businessmen would have failed.

Peirce emphasizes that (1) and (2) must be understood modal propositions *de re*; he notes that "there is a great difference" between (4) and the *de dicto* proposition

(6) Under all possible circumstances there is some married woman *or other* who would commit suicide, or else her husband would not have failed.

It is clear that the *de dicto* formulation of (1) and (2) would not solve the problem: If the extensional readings of (1) and (2) are logically equivalent, their modalized *de dicto* counterparts are logically equivalent as well. On the other hand, (4) and (5) are not logically equivalent: The former entails the latter but not conversely. For example, (5) is true but (4) is false if there are two married women, Alicia and Delia, and two possible circumstances or situations, s and u, such that:

(7) In s, Alicia's husband fails in business, Delia's husband does not fail in business, and neither Alicia nor Delia commits suicide,

and

(8) In u, both husbands go bankrupt and Alicia commits suicide, but Delia does not commit suicide.

According to (7) and (8), there is a married woman who under all possible conditions would commit suicide or not all married businessmen would have failed (i.e., if all married businessmen had failed), viz., Alicia, but there is no married woman who under all possible conditions would commit suicide in case her husband were to fail. Both Alicia and Delia can (in suitable circumstances) continue to live despite the husband's failure. In other words, (5) does not entail (4).

As Peirce observed, the conditionals in question must be construed as *de re* conditionals. The concept of possibility in (4) and (5) need not be epistemic possibility: The modal difference between Alicia and Delia can be a real (psychological) difference between the two women. Thus Peirce's example illustrates simultaneously the non–truth-functional character of conditional propositions and the reality of unactualized possibilities. (For a discussion of Peirce's example from the standpoint of his pragmatism, see Wennerberg 1962, 143–144.)

Peirce regards hypothetical propositions as strict rather than material conditionals. In the example discussed, the conditionals under consideration are formulated as indicative conditionals. In the recent work on the logic of conditionals, it has been argued that many conditional propositions (for example, counterfactual and subjunctive conditionals) are *variably* strict rather than strict conditionals (Lewis 1973, 13). This means that the "strictness" of the conditional depends on its antecedent and the context of evaluation: "If (it were the case that) P, (it would be the case that) Q" is true in a given situation or point of evaluation w if and only if Q holds in all *selected* worlds where P is true. The selection depends on the point of evaluation; for example, the selected worlds may be those that resemble w as much as possible (that is, as much as the truth of the antecedent permits; see Lewis 1973, 8–21), or they may be worlds where certain boundary conditions or general laws hold. Peirce's concept of range of possibility can easily be adjusted to fit the truth-conditions of variably strict conditionals. As was observed, Peirce suggests that the range of possibility can vary from case to case, and "is in one case taken wider, in another narrower." In the case of extensional ("first-intentional") logic, it is limited to the actual state of things (CP 3.375). The relevant range of possibility may also depend on what is actually taken to be the case, as in the following example of a counterfactual conditional:

> To say that if Napoleon had been in his best trim he would have won the battle of Waterloo, so far as it means anything, means that taking all the different possible courses of events that might reasonably be admitted as such by taking into consideration the variations of power shown by Napoleon during his life, while external circumstances remain substantially as they were, every such possible course of events would either be one in which Napoleon was not in his best trim or would be one in which he would have won the battle of Waterloo. (MS 284, 29)

Peirce makes here a distinction between the antecedent of the conditional and the circumstances which are "external" to it, that is, the circumstances which are taken to be the same throughout the range of possibility. What is regarded as "external" depends of course on the antecedent; therefore different conditionals are evaluated in terms different (sets of) courses of events, that is, different ranges of possibility. Here Peirce's semantic analysis comes close to the view that counterfactual conditionals are variably strict rather than strict (or necessary) conditionals.

References

van den Berg, Harmen. 1993. Modal logics for conceptual graphs. In *Conceptual Graphs for Knowledge Representation: First International Conference on Conceptual Structures, ICCS'93*, eds. G. W. Mineau, B. Moulin, and J. F. Sowa, 411–429. Berlin and New York: Springer Verlag.

Frege, Gottlob. 1879. *Begriffschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Verlag von Louis Nebert. Reprinted in Gottlob Frege, *Begriffsschrift und andere Aufsätze*, ed. I. Angelelli. Darmstadt: Wissenschaftliche Buchgesellschaft, 1977.

Hilpinen, Risto. 1992. On Peirce's philosophical logic: Propositions and their objects. *Transactions of the Charles S. Peirce Society* 28: 467–488.

Hilpinen, Risto. 1998. On C. S. Peirce's pragmatic theory of meaning. In *The Role of Pragmatics in Contemporary Philosophy: Proceedings of the 20th International Wittgenstein-Symposium*, eds. P. Weingartner, G. Schurz, and G. Dorn, 151–164. Vienna: Hölder-Pichler-Tempsky.

Hintikka, Jaakko. 1973. *Time and Necessity: Studies in Aristotle's Theory of Modality*. Oxford: Clarendon Press.

Knuuttila, Simo. 1982. Modal logic. In *The Cambridge History of Later Medieval Philosophy*, eds. N. Kretzmann, Anthony Kenny, and Jan Pinborg, 342–357. Cambridge: Cambridge University Press.

Lewis, David K. 1973. *Counterfactuals*. Oxford: Blackwell.

Moore, Edward C. 1964. The influence of Duns Scotus on Peirce. In *Studies in the Philosophy of Charles Sanders Peirce: Second Series*, eds. E. C. Moore and R. Robin, 401–413. Amherst: University of Massachusetts Press.

Øhrstrøm, Peter. 1996. Existential graphs and tense logic. In *Conceptual Structures: Knowledge Representation as Interlingua: 4th International Conference on Conceptual Structures, ICCS'96*, eds. Peter W. Eklund and Graham Mann, 213–217. Berlin: Springer Verlag.

Øhrstrøm, Peter, and Per F. V. Hasle. 1995. *Temporal Logic: From Ancient Ideas to Artificial Intelligence*. Dordrecht: Kluwer.

Peirce, Charles Sanders. 1883. Note B: The logic of relatives. *Studies in Logic by Members of the Johns Hopkins University*, ed. C. S. Peirce, 187–203. Boston: Little, Brown. Reprint edition published in the series *Foundations of Semiotics*, vol. 1. Amsterdam/Philadelphia: John Benjamins, 1983. Reprinted in CP 3.328–358 and in WCSP 4, 453–466.

Peirce, Charles Sanders. 1885. On the algebra of logic: A contribution to the philosophy of notation. *American Journal of Mathematics* 7: 180–202. Reprinted in CP 3.359–403 and in WCSP 5, 162–190.

Peirce, Charles Sanders. [1903] 1997. *Pragmatism as a Principle and Method of Right Thinking: The 1903 Harvard Lectures on Pragmatism*, ed. Patricia Ann Turrisi. Albany: State University of New York Press.

Peirce, Charles Sanders. 1905a. What pragmatism is. *Monist* 15: 161–181. Reprinted in CP 5.411–437.

Peirce, Charles Sanders. 1905b. Issues of pragmaticism. *Monist* 15: 481–499. Reprinted in CP 5.438–463.

Peirce, Charles Sanders. 1906. Prolegomena to an apology for pragmaticism. *Monist* 16: 492–546. Reprinted in CP 4.530–572.

Peirce, Charles Sanders. 1931–1935. *Collected Papers of Charles Sanders Peirce*, vols. 1–6, eds. C. Hartshorne and P. Weiss. Cambridge, Mass.: Harvard University Press.

Peirce, Charles Sanders. 1958. *Collected Papers of Charles Sanders Peirce*, Vols. 7–8, ed. Arthur Burks. Cambridge, Mass.: Harvard University Press.

Peirce, Charles Sanders. 1982–2000. *Writings of Charles S. Peirce: A Chronological Edition*, ed. M. H. Fisch, N. Houser, et al., vols. 1–6, Bloomington: Indiana University Press.

Roberts, Don D. 1973. *The Existential Graphs of Charles S. Peirce*. The Hague: Mouton.

Robin, Richard S. 1967. *Annotated Catalogue of the Papers of Charles S. Peirce*. Amherst: University of Massachusetts Press.

Wennerberg, Hjalmar. 1962. *The Pragmatism of C. S. Peirce. An Analytical Study*. Library of Theoria no. 9. Lund: CWK Cleerup and Ejnar Munksgaard.

Zeman, J. Jay. 1964. *The Graphical Logic of C. S. Peirce*. Ph.D. diss., University of Chicago.

Zeman, J. Jay. 1997. Peirce and Philo. In *Studies in the Logic of Charles Sanders Peirce*, eds. N. Houser, D. D. Roberts, and J. Van Evra, 402–417. Bloomington: Indiana University Press.

13

Logic and Semantics in the Twentieth Century

Gabriel Sandu and Tuomo Aho

1. Semantics in the Twentieth Century

Semantics is broadly understood as the study of the meanings of the expressions of a given language. It can be approached from a theoretical as well as an empirical perspective. As a subfield of linguistics, it studies the meaning relations between different expressions (synonymy, homonymy, etc.), the argument structure of compound expressions in a given language, in contrast to syntax. Several examples of such semantic theories will be analyzed in chapter 16. Here, however, we shall largely be concerned with logical semantics, that is, the structural meaning of logical expressions like connectives (*it is not the case that*, *either or*, *if ... then*), quantifiers (*there is*, *for all*, etc.), and modalities (*it is necessary that*, *it is possible that*). The main focus will be on truth-theoretical semantics for formalized languages, a tradition emerging from Carnap's and Tarski's work in the first half of the last century that specifies the meaning of these expressions in terms of the truth-conditions of the sentences in which they occur. For a more detailed sketch of the developments which led to Carnap's and Tarski's theories, the reader is referred to chapter 16. The focus of the present chapter will be on Tarski-style definitions of the semantics of a given language in a stronger metalanguage, Tarski's impossibility results, and attempts to overcome them in the post-Tarskian tradition.

The birth of the "semantic tradition" was a slow and complicated process. It arose in the context of the marriage between mathematics and philosophy at the turn of the century that strove to put the former on secure foundations purified of Kantian intuitions. Language was experienced to have an objectivity of its own, every sentence bearing logical relations to other sentences in virtue of the meanings of its expressions encoded in its logical form. But not any language would do, only that which is regimented after a process of logical analysis which renders it transparent enough to be the medium of the pure laws of rational discourse. Semantics has its source in this process of logical analysis undertaken by Frege and Russell and then continued through the work of Wittgenstein by Carnap and the logical positivists. The start was shaky, clouded by paradoxes and by what has been called the universalist conception of logic and language (see the following): the impossibility of adopting a perspective which would allow one to study systematically the relations between language and reality. Although many interesting things were said about the meaning of proper names, concept words, quantifiers, and above all, the truth of statements, the reflections of Frege, Russell, and Wittgenstein never took the form of a general and systematic investigation. It is no wonder that in the early beginnings of his work on truth Tarski could write that mathematicians regard semantic notions with distrust. Little by little, the results emerging in metamathematics from Gödel's, Hilbert's and the Polish logicians' work imposed a new conception of studying formal languages based on compositional methods. The development of the model-theoretic point of view was for a long time, by the influence of Tarski and Carnap, restricted to the study of formal languages. Tarski's student, Montague, broke with this tradition, establishing English as a formal language, and around the same time, Davidson transformed Tarski's truth-definition into a theory of meaning for natural languages.

It is perhaps useful to offer here, as a digression, some remarks about the philosophical problems of logical semantics in general. It must be remembered that there are plenty of philosophically relevant topics even in the technical results of the semantic theory. For instance, what is the exact philosophical import of the incompleteness theorems, how strong conclusions can be derived from the dichotomy of object language and metalanguage, and what can be learned from the variety in methods of proving central theorems? Nevertheless, when we shall consider some philosophical issues we are going to restrict ourselves to a few very general and elementary issues without going to details at all. (See Haack 1978.) Obviously, one such basic question is about what the entities are for which semantics is ultimately built.

The linguistic units that logic has been concerned with have, with few exceptions, been "assertive" or "declarative." But the entities in this class, those which logic operates with, have traditionally been explained in several ways, and twentieth-century logicians have felt the need for clarification of the different senses. Nowadays it has become standard to speak about statements, sentences, and propositions. The explication of these concepts is important because there is disagreement about which of them is the true field of logic. Can some of them be eliminated or derived from others? In other words, is there any reason to suppose that the logical relations basically concern one of these fields?

Sentence is perhaps the most definite of these three. The grammar of a given language defines, on formal grounds, some expressions as well-formed sentences. (Natural languages allow borderline cases, whereas formal languages are fully definite in this respect.) The particular occurrences of sentences, or alternatively the contents of such occurrences, are *statements*. This notion is tricky because of indexical or deictic elements: Grammatically different sentence-forms often conduct the same statement if the indexicals are suitably changed. Thus, sentences differing with respect to “I”, “you”, “this”, “that”, “here”, and so on can make the same statement, if the speaker, place, and so on also vary accordingly. In logical theory, Russell already was interested in indexicals, their systematic study started in the 1950s (Bar-Hillel 1954), and Montague brought the matter into the scope of formal semantics. An additional complication comes from the assumption that sentences in different languages can serve the same statement. (Note, moreover, that the linguistic acts or “utterances” must not be confounded with their products or sentence-tokens.)

Propositions are the most difficult of the suggested items; that is, the old term is used in especially many and philosophically loaded senses. Some authors have wanted to eliminate propositions from ontology altogether as hopelessly obscure things. Quine is the most important example here. According to him, logic can only be concerned with sentences that are fully definable. (See Quine 1960, §40.) Though most others have not been willing to accept his argument, the notion of proposition has proved to be highly difficult. It has apparently been interpreted mainly according to three aspects. In one sense, a proposition can be the meaning of a sentence; second, it is said to be the entity that basically is true or false, the so-called truth-bearer; third, it can be the content of a propositional attitude. These three senses seek to explain propositions either by means of sentences, truth, or communicable attitudes. (Hence they have an affinity to syntax, semantics, and pragmatics, but the distinctions do not coincide.)

Let us try to say something about each meaning or aspect of the notion of a proposition. All the interpretations are classical and involved in the whole logical tradition. (For a learned historical summary with bibliographical references see Nuchelmans 1989.) It was customarily taken to be self-evident and unproblematic that sentences—unless they are “senseless”—have a meaning and that therefore there is something that is “the meaning” of the sentence. Frege wanted to give a clear and conscious expression of this idea with his general theory of *Sinne*, postulating that sentences have their senses in exactly the same way as nominal and predicative expressions. The senses of sentences were abstract and objectively existent entities, nonmental and language-independent “thoughts,” *Gedanke*. Obviously this guarantees the existence and identity of propositions, but also brings well-known difficulties, both ontological and logical. Frege’s approach was a bold idea that was not at all understood for a long time. Moore (1899) made an attempt of a somewhat parallel turn to propositions as definite conceptual meanings of sentences, but for the most part, early twentieth-century philosophy still used sentence meanings fairly nonchalantly.

Among empiricists, a popular assumption was naturally that the meanings could be understood positivistically, connecting them to patterns of verification, although this was seldom done in any clear-cut way. The awareness of problems connected to an accurate formulation of meanings then grew in the 1940s, provoked among other things by the development of linguistic theory. Russell (1940) may have been the last ambitious sample of the old positivist conception of meanings. Carnap (1947) takes the matter much more seriously, wishing to explicate the notion of meaning in the language L by use of "semantical rules" of L. Carnap's theory was intended to utilize Fregean ideas, but to avoid the strong metaphysical assumptions about abstract and unanalyzable thought-entities. However, Carnap must then assume as given the absolute semantical rules that lead to the meaning for each sentence, and it was asked if any true progress had been achieved.

Carnap proceeded by calling the class of sentences that are equivalent because of the semantic rules a proposition. This was one formulation of the thought that a proposition is what synonymous sentences have in common; the thought is of course old and independent of any mode of presentation. In other words, the same proposition is expressed in all utterances that are synonymous or have the same meaning. It could also be said that a proposition is what remains intact in correct translations. This intuitively persuasive idea was long accepted—and it dominated the theory of linguists—but it was seriously questioned when Quine claimed that there is simply nothing constant in translations. That was a result from his famous thesis of the indeterminacy of translation (Quine 1960): There is never an objective reason to say that two sentences are translations of each other, even if they are sentences in the same language, since there could be many schemes or "manuals" of translation that are mutually incompatible but still match all physical evidence. The radical thesis caused a lot of debate that is still going on in considerably sophisticated forms. However, Quine's original straightforward thesis appears to have been pretty convincingly criticized. (See Kirk 1986.)

In one sense, propositions are said to be contents of propositional attitudes like belief, knowledge, will, among others—the class of these attitudes is somewhat vague. (It is best in accord with classical tradition to speak of a content, though in the twentieth century many authors have called it the object of attitude, thus implicitly requiring that the attitude is a relation between two things. Such an assumption is not necessary.) Supposing some kind of intentionality of mental activity, it seems unquestionable that a propositional component is necessary for characterizing those attitudes. Also, a statement is a way of informing the audience of one's attitude. And it has been argued that the notion of propositional meaning is dependent on the notions of judgment and understanding, which belong together with propositional attitudes. Such debates from the time of Frege, Brentano, and Husserl have been revived as fundamental questions after Grice and Dummett. But this sense leads us deeply to the philosophy of mind and even to metaphysics. Therefore it is no wonder that many philosophers have found these notions dubious and, abandoning all accounts based on them, have expelled intentions from semantics. Thus, for instance, Quine implied clearly that the logical incomprehensibility of the intentionality doctrine is a sufficient reason for rejecting it entirely.

Modern supporters of propositions often want to understand them in a modest way, without any supposition of a special metaphysical type of objects. In this spirit it is argued that, proceeding with due care, one can reasonably speak about the common meaning of several utterances, or the common content of the beliefs of several people, with rather weak ontological commitments.

In the third sense, propositions are *truth-bearers*. This thesis obviously depends on what the truth-bearers are, that is, what entities in the first place are true or false (or have some other truth value). At least sentences, propositions, statements, judgments, and beliefs are defended. However, the debate here appears to be partly inessential. When, say, some theorist asserts that truth is a property of sentence-tokens and another claims that it is a property of propositions, they probably disagree mainly about the correct criteria for regarding one use of the word as more basic than another. (See Kirkham 1992, ch. 2.) But of course, if propositions are defined as truth-bearers, this use is seen as fundamental. The most successful formal truth theory, Tarski's definition, was written for sentences, but it is questionable if something can be concluded from this fact—though it is true that an abundant prehistory of Tarski includes discussion just about truth-bearers (Rojszczak 2005). Anyway, the question about truth-bearers is seen to be strongly bound to the general issues of *truth-theory*.

2. Frege and Russell

2.1. Frege: Senses, Concepts, and Extensions

Frege had in the *Begriffsschrift* (1879) a function-argument distinction: In §9 he compares "the number 20" with "positive integer" and remarks that the former, unlike the latter, corresponds to an independent representation. The distinction was further extended in *Die Grundlagen der Arithmetik* (1884) (*The Foundations of Arithmetic*) to one between *object*, *concept*, and its *extension*. The explicit definition of concepts as functions which map objects into the set of truth values *True* and *False* as well as a rigorous distinction between sense (*Sinn*) and reference (*Bedeutung*) is introduced in his *Funktion und Begriff* (1891).

The distinction between concepts and the objects which are their extensions is for Frege a categorial one: The former are (unsaturated) functional entities while the latter are (saturated) objects. In his review of Husserl's *Philosophie der Arithmetik* (1894), Frege stated unambiguously the necessary and sufficient criterion for the identity of concepts: the identity of their extensions.

Frege needed the distinction between concepts and their extensions in the *Grundlagen* for the definition of "The number that belongs to the concept F" as the extension of the concept "concept equinumerous to the concept F", where a concept G is called equinumerous to a concept F if the possibility exists of a one–one correlation. He did not say much about what kinds of entities extensions are until the first volume of *Grundgesetze der Arithmetik* (1893), where he formulates an explicit criterion for their identity, his famous Axiom V: "Two concepts F and G have the same extension if and only if whatever falls under the concept F falls under the concept G, and *vice versa*." Frege construed natural numbers as extensions of certain concepts thinking that in this way he would reduce arithmetic to logic. It turned out, however, that Frege's system was contradictory. Russell came to learn indirectly of the work done by Frege from Peano with whom Frege had been corresponding from 1894 to 1896. In *My Philosophical Development* (1959, ch. VII), Russell remembers how in the spring of 1901 he discovered the paradox which bears his name, of the class of classes which are not members of themselves. In 1902 at the time when the second volume of Frege's *Grundgesetze* was in press, Russell communicated it in a letter to Frege. (The correspondence between Frege and Russell and Russell's solution to the paradoxes are detailed in chapter 9.)

From the point of view of the present chapter, it is important to notice that at this stage Frege took sentences (we shall use interchangeably "sentences" and "propositions") to be complex names which refer to truth values. He adopted the principle of compositionality, which states that the semantic value (reference, extension) of a complex expression is completely determined by the semantic values of its parts (as given by the syntactic analysis of the sentence). An immediate consequence of this principle is that substitutivity of two expressions with the same semantic value must leave the value of a complex expression unchanged. These two assumptions together led Frege to conclude that all true sentences refer to the same thing and so do all false ones:

> [the reference of a sentence] must remain unchanged when a part of a sentence is replaced by an expression with the same reference... What feature except the truth value can be found that belongs to ... sentences quite generally and remains unchanged by substitutions of the kind just mentioned?
>
> If now the truth value of a sentence is its reference, then on the one hand all true sentences have the same reference and so, on the other hand, do all false sentences. (Frege 1892a, 64–65 in 1980)

The idea that all true sentences refer to one entity is very important in the history of twentieth century semantics. We return to it after contrasting it with the idea that each sentence stands for a fact.

2.2. Russell

The logical system developed by Russell in *The Principles of Mathematics* (1903) and in the articles preceding "On Denoting" (1905) contains, as witnessed by Russell in appendix A of his *Principles*, many of the doctrines set forth by Frege. For Russell like for Frege, the meaning of a sentence is a proposition, a mind-independent entity that has constituents. A Russellian proposition is a combination of concrete things (Frege's *Gegenstand*) and concepts, which matches closely the grammatical structure of the sentence, and is to be found, from the vantage point of Frege, at the level of reference. Every word occurring in a sentence has a meaning that appears as a constituent in the corresponding proposition (*Principles*, §51). What this meaning is varies from an expression to another. For a singular like "Socrates", it is the thing which bears it; for verbs and adjectives, it is a concept (ibid., §46, §48). Thus the Russellian (singular) proposition that is the meaning of the sentence "Socrates is human" contains as constituents the concrete individual Socrates and the concept of being human, while its Fregean counterpart includes two abstract entities, the sense of the singular term "Socrates" and the sense of the concept word "human." Although Russell was well aware of Frege's distinction between sense and reference, he did not see any compelling reason to adopt it for singular terms. He adopted it, however, for descriptions, which have both a meaning and a denotation, a move that allowed him, among other things, to explain the puzzle about identity ("The Morning Star is the Evening Star") in the same way as Frege did. Later on, however, in "On Denoting" (1905), he ended up by rejecting the two-stage Fregean analysis (see chapter 9) and proposed the well-known method of contextual elimination of definite descriptions. According to it, the sentence "The author of Waverley is Scotch" is a conjunction of three sentences:

1. At least one individual has written Waverley;
2. At most one individual has written Waverley; and
3. Anybody who has written Waverley is Scotch.

Thus in none of the sentences (1), (2), and (3) is the expression "the author of Waverley" present any longer.

To resume, in "On Denoting" Russell indicated how denoting concepts can be replaced by expressions that do not denote, but in the same time he had to give up his earlier thesis in *The Principles* according to which the grammatical structure of the sentence reflects the structure of the proposition. Once the grammatical structure of the sentence is no more faithful to the structure of the proposition, the purpose of the philosophical analysis is to find the symbolic sentence that adequately represents it. Embracing such a view made it possible for Russell to go into a direction opposite to Frege to reach a conception according to which true sentences stand up for facts. In "The Philosophy of Logical Atomism" (1918) he writes, "the world contains facts, which are what they are whatever we may choose to think about them, and . . . there are also beliefs, which have reference to facts, and by reference to facts are either true or false" (182).

Thus, for Russell, facts, as complexes of particulars and universals, are part of the real world. They are the sort of things expressed by a whole sentence and not by a single name. Most important, facts are the kind of entities true propositions and beliefs correspond to.

3. Slingshot Arguments

The crucial difference between Frege and Russell is that the former treats definite descriptions referentially, whereas the latter treats them quantificationally. The difference turns out to be philosophically very significant, for on Frege's account all true sentences correspond to one single entity (the truth value *True*) and all false sentences correspond to the truth value *False*, while on Russell's account each true sentence corresponds to a fact. Perhaps nobody saw more clearly the distinction between the two accounts than Kurt Gödel in his review of Russell's work. We quote him at length:

> An interesting example of Russell's analysis of the fundamental logical concepts is his treatment of the definite article "the". The problem is: what do the so-called descriptive phrases ... denote or signify [note: I use the term "signify" in the sequel because it corresponds to the German word "bedeuten" which Frege, who first treated the question under consideration, first used in this connection.] and what is the meaning of sentences in which they occur? The apparently obvious answer that, e.g., "the author of Waverley" signifies Walter Scott, leads to unexpected difficulties. For, if we admit the further apparently obvious axiom, that the significance of a complex expression, containing constituents which have themselves a signification, depends only on the significance of these constituents (not on the manner in which this signification is expressed), then it follows that the sentence "Scott is the author of Waverley" signifies the same thing as "Scott is Scott"; and this again leads almost inevitably to the conclusion that all true sentences have the same signification (as well as all false ones). Frege actually drew this conclusion; and he meant it in an almost metaphysical sense. (Gödel 1944, 128–129)

In other words, if one sticks to the principle of compositionality and treats, in addition, definite descriptions referentially, then, under few obvious assumptions, one cannot avoid the conclusion that all true sentences refer to the same thing. Russell escaped this conclusion by not treating definite descriptions referentially, but as Gödel pointed out,

> As to the question in the logical sense, I cannot help feeling that the problem raised by Frege's puzzling conclusion has only been evaded by Russell's theory of descriptions and that there is something behind it which is not yet completely understood. (Gödel 1944, 130)

In a note to the first quotation, Gödel indicates the further assumptions needed for the proof of the statement that all true sentences refer to the same fact:

G1 "$\varphi(a)$" and the proposition "a is the object which has the property φ and is identical to a" mean (refer to) the same thing.

G2 every proposition speaks about something, that is, it can be brought to the form "$\varphi(a)$".

G3 for any two objects a, b there exists a true proposition of the form "$\varphi(a, b)$" as e.g., "$a \neq b$" or "$a = a \wedge b = b$". (Gödel 1944, 129)

With these assumptions at hand, it is straightforward to show that all true sentences refer to one and the same entity (stand for the same fact). Gödel himself did not explicitly give the proof, but many logicians after him did that. We follow here the proof given in Neale (1995).

Let us agree to use "$\iota x(x = a \wedge \varphi(x))$" as a symbolization of the definite description "the object which has the property φ and is identical to a."

We illustrate Gödel's proof with two true atomic sentences, "Fa" and "Gb." We suppose also that "$a \neq b$" is also true (the premiss "$a = b$" being true leads to the same conclusion). We show that all these three sentences correspond (refer to) to the same fact.

1. "Fa" corresponds to f_1, Premise.
2. "Gb" corresponds to f_2, Premise.
3. "$a \neq b$" corresponds to f_3, Premise.
4. "$a = \iota x(x = a \wedge Fx)$" corresponds to f_1, (1) and G1.
5. "$a = \iota x(x = a \wedge x \neq b)$" corresponds to f_3, (3) and G1.
6. "$a = \iota x(x = a \wedge Fx)$" and "$a = \iota x(x = a \wedge x \neq b)$" correspond to the same fact.

(Here we use the assumption that definite descriptions contribute with their reference, and the principle of compositionality: The former entails that the two descriptions contribute with the same individual, while the latter ensures that there are no further ingredients in the corresponding facts than those contributed by "a", "$=$", and the relevant descriptions.)

7. $f_1 = f_3$ (4), (5), (6).
8. "$b = \iota x(x = b \wedge Gx)$" corresponds to f_2, (2) and G1.
9. "$b = \iota x(x = b \wedge x \neq a)$" corresponds to f_3, (3) and G1.
10. "$b = \iota x(x = b \wedge Gx)$" and "$b = \iota x(x = b \wedge x \neq a)$" correspond to the same fact (analogous to (6)).
11. $f_2 = f_3$ (8), (9), (10).
12. $f_1 = f_2 = f_3$. (7), (11).

Arguments of this kind, called *slingshots*, have been used over and over again in twentieth-century semantics. The most notorious ones are due to Church (1943, 1956), Quine (1953, 1960) and Davidson (1967, 1990).

Church (1943) used a slingshot argument in his review of Carnap's book *Introduction to Semantics* (1942). In the book, Carnap departs from his teacher Frege by taking sentences to designate propositions or states of affairs that are complex entities and not Fregean truth values. In his review, Church uses a Gödelian argument to show that in Carnap's system all true sentences designate the same proposition. (In response to this argument, Carnap [1947] takes the references of sentences to be truth values.)

Church's slingshot argument departs in two significant ways from Gödel's. First, he replaces G1 by the stronger principle:

SLE Logically equivalent sentences designate the same proposition (state of affairs).

In addition, instead of using Gödel's definite description operator "$\iota x\varphi$", Church used "$\lambda x\varphi$" with the meaning "the class of all x such that φ." Here we shall use the set theoretical notation $\{ x : \varphi(x) \}$ as a term that denotes the class of entities which have the property φ, and describe a slingshot argument that is common to both Church and Davidson.

Davidson uses the slingshot against the correspondence theory of truth to show that any two sentences correspond to the same fact. Once again, the underlying assumptions are the principle of compositionality (PC), (SLE), and the principle that the contribution of a definite description to a fact is the individual that satisfies it (hence two coreferential definite descriptions are interchangeable; we denote this principle by SCT).

Two observations are essential in the proof:

O3 "φ" and "$\{ x : x = x \wedge \varphi \} = \{ x : x = x \}$" are logically equivalent.

(Indeed, observe that when φ is true, then the term "$\{ x : x = x \wedge \varphi \}$" denotes the class of individuals x such that x is identical to itself, and φ. Because φ is true, this class is the universal one, that is, the denotation of "$\{ x : x = x \}$". Thus "$\{ x : x = x \wedge \varphi \} = \{ x : x = x \}$" is true. For the converse, assume that "$\{ x : x = x \wedge \varphi \} = \{ x : x = x \}$" is true. Whence the true terms must denote the same entity and that is possible only if φ is true.)

O4 Whenever φ and ψ are true, the descriptions "$\{ x : x = x \wedge \varphi \}$" and "$\{ x : x = x \wedge \psi \}$" designate the same class.

The starting assumption is that φ and ψ are true.

13. "φ" corresponds to f_1, Premise.
14. "$\{ x : x = x \wedge \varphi \} = \{ x : x = x \}$" corresponds to f_1, (13), and (SLE).
15. "$\{ x : x = x \wedge \psi \} = \{ x : x = x \}$" corresponds to f_1, (14), (PC), and (SCT).
16. "ψ" corresponds to f_1, (15), and (SLE).

The philosophical signification of the slingshot is undeniable. It shows that nonextensional operators or connectives like "corresponds to the fact that", "signifies the proposition that", "it is necessary that" collapse to extensional, truth-functional operators whenever we assume the principles mentioned above (relativized to the right context). To illustrate, we reproduce a slingshot argument due to Quine which shows that if we substitute *salva veritate* logically equivalent sentences inside a modal operator, and we do the same thing for coreferential definite descriptions, then the operator collapses to an extensional, truth-functional one (an extensional operator is one that allows for substitutivity of materially equivalent sentences *salva veritate*). We use Quine's notation "δp" as an abbreviation for the definite description "$\iota x[(x = 1 \wedge p) \vee (x = 0 \wedge \neg p)]$."

As in the Church-Davidson slingshot, two observations are essential in the proof:

O5 "φ" and "$\delta\varphi = 1$" are logically equivalent.

O6 Whenever "φ" and "ψ" are true, "$\delta\varphi$" and "$\delta\psi$" designate the same individual.

Here is the argument (Quine 1960):

17. "φ" and "ψ" are true, Premise.
18. $\Box\varphi$, Premise.
19. $\Box(\delta\varphi = 1)$, (18), (SLE).
20. $\delta\varphi = \delta\psi$, (17), O6.
21. $\Box(\delta\psi = 1)$, (19), (20).
22. $\Box\psi$, (21), (SLE).

The reader may consult Burge (1986), Olson (1987), Neale (1995), and Stainton (forthcoming) for further discussions of slingshot arguments.

4. From the Universality of Logic to the Model-Theoretical View

Frege speaks of his symbolic system in the introduction of the *Begriffsschrift* (1879) as a *language*. It is a universal language intended to make transparent the laws of rational discourse. Frege, like Wittgenstein later on, is held to be a representative of what has been called the *universalist conception* of logic and language: In his conceptual language, quantifiers range over all possible objects from the absolute universe of discourse. The universe cannot be independently varied nor can the language and its relation to the universe be the subject of systematic explanations made in another (meta)language, because otherwise it would not be universal. For this reason, Frege's categorial distinction between concepts and objects and his remarks about the predicative character of the former are not to be taken as a full-fledged semantic *theory*, that is, as a systematic description of the semantic interpretations of the expressions of one's language, but rather as *elucidations*; the same remarks apply to Wittgenstein's *Tractatus*. This comes out quite clearly several times in Frege's work. To take a paradigmatic example, Frege remarks in "The Thought" that the property of a sentence being true cannot be defined because it is presupposed in one way or another by any attempts to define it. For suppose, he writes, that we could devise such a reductive definition

$$(*) \qquad P \text{ is true} =_{df} X,$$

where truth does not occur in X. Then, to be able to say whether P is true or not, we must be able to evaluate whether X. If we are entitled to assert X, then we can claim that P is true. But if we are entitled to assert X, Frege goes on, then we are entitled to assert that X is true, and we end up in relying on what we wanted to define in the first place. These considerations show how overwhelming the universalist conception of language was: Frege would not think, like Tarski later on, that a distinction between object language and metalanguage would avoid the circularity of the truth-definition by taking "X is true" to be an assertion in the metalanguage.

Wittgenstein reaches a similar conclusion. The truth of a proposition cannot be defined because when one tries to define it, one always ends up by repeating the proposition itself. "For what does it mean to say that a proposition is true? 'p' is true = p. This is the answer" (Wittgenstein 1978, appendix I, sec. 5). In a similar spirit, he writes: "The limit of language is shown by the impossibility of describing the fact that corresponds to an assertion without repeating the assertion itself" (Wittgenstein 1980, 10). The clue to all this lies, according to Wittgenstein, in the impossibility to go over the limits of language itself: "The impossibility to express in language the conditions of correspondence between a meaningful proposition, a thought and reality, this is the solution of the puzzle" (Wittgenstein 1978, 265).

In his introduction to the *Tractatus* Russell takes seriously Wittgenstein's claims to the effect that it is impossible to describe the syntactic and semantic properties of our language in that language itself. Russell agrees that each language has a structure that cannot be described in that language itself, but he suggests that there may be another language in which the structure of the first language can be described and yet another one in which the structure of the second language may be described and so ad infinitum. A systematic development of these ideas still needs to await some years.

Frege's and Wittgenstein's universalist conception of logic and language were shared largely by early Carnap and the members of the Vienna Circle, among whom there was a general agreement that metadiscourse about a particular object language must be excluded. These views began to be shaken little by little by Gödel and by Hilbert's results on independence, consistency, and completeness, for they showed that systematic investigations on the syntax and the semantics of a language could be obtained, contrary to what the universalists had previously asserted (see chapter 9).

Carnap's early work may be seen as an attempt to reconcile the one-world and one-language view of the universalists with the metalinguistic views of Hilbert's school. Carnap's way to reconcile them was to show that metalinguistic results could be formulated within the object language itself. The first such systematic attempts are contained in his lectures *Metalogik* (1931) where he constructed a very simple language, a precursor of *Language I* of the *Logical Syntax* to demonstrate his philosophical point. (These are contained in the section 18 of the *Logical Syntax*). Carnap thought that both approaches were wrong and proposed to show that one language was sufficient. He did not repudiate metalinguistic methods, but showed instead how the metalanguage could be formulated within the object-language (Carnap 1937, 53). Although the metalanguage of Carnap's lectures can be expressed in it, this metalanguage lacks the resources to prove important metatheoretical results like consistency or completeness, which can be proved only in a stricter metalanguage (in which one can carry transfinite induction), a fact Carnap became aware of later on (Carnap 1937, 38). So Carnap's project was doomed to fail.

Under Tarski's influence, the universal language of *Metalogik* started to lose its central role. From 1932 on, Carnap started gradually to desert the logicist ideal of universal language and began to think that the choice of a language is conventional. As a consequence, he gave up Frege's idea that truth is undefinable and primitive. The meaning of sentences is now explained in terms of the semantic meanings of its parts, which are specified in the metalanguage, following Tarski's method. The details of this development are described in Oberdan (1990, 1992).

5. Frank Ramsey

Perhaps the most ambitious endeavor to break up with the earlier tradition is that of Ramsey. His article "Facts and Propositions" (1927) made him one of the first proponents of the redundancy theory of truth, expressed in

The belief that p is true if and only if p.

However, in the manuscript *On Truth*, which has been published posthumously, Ramsey visibly tries to convert his conception into a precise definition in a way that clearly shows the ingredients of an inductive truth-definition. That is, he thought such an inductive definition should run through the complexity of p, but, interesting enough, he did not see any hope of limiting the form of p to something manageable in natural languages. He noticed that

> A man may be believing that all A are not B, that if A are B, either all C are D or some E are F, or something more complicated. (Ramsey 1991, 9)

Ramsey's reflections remind one of Tarski's remarks on natural languages (see following). Tarski gave up the attempt to define truth for "colloquial languages" on account of these languages being too vague and irregular to allow for an inductive definition. Instead he focused on formalized languages. There was, however, another development that prompted Tarski to look away from natural languages into the direction of formal languages: The universality of the former render them liable to the Liar paradox. Tarski's solution to this paradox was the distinction between object language and metalanguage. Since such a distinction does not seem to make sense for natural languages, formal languages offered him a fruitful instrument to work with.

6. Tarski's Truth-Definition

6.1. The Tradition of Correspondence

The most popular classical theory of truth has always been the *correspondence theory*. We can count the "semantic theory" formulated by Tarski as a subspecies of correspondence theories if we make a division between two types: Some forms of correspondence theory do require a metaphysically real or factual correspondence between linguistic entities and nonlinguistic world, other versions do not take a stand on such issues. In that weaker case the main challenge to the array of correspondence theories comes from various *deflationary* conceptions.

The theories of "weak correspondence" are metaphysically neutral. Their pursuit is to express the simple *logical* insight of Aristotle: "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true" (*Metaphysics* 1011b26–27). In his great treatise, Tarski especially stated his aim to formulate this principle completely, and as we shall see, this is what he achieved excellently. On the other hand, he said that he did not want to defend anything like a correspondence theory. But then he apparently was thinking of correspondence theory in the other sense, as "strong correspondence." The Tarskian semantic theory can probably be regarded as a purified correspondence theory. But perhaps this is not the case with Tarski's own first version of it, where he wanted to avoid all *semantic* concepts like "mean" and "express" and to provide all expressions only with extensions. (Compare Haack 1978, 112–114, and Kirkham 1992, 167–173.) The Tarskian correspondence has its well-known virtues but also the much debated formal restrictions, in the first place those resulting from the distinction between object-language and metalanguage.

It is interesting that some readers took the new semantic or "minimal" correspondence as a highly informative and even inspiring insight, whereas others thought that it was a desperate admission of failure. In spite of its careful formulation, and in spite of the fact that Tarski himself started its explanation in more philosophical terms in Tarski (1944), the definition has often been misunderstood. Thus nonprofessional public, more often than not, confuses the definition with the T-condition of material truth. In fact the crucial point in the definition procedure is the satisfaction relation, together with the recursive apparatus. Definitely the semantic or Tarskian theory is not "vacuous" or "trivial." (Concerning various objections to Tarski's theory see, e.g., Kirkham 1992, ch. 6.)

The semantic theory was not the only "weak" correspondence theory of the twentieth century. Thus J. L. Austin, who was reluctant to all formalized semantics, sketched his own version which emphasized the conventional nature of the correspondence. He argued that numerous linguistic conventions of meaning correlate expressions with appropriate situations which can make them true (Austin 1950). On the other hand, "strong" correspondence theories assume that truth depends on some actual relations of congruence, adequacy, or something like that—in fact *adaequatio* was the most traditional term, while "correspondence" is new, perhaps introduced by Russell. Twentieth-century discussion about strong correspondence begins with Russell's logical atomism. Russell (1912, ch. 12) already explains that a's belief that $P(b, \dots, n)$ is a relation $B(a, P, b, \dots, n)$, and that it is true if the relation $P(b, \dots, n)$ actually obtains. He so wants to reduce truth to a correspondence between metaphysically interpreted relations, and this idea is developed in the writings of his logical atomist phase.

A more advanced stage of a similar idea was in the picture theory of Wittgenstein's *Tractatus* (1922). According to it, a proposition, as a constellation of its elements, is structurally isomorphic with a state of affairs; it shows a state of affairs and claims that this state of affairs obtains. (See *Tractatus* 4.022 and 4.024.) Obviously this has immediate philosophical implications for the theory of truth, since a true sentence and a fact will stand in a strong pictorial correspondence, an isomorphism. But the picture theory is very enigmatic and not primarily logical, so we will not discuss it here, but only refer to the abundant literature on *Tractatus*. In fact the picture theory itself had rather little influence in logic, though it is by no means absurd or outdated. One drawback in all pictorial or isomorphist theories, however, lies in the difficulties with the metaphysics of facts.

It is clear that the so-called classical alternatives to the correspondence theory of truth were motivated mostly by philosophical considerations. Their real logical and semantic content has often been vague and hard to estimate. So we shall not try to say anything informative about them here, though they have had influential supporters during the twentieth century. The pragmatic theory has its defenders who build on the remarks of the founding fathers Peirce, James, and Dewey. In fact, these three seem to have had quite different ideas about truth, and expressing them clearly has turned out a difficult task. A common central feature in them, however, is the close conceptual connection between truth and acceptable beliefs. Truth has also been a central subject in the new versions of pragmatism. (Concerning Rorty and other recent developments, see Habermas 1996.) Coherence theory, which replaces semantic truth with some variant of syntactic coherence, was popular among idealists and has occasionally been advocated later; so Neurath (1933), and the recent "internal realism" sometimes comes close to the same position. (The fullest logical text is Rescher 1973.)

6.2. The Absolute Conception of Truth

Tarski presented his Polish version of *The Concept of Truth in Formalized Languages* in front of the Warsaw Philosophical Society in March 1931. It was published in Polish in 1933, translated in German in 1935, and then in English in 1956. His goal was the foundation of scientific semantics, where one does not mention any kind of abstract entities besides those mentioned by physics, mathematics, and logic. In this Tarski was on the same side as the logical positivists with whom he shared the common ideal of the unity of science.

Tarski opens his famous article by telling us that it is devoted to the definition of truth. This is not entirely correct, because in the end Tarski does not define truth but truth-in-L, that is, truth relative to a given language. In other words, truth is a linguistic predicate, which applies to sentences of a given language, for example, "Snow is white" is true-in-English.

Tarski required any definition of truth-in-L to be *materially adequate and formally correct.* (The reader is referred to chapter 9 for an explanation of these two notions as well as a detailed exposition of the background which led to Tarski's theory.) Here it is sufficient to recall that a formally correct definition of Tr is materially adequate if it obeys *Convention T*, that is, it has the following two consequences:

- All sentences which are obtained from the scheme

  $$Tr(x) \text{ if and only if } p$$

  by substituting for the symbol x the name (structural description) of any sentence of the language L and for the symbol p the expression which forms the translation of this sentence into the metalanguage;
- The sentence "for any x: if $Tr(x)$ then $S(x)$" where "$S(x)$" stands for "x is a sentence of L." In other words, it is required that the predicate Tr applies only to sentences. (Tarski 1956, 188)

Tarski believed that a truth-definition that satisfies the requirement of material adequacy captures the correspondence theory of truth encoded in the Aristotelian doctrine: "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true." He was, however, fully aware that the correspondence theory of truth and the philosophical discussions around it are too vague to serve as a foundation for a semantic definition of truth. Instead he devoted his efforts to showing that for a great deal of languages (formalized languages of the deductive sciences), it is possible to construct in the metalanguage a definition of truth which is both formally correct and materially adequate.

At the beginning of his treatise, Tarski deals with the problem of truth in natural ("colloquial") languages and shows that, on pain of inconsistency, they cannot have a truth-predicate which is materially adequate. (See following discussion.) In addition to inconsistencies arising from the paradoxes, Tarski mentions another reason which renders him skeptical about the possibility of developing a systematic theory of truth for natural languages: the impossibility of applying an inductive (compositional) method that first defines truth for atomic sentences and then reduces somehow the truth of complex sentences to that of atomic ones. This impossibility is connected to the fact that our colloquial language is not something finished, closed, or bounded by clear limits, and thereby one cannot specify structurally the expressions which constitute its sentences.

6.3. Tarski's Truth-Definition for Formalized Languages

In contrast to natural languages, formalized languages have certain features which render a definition of truth-sentence possible, namely:

1. All the signs from which the expressions of the language are formed may be given in a structural way.
2. The class of sentences of the languages may be specified by means of purely structural properties.
3. Each formalized language has grown together with a formal theory which is given by: (i) a finite list of axioms; (ii) a finite list of rules of inference; and (iii) on the basis of (i) and (ii) a characterization of the notion of a *provable* or *asserted* sentence. (ibid., 166)

Tarski points out that formalized languages do not have the universality of the natural languages, and for this reason, the description of their syntax (i.e., (1), (2), and (3)) must be undertaken in another language, the metalanguage, which must contain three kinds of expressions:

i. Expressions of a general logical kind;
ii. Expressions having the same meaning as all the constants of the object language;
iii. Expressions of a structural descriptive kind (terms) which denote the signs and the expressions of the object language. (ibid., 210–211)

The need of the expressions in group (i) is obvious. The expressions in group (ii) ensure that every meaningful expression of the object language has a translation in the metalanguage; and the terms belonging to group (iii) ensure that every meaningful expression in the object language has a name (structural description) in the metalanguage. The metatheory will have axioms which will ensure that all these requirements are met, that is, axioms which ensure that expressions of the general logical kind behave in the standard way; axioms which have the same meanings as the axioms of the science under investigation (or are logically stronger than them); and axioms which fix the properties of the structural descriptive names. As a result, one would have in the metatheory definitions of the syntax of the object language (definitions of the notions of term, formula, sentence) as well as a definition of the notion of being a theorem of the theory of the object language. All these details together with the definition of the notion of satisfaction of a formula by a sequence of objects have been described in chapter 9 (section 8.7) for the case in which the object language is that of the calculus of classes.

We recall one particular application of that framework. The structural descriptive name for the object language sentence "$\Pi x'\Sigma x'' I x' x''$" is $\cap_1 \cup_2 \iota_{1,2}$. The following T-biconditional is a consequence of the definition of satisfaction for the language of the calculus of classes (and the definition of *Tr* as satisfaction by every sequence): $\cap_1 \cup_2 \iota_{1,2}$ is true if and only if for all classes a there is a class b such that $a \subseteq b$.

The sentence "for all classes a there is a class b such that $a \subseteq b$" is the translation in the metalanguage of the sentence "$\Pi x'\Sigma x'' I x' x''$."

The T-biconditional does not yet give us the truth value of the sentence "$\Pi x'\Sigma x'' I x' x''$." To determine that, we have to know additional facts about the calculus of classes, namely, that the sentence on the right side is one of its theorems. This is indeed so, and we can conclude that $\cap_1 \cup_2 \iota_{1,2}$ is a true sentence (Tarski 1956, 196).

Let us, following Tarski, mention some of the remarkable properties his truth-definition has. The following are theorems of the metatheory (that is, their proofs are based on the general laws of logic, the special axioms of the metatheory, and the definitions of the concepts occurring in the theorems):

The principle of noncontradiction (*consistency*). For every sentence x of the calculus of classes, it is not the case that both x and its negation are true:

$$\forall x(Sen(x) \rightarrow \neg(Tr(x) \wedge Tr(Neg(x)))).$$

The principle of excluded middle. Every sentence of the calculus of classes is either true or its negation is true:

$$\forall x(Sen(x) \rightarrow Tr(x) \vee Tr(Neg(x))).$$

Tarski explicitly points out that the notion of formal proof as it emerges from Hilbert's and Gödel's works obeys only the principle of consistency but not the principle of excluded middle: Gödel has shown there are sentences of arithmetic which are undecidable (ibid., 199).

One of the most interesting properties of Tarski's notion of truth is the possibility it opens for proving a principle of soundness. When the calculus of classes is formalized, that is, when certain sentences are selected as axioms, rules of inference are fixed, and the notion of theoremhood ($Pr(x)$: x is a theorem of the formalized theory) is defined in the standard way, then all theorems are true (Tarski's Theorem 5):

The principle of soundness: $\forall x(Pr_{CL}(x) \rightarrow Tr(x))$.

7. Rudolf Carnap

7.1. Carnap's Semantical Frameworks

Tarski's insight according to which the formal method of syntax must be supplemented by a rigorous analysis of semantical concepts, exercised a deep influence on Carnap, who in a couple of books, *Introduction to Semantics* (1942) and *Meaning and Necessity* (1947) extended Tarski's methods. As Carnap notices in his preface to the *Introduction*, he demarcates himself from Tarski by laying a greater emphasis on the distinction between interpreted language systems and uninterpreted calculi (recall that for Tarski all formalized languages are interpreted), and more dramatically by stressing "the distinction between factual truth, dependent upon the contingency of facts, and logical truth, independent of facts and dependent merely on meaning as determined by semantical rules" (Carnap 1942, xi).

Carnap's semantical systems S consist of a set of rules that give the truth-conditions of the sentences of an object language in a metalanguage (usually English) and thereby determine the meaning of these sentences. The rules are of three kinds:

- Formation rules, which define the sentences of S.
- Designation rules, which determine, for some of the expressions of S, their designata.
- Rules of truth, which specify for each sentence of S the set of necessary and sufficient conditions for its truth.

Here is a simple semantical system restricted to the quantifier-free part of predicate logic:

1. Classification of signs:

   Individual constants, *in*: "a" and "b".

   Predicates, *Pr*: "P" and "Q".

   Other signs: "$\neg$", "$\vee$".

2. Rules of formation: An expression is a sentence if it has one of the following forms: $Pr(in)$, $\neg A$, $B \vee C$.
3. Rules of designation:

   i. "a" designates Chicago.

ii. "b" designates New York.
iii. "P" designates the property of being large.
iv. "Q" designates the property of having a harbor.

4. Rules of truth: The sentence A is true in S, if
   a. A has the form $Pr(in)$ and the object designated by in has the property designated by Pr.
   b. A has the form $\neg B$, and B is not true.
   c. A has the form $B \vee C$, and at least B is true or C is true.

   (Carnap 1942, 32)

As we see, Carnap assigns a designation to both individual constants (Tarski's formalized languages do not contain individual constants) and predicate symbols (to be compared with Tarski's nominalism!) Carnap's rules of truth match closely Tarski's recursive truth-definition. In addition, he formulates a condition of adequacy that is literally Tarski's condition of material adequacy (Convention T). Thus Carnap's semantical rules are designed in such a way as to yield adequate truth-conditions for the sentences of the system, for example,

"$Q(a)$" is true if and only if Chicago has a harbor.

As in the case of Tarski, these rules do not yet suffice to determine the truth value of the relevant sentences: "in order to find this we must know certain facts in addition to the rules. This would lead us outside of semantics into empirical science, in this case into geography" (Carnap 1942, 33). But Carnap emphasizes that his rules of truth yield the meaning of each sentence to which they apply.

7.2. Truth and L-Truth: The Analytic-Synthetic Distinction

Perhaps Carnap's most considerable departure from Tarski is his distinction between factual and logical truth (L-true). Logic, being concerned with the latter, which is a special kind of truth, becomes a special part of semantics.

The separation between the two notions is based on the distinction between logical and descriptive expressions. The latter are usually names and predicates, while the former are the sentential connectives, the standard quantifiers, the element-class relation "ϵ," the necessity operator and all the expressions definable from these ones. As we saw in the presentation of a semantical system, it is only the descriptive expressions that receive a designatum while each logical expression prompts a rule of truth.

Once a semantical system S is fixed, the sentences which are L-true in it get fixed, too: They are the sentences which are true in S solely in virtue of the semantical rules without any appeal to facts. Thus, to take one of Carnap's examples, "$P(a) \vee \neg P(a)$" is L-true. More exactly, in the propositional system we have described, a sentence is L-true if its truth table shows T for each distribution of truth values for its components (Carnap 1942, 81). This definition works, however, only for the particular semantic system sketched above. Carnap is aware that a more general definition which would work for arbitrary semantical systems is much harder to attain. In a note (p. 85) he sketches how this notion could be made more precise using logical necessity, but a systematic development of this idea is undertaken only later on in *Meaning and Necessity*.

Disregarding many of the details, one can extract two important elements which remain constant throughout Carnap's logical writings (*The Logical Syntax of Language*, *Introduction to Semantics*, *Formalization of Logic*, *Meaning and Necessity*):

i. The separation of sentences that are true or false in virtue of the meaning of their terms only, and not on the basis of extralinguistic facts. These sentences that he calls L-determinate (either L-true or L-false) came to be known as analytic sentences, in contradistinction to synthetic sentences whose truth is based on extralinguistic facts.

ii. The relativity of the distinction to the semantical framework.

Carnap's famous slogan expressed in *The Logical Syntax of Language*, "In logic, there are no morals. Everyone is at liberty to build up his own logic, that is, his own form of logic, as he wishes," has been taken to express his semantic conventionalism which leaves room for the pragmatic choice of linguistic frameworks, on condition that the rules of the framework are clearly stated.

Carnap's distinction came later under Quine's attack. Perhaps Quine's most forceful criticism is that all the statements in science are revisable in the light of new empirical data, including the laws of logic (Quine 1953, 43). This shows Quine to have taken Carnap's L-truth to be somehow unrevisable, a view that perhaps neglects (ii) above and the possibility it opens that one can always countenance a new semantical framework. Section 13.3 takes up some of these questions.

8. Truth and Proof

8.1. Consistency Proofs

In the postscript to the English translation of his seminal article (Tarski 1956), Tarski adds some interesting parallels between his results and those of Gödel. He notices an interesting corollary of one of his earlier theorems, $\forall x(Pr_{CL}(x) \rightarrow Tr(x))$, when applied to the language of arithmetic and the theory associated with it, Peano arithmetic (PA). He points out that theorems of this kind allow one to prove the consistency of the deductive science (i.e., PA), while, on the other side, Gödel has shown that the consistency of PA cannot be proved solely by PA.

Let us look more closely at this result by first introducing some technical notation. The language of arithmetic L_{PA} consists of the logical symbols "$\neg$", "$\vee$", "$\exists$", and "$=$", and of the special symbols "0", "+", "$\cdot$", "S". Each natural number n has a name $\underline{n}$ in the object language, that is "$\underbrace{S(S(\ldots S}_{\text{n times}}(0)\ldots)$". One of the virtues of PA is to represent the syntax of L_{PA} in the object language, thanks to a technique due to Gödel. Gödel's arithmetization method consists in associating with each expression α of L_{PA} a natural number $\ulcorner\alpha\urcorner$ (the gödel number of α). The gödel numbering is done in such a way that two different expressions are assigned two distinct gödel numbers, and there is an effective procedure which, given a natural number, calculates which expression (if any) it is the gödel number of. Furthermore, the gödel number $\ulcorner\alpha\urcorner$ is represented in the object language by the numeral $\underline{\ulcorner\alpha\urcorner}$. Also, the notion of being a theorem of PA (the analog of the provability predicate for the calculus of classes) may be represented in PA itself, as Gödel has shown: L_{PA} contains a nontrivial expression (provability predicate) $Pr_{PA}(x)$ such that

$$\text{If } PA \vdash \varphi \text{ then } PA \vdash Pr_{PA}(\underline{\ulcorner\varphi\urcorner})$$

for every L_{PA}-formula φ. Thus the analog of principle of soundness for PA can now be stated as

(S) $$\forall x(Pr_{PA}(x) \rightarrow Tr(x)),$$

where "Tr" is the Tarskian truth-definition for L_{PA} stated in the metalanguage. (S) is provable from the axioms of PA (the truth-predicate is allowed to occur in the induction axiom), the axioms of the truth-theory and general principles of logic (just as the soundness principle for the theory of classes was). Let us denote the theory consisting of all these axioms and logical principles by $PA(Tr)$. Recall the statement $\neg Pr_{PA}(\underline{\ulcorner\neg 0 = 0\urcorner})$, which is taken to assert the consistency of PA (that we abbreviate by Con). Now we have:

$PA(Tr)$ proves the consistency of PA:

$$PA(Tr) \vdash \neg Pr_{PA}(\underline{\ulcorner\neg 0 = 0\urcorner}).$$

The proof is straightforward. $PA(Tr) \vdash Pr_{PA}(\underline{\ulcorner\neg 0 = 0\urcorner}) \rightarrow Tr(\underline{\ulcorner\neg 0 = 0\urcorner})$ follows from (S). On the other side, $PA(Tr) \vdash Tr(\underline{\ulcorner\neg 0 = 0\urcorner}) \leftrightarrow \neg 0 = 0$ is one of the T-biconditionals. But $PA(Tr) \vdash 0 = 0$, whence $PA(Tr) \vdash \neg Tr(\underline{\ulcorner\neg 0 = 0\urcorner})$, which together with (S) entails $PA(Tr) \vdash \neg Pr_{PA}(\underline{\ulcorner\neg 0 = 0\urcorner})$, that is, $PA(Tr) \vdash Con(PA)$.

This result shows one major difference between Gödel's notion of formal proof and Tarski's notion of truth. Gödel's second incompleteness theorem shows that $PA \nvdash Con(PA)$ and $PA \nvdash \neg Con(PA)$, whereas Tarski's theory of truth proves $Con(PA)$:

> Moreover Gödel has given a method for constructing sentences which—assuming the theory concerned to be consistent—cannot be

> decided in any direction in this theory. All sentences constructed according to Gödel's method possess the property that it can be established whether they are true or false on the basis of the metatheory of higher order having a correct definition of truth. (Tarski 1956, 274)

This property of Tarski's theory of truth marks a clear advantage over its predecessors that can be spelled out using the technical notion of conservativity. Let T be a theory (set of sentences) formulated in the language L and T' be another theory formulated in the language L' such that $T \subseteq T'$ and $L \subseteq L'$. We say that T' is a *conservative extension* of T, if for every sentence φ of L:

$$T' \vdash \varphi \Longrightarrow T \vdash \varphi.$$

In other words, if T' is conservative over T, then T' does not prove any sentence in the language of T that T itself does not prove. The fact that $PA(Tr)$ proves the consistency of PA shows its nonconservativity over PA. This feature of Tarski's theory of truth is not shared by so-called disquotationalist or deflationist truth theories. The latter are usually formed by adding to the base theory (PA) all the T-biconditionals. If the base theory is PA, let us denote by $PA(DT)$ the theory obtained from adding to PA all T-biconditionals. Unlike $PA(Tr)$, $PA(DT)$ is conservative over PA (even in case we allow the induction scheme to contain the truth-predicate).

8.2. Tarski's Impossibility Result

Let us conclude the presentation of Tarski's contribution to semantics by resuming his legacy in the form of two main results that he himself mentions as the achievements of his work.

A positive result: For every formalized language a formally correct and materially adequate definition of true sentence can be constructed in a richer metalanguage (i.e., metalanguage of higher order).

A negative (impossibility) result: If the order of the metalanguage is at most equal to that of the language itself, such a definition cannot be constructed.

To be fully aware of what the impossibility result amounts to, let us see how it applies to the language of arithmetic L_{PA}. The syntax of this language, that is, such notions as being a term, a formula, a sentence of L_{PA} are defined in a metalanguage, but then, due to the Gödelian technique mentioned earlier, these definitions can be actually represented in L_{PA} itself. (This is another way of saying that L_{PA} defines its own syntax.)

In the next stage, one formulates the truth-definition in the metalanguage, which, as Tarski points out in his conclusions, must be of higher order than the object language. In the present case, the metalanguage will be a second-order language.

The truth-definition will first take the form of a set of axioms that show how the satisfaction of a complex L_{PA}-formula depends on the satisfaction of its simpler parts. Given that every natural number has a name in L_{PA}, we can dispense with the notion of satisfaction here. Let X be a second-order variable. Consider the following axioms, where we may think of $X(x)$ as "x has the property X":

1. If $X(x)$, then x is a sentence (of L_{PA});
2. If x is an atomic sentence of the form $t_1 = t_2$, then $X(x)$ iff the value of t_1 is identical to the value of t_2;
3. If y is a sentence and x is a negated sentence of the form $\neg y$, then $X(x)$ iff not $X(y)$;
4. If y is a sentence and z is a sentence and x is the disjunction of y and z, then $X(x)$ iff $X(y)$ or $X(z)$;
5. If i is a number, y is a formula and x is formed from y by existentially quantifying over x_i, then $X(x)$ iff there is a closed term t such that $X(y/t)$ ((y/t) is the result of the substitution of t for x_i in y).

Let $\Phi(X)$ denote the conjunction of these axioms. The explicit truth-definition is the second-order formula

$$\forall X(\Phi(X) \rightarrow X(x)),$$

or rather

$$Tr(x) \leftrightarrow \forall X(\Phi(X) \rightarrow X(x)),$$

which says that x is true if it has the property X for every X which satisfies the axioms (1)–(5)). Equivalently,

$$Tr(x) \leftrightarrow \exists X(\Phi(X) \wedge X(x)).$$

Tarski's impossibility result tells us that it is impossible to replace "$\forall X(\Phi(X) \rightarrow X(x))$" (or "$\exists X(\Phi(X) \wedge X(x))$," for that matter) with a first-order formula in the language of arithmetics.

8.3. Truth in a Model

The notion of truth in a model is usually traced back to a remark Tarski makes on page 199 of his famous paper where he mentions the idea of defining truth in a special domain which is a subset of the universal domain. (See chapter 9.) It emerged gradually from Tarski's and Carnap's work, and is now a standard notion in logical textbooks. It takes the form of a model $M = (D, I)$ for a given object language, that is, a domain of discourse (universe) D which is a set, and an interpretation function I which fixes the reference and denotations in the universe of M of the constant, predicate, and function symbols of the object language. One can recognize here some of the ingredients of Carnap's semantical systems: The interpretation function I assignes a "designatum" for each "descriptive" expression of the object language. For simplicity, we consider only the case in which each individual a of the domain D has a name $\underline{a}$ in the object language. Each closed term t in the language has a semantic value t^M which is an element of D. Thus for each constant symbol c, $c^M = I(c)$, and if the values of the closed terms $t_1, \ldots, t_n$ are $t_1^M, \ldots,$ and t_n^M, respectively, then the value of the closed term $ft_1, \ldots, t_n$ is the value of the function which is the interpretation of f applied to the arguments $t_1^M, \ldots,$ t_n^M, that is, $I(f)(t_1^M, \ldots, t_n^M)$.

The notion to be defined (in the metalanguage) is not any longer *truth-in-L* but *truth-in-M* ($M \vDash \varphi$). The definition is given by induction on the complexity of the sentence φ:

1. $M \vDash R(t_1, \ldots, t_n) \Leftrightarrow (t_1^M, \ldots, t_n^M) \in I(R)$,
2. $M \vDash t_1 = t_2 \Leftrightarrow t_1^M = t_2^M$,
3. $M \vDash \neg\varphi \Leftrightarrow M \nvDash \varphi$,
4. $M \vDash \varphi \vee \psi \Leftrightarrow M \vDash \varphi$ or $M \vDash \psi$,
5. $M \vDash \exists x_n \varphi \Leftrightarrow$ there is $a \in D$ such that $M \vDash \varphi(\underline{a})$.

Notice that the symbols M, D, I, and $\in$ belong to the metalanguage, and (1)–(5) are statements in the metalanguage. Let M be a model for a language L and *Th* a set of sentences (i.e., a theory) of L. We say that M is a model of *Th* if every sentence of *Th* is true-in-M. A first-order order theory may have more than one model. This is the case with *PA*. It has a standard model $N = (\mathbb{N}, I)$ whose universe $\mathbb{N}$ is the set of natural numbers, and in addition $I(0) = 0$, $I(S)$ is the ordinary successor function on $\mathbb{N}$, and $I(+)$ and $I(\times)$ are the usual operations of addition and multiplication, respectively. Every individual n in the universe has a name in L_{PA}, the numeral $\underline{n}$. It is well known that *PA* has also nonstandard models, that is, models whose universe contains individuals distinct from every natural number.

Tarski's technique of defining truth in the metalanguage also applies to the notion of truth in a model. Recall the second-order formula $\forall X(\Phi(X) \to X(x))$ which is the explicit truth-definition for L_{PA}. Let us denote it by $\Delta(x)$. It is routine to check that $\Delta(x)$ defines truth-in-$\mathbb{N}$ for the language L_{PA}, that is, the condition of material adequacy is fulfilled:

$$\mathbb{N} \vDash \Delta(\underline{\ulcorner \varphi \urcorner}) \leftrightarrow \varphi$$

for each first-order L_{PA}-sentence φ. Actually it may be shown that the formula $\forall X(\Phi(X) \to X(x))$ defines truth-in-M for every model M of *PA*.

Tarski's impossibility result applies to the present case, too: $\Delta(x)$ cannot be replaced by a first-order formula in the language of L_{PA}.

The reader is referred to Philippe de Rouilhan and Serge Bozon (2006) for an elaboration of the distinction between truth-in-a model versus truth in all models in Tarski's work.

9. Partial Interpretations

Tarski's proposal can be viewed as a hierarchical account of truth for formal languages. According to it, one does not have truth *simpliciter* but truth indexed with a language L_0. In the general case, one starts with a truth-free base language L_0. The truth-predicate Tr_0 for L_0 is given in the metalanguage L_1 which contains names for the sentences of L_0 in the way indicated (except that the language of arithmetics contains names for its expressions in the object languages itself). Thus if φ_1 and φ_2 are sentences of L_0, and $\ulcorner\varphi_1\urcorner$ and $\ulcorner\varphi_2\urcorner$ are their standard names in the language L_1, "$Tr_0(\ulcorner\varphi_1\urcorner) \wedge \neg\, Tr_0(\ulcorner\varphi_2\urcorner)$" is an assertion in L_1 saying that φ_1 is true (in L_0) and φ_2 is false (in L_0). But now if one wants to speak about the truth and falsity of the sentences of L_1, we have to go to the metalanguage L_2 of L_1 and introduce a truth-predicate Tr_1 for L_1. If $\ulcorner Tr_0(\ulcorner\varphi_1\urcorner) \wedge \neg\, Tr_0(\ulcorner\varphi_2\urcorner)\urcorner$ is the standard name of "$Tr_0(\ulcorner\varphi_1\urcorner)\wedge\neg\, Tr_0(\ulcorner\varphi_2\urcorner)$" in L_2, then $Tr_1(\ulcorner Tr_0(\ulcorner\varphi_1\urcorner)\wedge\neg\, Tr_0(\ulcorner\varphi_2\urcorner)\urcorner)$ asserts in L_2 that "$Tr_0(\ulcorner\varphi_1\urcorner) \wedge \neg\, Tr_0(\ulcorner\varphi_2\urcorner)$" is true (in L_1).

There are various difficulties associated with this proposal, of which probably the most famous is due to Kripke (1975). According to Kripke, it is perfectly meaningful of, say, Nixon to say that

Everything that Dean says is false

and of Dean to say

Everything Nixon says is true.

In this example, it seems that Nixon's assertion occurs in the metalanguage of Dean's discourse, and Dean's assertion in the metalanguage of Nixon's discourse, a situation that cannot be consistently modeled in Tarski's hierarchy of levels.

There have been various attempts to overcome the limitations of Tarski's account by defining the truth-predicate in the object languages themselves. We shall sketch two of them, each allowing for truth value gaps: Kripke's fixed point construction and *IF*-languages (Hintikka and Sandu 1989).

9.1. Partial Interpretations: Kripke

Kripke (1975) shows that there are languages which define their own truth-predicate (in a sense to be defined) in certain models. He shows that there is a model M of a language L that defines its own syntax and contains a truth-predicate $Tr(x)$ whose extension coincides with the class of sentences of L true in M and whose counterextension coincides with the class of sentences false in M. The result in the next section shows that there are languages which define truth in all models (abstract concept of truth). Let us say right away that none of these results contradicts Tarski's impossibility result: In both cases, the law of excluded middle fails.

We shall present Kripke's theory for the first-order language L_{PA} of arithmetic. We add to it a truth-predicate Tr and form the extended language $L^+_{PA} = L_{PA} \cup \{Tr\}$. On the interpretational level, recall the standard structure $(\mathbb{N}, I)$ for L_{PA}. We add to it two new elements I^+, I^- and form the structure $M = (\mathbb{N}, I, I^+, I^-)$ for the language L^+_{PA}. I^+, I^- are functions which interpret the truth-predicate in a partial way: $I^+(Tr)$ is a set of natural numbers, the extension of Tr; $I^-(Tr)$ is a set of natural numbers, the counterextension of Tr, disjoint from $I^+(Tr)$. If we now recall that sentences φ of $\mathcal{L}^+$ are associated with natural numbers $\ulcorner\varphi\urcorner$, then the universe $\mathbb{N}$ may be seen as divided into (a) sentences that belong to the extension $I^+(Tr)$ of the truth-predicate; (b) sentences that belong to its counterextension $I^-(Tr)$; and (c) nonsentences.

Truth and falsity of sentences of $\mathcal{L}^+$ in the partial model $M = (\mathbb{N}, I, I^+, I^-)$ are determined by the *Strong Kleene valuation scheme*: We define simultaneously two semantic values, truth-in-M, $M \models^+ \varphi$, and false-in-M, $M \models^- \varphi$. Here is the definition:

1. $M \models^+ R(t_1, \ldots, t_n) \Leftrightarrow (t_1^M, \ldots, t_n^M) \in I(R)$,
2. $M \models^- R(t_1, \ldots, t_n) \Leftrightarrow (t_1^M, \ldots, t_n^M) \notin I(R)$,
3. $M \models^+ Tr(t) \Leftrightarrow t^M \in I^+(Tr)$,
4. $M \models^- Tr(t) \Leftrightarrow t^M \in I^-(Tr)$,
5. $M \models^+ t_1 = t_2 \Leftrightarrow t_1^M = t_2^M$,
6. $M \models^- t_1 = t_2 \Leftrightarrow t_1^M \neq t_2^M$,
7. $M \models^+ \neg\varphi \Leftrightarrow M \models^- \varphi$,
8. $M \models^- \neg\varphi \Leftrightarrow M \models^+ \varphi$,
9. $M \models^+ \varphi \vee \psi \Leftrightarrow M \models^+ \varphi$ or $M \models^+ \psi$,
10. $M \models^- \varphi \vee \psi \Leftrightarrow M \models^- \varphi$ and $M \models^- \psi$,
11. $M \models^+ \exists x_n \varphi \Leftrightarrow$ there is $n \in \mathbb{N}$ such that $M \models^+ \varphi(\underline{n})$,
12. $M \models^- \exists x_n \varphi \Leftrightarrow$ for every $n \in \mathbb{N} : M \models^- \varphi(\underline{n})$.

Kripke motivates his construction in terms of a learning process which takes place in stages. The initial stage 0 is illustrated by a model $M = (\mathbb{N}, I, I^+, I^-)$ in which both the extension and the anti-extension of the truth-predicate are empty (i.e., $I^+(Tr) = I^-(Tr) = \varnothing$), and thus in conformity with the Strong Kleene schema, there is no sentence of the form $Tr(\underline{\ulcorner\varphi\urcorner})$, which is true, and no sentence of the form $\neg\, Tr(\underline{\ulcorner\varphi\urcorner})$, which is false at this stage. At stage 1, the extension of the truth-predicate contains all sentences that are true at stage 0, and $I^-(Tr)$ all the sentences that are false at that stage. As the extension of the truth-predicate is empty at stage 0, the sentences true there do not contain the truth-predicate but are all the semantic-free sentences true in the ground model. Thus, at stage 1, $I^+(Tr)$ contains sentences like $\ulcorner 0 = 0\urcorner$, $\ulcorner 2+1=3\urcorner$, $\ulcorner \exists x x + 0 = 1\urcorner$, and so on, and $I^-(Tr)$ sentences like $\ulcorner 0 = 1\urcorner$, and so on. But as $I^+(\mathit{Tr})$ contains $\ulcorner 0 = 0 \urcorner$ at stage 1, the Strong Kleene scheme tells us that $\mathit{Tr}(\underline{\ulcorner 0 = 0 \urcorner})$ holds at that stage, which means that $\ulcorner \mathit{Tr}(\underline{\ulcorner 0 = 0 \urcorner}) \urcorner$ will enter into the extension of the truth-predicate at stage 2. By a repeated process, we learn at every stage that more and more sentences are true, and the same goes for the false ones.

The transition from one stage to another is formally captured by Kripke's jump operator ρ whose arguments and values are pairs $\langle E^+, E^- \rangle$ where E^+ and E^- are the extension $I^+(\mathit{Tr})$ and respectively the anti-extension $I^-(\mathit{Tr})$ of the truth-predicate. Intuitively $\rho(\langle E^+, E^- \rangle) = \langle P^+, P^- \rangle$ means that the "learning process" has moved from the stage $(\mathbb{N}, I, I^+, I^-)$ where $I^+(\mathit{Tr}) = E^+$ and $I^-(\mathit{Tr}) = E^-$, to the stage $(\mathbb{N}, I, I_1^+, I_1^-)$ where $I_1^+(\mathit{Tr}) = P^+$ and $I_1^-(\mathit{Tr}) = P^-$. The construction guarantees that ρ is monotonic, that is, if $E^+ \subseteq P^+$ and $E^- \subseteq P^-$, then $\rho(\langle E^+, E^- \rangle) \subseteq \rho(\langle P^+, P^- \rangle)$, where the latter is just a notational device to say that the left member of $\rho(\langle E^+, E^- \rangle)$ is included in the left member of $\rho(\langle P^+, P^- \rangle)$ and the same holds for the right members. The monotonicity of ρ ensures that when moving from one stage to another more sentences are put into either the extension or counterextension of Tr (or both). Thus,

$$\langle \varnothing, \varnothing \rangle \subseteq \rho(\langle \varnothing, \varnothing \rangle) \subseteq \rho(\rho(\langle \varnothing, \varnothing \rangle)) \subseteq \cdots$$

At some point, the construction stops as there are no sentences to be put there any longer. In technical terms, we say that ρ has a fixed point, that is, there is a pair $\langle E^+, E^- \rangle$ such that $\rho(\langle E^+, E^- \rangle) = \langle E^+, E^- \rangle$. The existence of the fixed point has important consequences, as we have for every sentence φ

$$(\mathbb{N}, I, E^+, E^-) \vDash^+ \mathit{Tr}(\underline{\ulcorner \varphi \urcorner}) \Longleftrightarrow (\mathbb{N}, I, E^+, E^-) \vDash^+ \varphi$$
$$(\mathbb{N}, I, E^+, E^-) \vDash^- \mathit{Tr}(\underline{\ulcorner \varphi \urcorner}) \Longleftrightarrow (\mathbb{N}, I, E^+, E^-) \vDash^- \varphi.$$

For a monotonic ρ, Kripke's construction does not yield only one fixed point, but many. We do not have the space here for detailing the philosophical literature which defends the least fixed point as being the most adequate for representing truth. The reader is referred to Haack (1978) and Kirkham (1992) for two representative samples of such a view.

9.2. Partial Interpretations: IF-Languages

An attempt to overcome Tarski's second impossibility result (i.e., languages defining truth in all models) is given within the so-called *IF*-languages introduced in Hintikka and Sandu (1989) (see Hintikka 1996; Hodges 1997). These languages express more quantifier dependences and independences than ordinary first-order languages whose extensions they are. More exactly, the object language contains, in addition to first-order formulas, formulas of the form

(1) $$\forall x_0 \exists x_1 \forall x_2 (\exists x_3 / \{x_0, x_1\}) \varphi(x_0, x_1, x_2, x_3),$$

where φ is a first-order formula. The slash is an outscoping device intended to express the fact that an existentially quantified variable is functionally dependent on all the variables that precede it, except the ones that are slashed. Thus x_1 is functionally dependent on $\{x_0\}$ and x_3 is functionally dependent on $\{x_0, x_1, x_2\} - \{x_0, x_1\} = \{x_2\}$. The truth of (1) is given by asserting that there are certain (Skolem) functions that express the dependences between existential and universal quantifiers and which satisfy the formula φ. In the present example, this means there are two functions, the first one having x_0, and the second one having x_2 as an argument, such that $\varphi(x_0, f(x_0), x_2, g(x_2))$ holds:

$$\begin{aligned}&\forall x_0 \exists x_1 \forall x_2 (\exists x_3/\{x_0, x_1\}) \varphi(x_0, x_1, x_2, x_3)\\ \Longleftrightarrow\ &\exists f \exists g \forall x_0 \forall x_2 \varphi(x_0, f(x_0), x_2, g(x_2)).\end{aligned}$$

The motivation for this translation goes back to a game-theoretical interpretation due to Henkin (1961): There are two teams of players, the universal team consisting of two players, $\forall_0$ and $\forall_2$, and the existential team, consisting of two players, $\exists_1$, $\exists_3$, choosing elements from the universe of the model in which the nonlogical expressions of φ are interpreted. Thus $\forall_0$ chooses a, $\forall_2$ chooses b, $\exists_1$ chooses c, and $\exists_3$ chooses d. The slash indicates the fact that when choosing d, $\exists_3$ ignores the choice of a and that of c, knowing thus only the earlier choice of b. The two functions represent the strategies of the existential players: The sentence $\forall x_0 \exists x_1 \forall x_2 (\exists x_3/\{x_0, x_1\}) \varphi(x_0, x_1, x_2, x_3)$ is true (in a model) if each one of the existential players has a strategy so that together they constitute a win (i.e., make φ true) against any choices of the universal players.

We shall abbreviate the prefix $\forall x_0 \exists x_1 \forall x_2 (\exists x_3/\{x_0, x_1\})$ by $Hx_0x_1x_2x_3$. Recall the first-order language of arithmetics, L_{PA}. We enrich it with formulas of the form $Hx_0x_1x_2x_3\varphi$, where φ is a first-order formula of L_{PA}. A Tarski-style truth-definition can now be given for the new language, using our earlier truth-definition $\exists X(\Phi(X) \wedge X(x))$ for L_{PA} that we shall abbreviate by $\Phi(x)$. The truth-definition has two clauses:

a. If x is a first-order sentence of L_{PA}, then $\Phi(x)$.

b. If y is an L_{PA}−formula and x is an *IF* sentence of the form $Hx_0x_1x_2x_3y$, then there are functions f and g such that for all closed terms t_1 and t_2: $\Phi(y(t_1, f(t_1), t_2, g(t_2)))$.

(Here $y(t_1, f(t_1), t_2, g(t_2))$ is the result of the substitution of $t_1, f(t_1), t_2$ and $g(t_2)$ for x_0, x_1, x_2, and x_3, respectively.) It can be shown that for every model M of PA we have

$$M \vDash \Phi(\underline{\ulcorner \varphi \urcorner}) \Longleftrightarrow M \vDash \varphi,$$

for every *IF* sentence φ.

There are several steps that show the formula $\Phi(x)$ to be equivalent (in every model of PA) to an *IF* formula $\Theta(x)$. The reader is referred to Sandu (1996, 1998) and Hyttinen and Sandu (2000) for some of the missing details.

A couple of things need to be said in this connection.

Kripke (1975) shows that a certain first-order language L defining its elementary syntax contains a truth-predicate Tr for itself. The result presented in this section shows the same thing for *IF*-languages. There is, however, a major difference between the two results that can be spelled out by way of a distinction introduced in de Rouilhan and Bozon (2006) between a definition being *e-adequate* (yielding the right extension of the truth-predicate in the model under consideration) and a definition being *i-adequate* (fixing the correct extension of the truth-predicate in every model that satisfies the principles of elementary syntax of L). The property of i-adequacy guarantees that the definition fixes the intension of the predicate to be defined (modulo the principles of elementary syntax). We recall in this context Carnap's semantical systems and his emphasis on his rules of truth fixing the meaning of the sentences to which the truth-predicate applies.

We are now in a position to see that Tarski's truth-definitions for formalized languages are i-adequate (and of course e-adequate), whereas Kripke's construction is only e-adequate. The truth-definition for the *IF*-languages, being a Tarski-style truth-definition, is also i-adequate.

Finally, let us emphasize that results similar in spirit to the one presented in this section have been given before. Myhill (1950) and Smorynski (1977) proved the existence of nonclassical languages containing an adequate truth-predicate for themselves. Like *IF*-languages and Kripke's languages, none of them are closed under both universal quantification and classical negation.

9.3. Some Problems with the Three-Values Approach

Several problems have been pointed out in connection with the three-values approach. Perhaps the most debated one is that these languages are not able to express certain notions used in the metalanguage, like the notion of complementary negation and the classical notion of implication. $\varphi \rightarrow \varphi$ is not valid in Kripke's system or in *IF*-logic given the fact that $\psi \rightarrow \theta$ is defined as $\neg\psi \vee \theta$ and $\neg$ is not contradictory negation. Thus, in Kripke's system $\varphi \rightarrow \varphi$ does not hold for sentences that are neither true nor false (like the Liar) and in *IF*-logic $\varphi \rightarrow \varphi$ fails even for very simple sentences like $\forall x_0(\exists x_1/\{x_0\})x_0 = x_1$. For this reason not all instances of the T-biconditionals $Tr(\ulcorner\varphi\urcorner) \leftrightarrow \varphi$ hold. The reader is referred to Beall (2005) for a discussion of the problems related to these issues in Kripke's system; to Feferman, de Rouilhan, and Bozon for a criticism of *IF*-logic along the same lines; and to Hintikka's replies to these criticisms, all contained in Hintikka's Schilpp volume (Auxier and Hahn 2006). We shall come back to some of these issues when discussing the paradoxes. For the moment let us mention one more question in connection with Kripke's fixed point construction. It has been pointed out that, being a partial interpretation, this construction does not lend itself to natural axiomatizations which would somehow have to be not classical but partial. Kripke himself was aware of this problem and suggested in his (1975) paper a method of dealing with it. There he played with the "closing off" fixed points construction. To go back to his fixed point model, one closes it off by putting all the sentences which are semantically undecided in that model into the anti-extension of the truth-predicate. The resulting model is a classical one in which the previously undecided sentences, including the Liar, are false. Following up this idea, Feferman (1991) devised an elegant construction that axiomatizes Kripke's fixed point models in classical logic. The resulting axiom system, called the Kripke–Feferman (*KF*) system has as its natural interpretations the closing off models suggested by Kripke.

10. Revision Theory of Truth: Herzberger and Gupta

This is a theory developed in the eighties by Herzberger and independently by Gupta. Unlike Kripke's theory and *IF*-logic, both approaches preserve classical logic. Here we shall focus on the variant due to Gupta and Belnap (1993), reviewed in Beall (2005). The spirit of the whole enterprise should be obvious from the title. In Kripke's case one starts with an empty extension and counterextension for the truth-predicate, and then one learns gradually that more and more sentences are true and false, until a saturation point is fixed. In the present case, one starts with an extension of the truth-predicate based on an initial guess, reaching the right extension after a gradual process of revisions.

Consider a model M with universe $\{a, b, c, d\}$ and the following information on the extensions of the predicates F and H:

i. $a, b \in F^M$,

ii. $c, d \notin F^M$,

iii. $a, c \in H^M$,

iv. $b, d \notin H^M$.

In addition, we are given the following condition on the extensions of F, H, and G:

$$Gx \leftrightarrow (Fx \wedge Hx) \vee (Fx \wedge \neg Hx \wedge Gx) \vee (\neg Fx \wedge Hx \wedge \neg Gx).$$

If the initial guess is that G^M is $\varnothing$, then we have the following calculations.

Since $a \in F^M$ and $a \in H^M$ then the first disjunct is true, whence $a \in G^M$. Given that $c \notin F^M$ and $c \in H^M$ and the initial guess, then the third disjunct is true, whence $c \in G^M$. A quick inspection shows that b falsifies all the three disjuncts, whence $b \notin G^M$. And the same holds of d. The initial guess has been revised to $a, c \in G^M$ and $b, d \notin G^M$. But now the process can go on and eventually lead to another revision.

The revision process should be now clear.

> At stage S_0 one starts with an initial guess h_0 of the extension of G, and revises it to G^{h_0};

At stage S_1 one's guess is G^{h_0} and the revised extension is G^{h_1}.

$\vdots$

At stage S_{n+1} one's guess is G^{h_n} and the revised extension is $G^{h_{n+1}}$, and so on.

There are different starting points. For instance, with an universe like $\{a, b, c, d\}$ there may be 16 initial guesses. And thus there may be 16 different sequences (which eventually can go on into the transfinite).

We call a sentence φ *stable* relative to a sequence s if there is a stage in s after which the truth value of φ remains constant.

We say that the sentence φ exhibits convergence if φ is stable relative to all sequences s with the same truth value.

The following pattern emerges in the present example.

1. Object a. If the initial guess is that $a \in G^M$, then $a \in G^{h_n}$ for all n. If the initial guess is that $a \notin G^M$, then $a \in G^{h_n}$ for all $n > 0$. Thus it can be checked that the following sequence obtains:

$$\langle S_0, \varnothing\rangle, \langle S_1, \{a, c\}\rangle, \langle S_2, \{a\}\rangle, \langle S_3, \{a, c\}\rangle, \langle S_4, \{a\}\rangle, \ldots$$

   So we may expect the sentence Ga to exhibit convergence, given that it is stable relative to two sequences.

2. Object b. If the initial guess is that $b \in G^M$, then $b \in G^{h_n}$ for all n. If the initial guess is that $b \notin G^M$, then $b \notin G^{h_n}$ for all n. Thus Gb is stable in both cases, but it fails to be convergent.

3. Object c. There is no stability with respect to Gc, and hence no convergence.

4. Object d. Identical to case a, except that Gd converges on falsity.

We reach the "categorical" judgments that a is G and that d is not G. On the other side, there are no categorical judgments to be made of b or c.

In the case of truth, Tarski biconditionals should be understood as having a hypothetical character providing a method for obtaining better and better approximations of the extension of the truth-predicate. Thus the role played in our earlier example by the circular definition of Gx is now played by Tarski's biconditionals. Given that some sentences may contain occurrences of Tr, it is clear that these biconditionals may be regarded as circular definitions. Starting with an arbitrary extension of the truth-predicate, the revision process results in a series of models M^*, M^{**}, M^{***}, ... that are constructed using the biconditionals by evaluating sentences in the previous member of the series. Each member in the series will be an improvement over its predecessor. In the end, a categorical judgment will be eventually reached. Some sentences will converge, others will stabilize but not converge, and still others will fail to stabilize. In general, the Liar sentence will turn out to be unstable. The reader is referred to Chapuis and Gupta (2000) for a collection of articles that deal with paradoxes in the context of the revision theory of truth.

11. Davidson: Tarski's Theory of Truth as a Theory of Meaning

Donald Davidson developed his views in a series of papers, which include *Truth and Meaning* (1967), *Semantics for Natural Languages* (1970), *In Defence of Convention T* (1973) all included in *Inquiries into Truth and Interpretation* (1984). The result is a theory of interpretation, which, one could say, is a combination of Tarski's theory of truth with Quine's views on radical translation: Tarski's object language becomes the language to be interpreted, the metalanguage becomes the language of the interpreter, and the process of interpretation itself becomes an empirical process subject to Quine's constraints on radical translation.

The starting point of Davidson's reflexions is the simple fact that we are able to interpret other people's linguistic utterances:

> Kurt utters the words "Es regnet" and under the right conditions we know that he has said that it is raining. Having identified his utterance as intentional and linguistic, we are able to go on to interpret his words: we can say what his words, on that occasion, meant. (Davidson 1984, 125)

Hence he finds it natural to ask: What do we know that enables us to interpret the words of others? The answer is: a theory of meaning which is the same thing as a theory of interpretation. Davidson sets to himself the task to formulate such a theory but before doing that, he displays three requirements any such theory should meet:

1. It should explain how the interpreter, which has finite knowledge, is able to interpret a potentially infinite number of sentences.
2. It should be informative: It cannot, for instance, be just a function which maps spoken utterances into understood utterances.
3. It should be empirically testable.

An adequate theory of meaning should entail for every sentence s of the language under investigation the sentence

$$s \text{ means that } p,$$

where "p" is to be replaced with a sentence in the language of the interpreter. The role of "p" is to give the "meaning" of the sentence s. The crucial question is, of course, what to put in the place of "p."

As in the case of the truth-predicate, one can think of "means that p" as denoting a predicate or a property, X, which applies to sentences. Thus on Davidson's view, we should have in the place of "p" a sentence that gives the necessary and sufficient conditions for the sentence s to have the property X. This step is described by Davidson in the following passage:

> As a final bold step, let us try treating the position occupied by "p" extensionally: to implement this, sweep away the obscure "means that," provide the sentence that replaces "p" with a proper sentential connective, and supply the description that replaces "s" with its own predicate. The plausible result is (T) s is T if and only if p. (Davidson 1984, 23)

In other words, after explaining the scheme and replacing intensional by extensional idiom, we end up with

(M) s has the property X (s is X) if and only if p.

But now we realize that the property X is coextensional with truth in Tarski's sense:

$$s \text{ is true if and only if } s \text{ is } X.$$

This can be seen in the following way: Suppose that s is true (in Tarski's sense). By Convention T we know that

$$s \text{ is true if and only if } p.$$

Whence p. But then from scheme (M) we get:

$$s \text{ is } X.$$

The argument in the other direction is established in a similar way. On the basis of the argument, Davidson reaches the conclusion that the theory of interpretation we are looking for is already on the market: It is Tarski's theory of truth, provided it undergoes several modifications.

- Tarski's definition of truth is not regarded as a definition any longer but as a *theory*. In other words, the definition of truth/satisfaction is replaced by a finite number of axioms, one for each predicate of the language. Its theorems are the T-sentences. We actually adopted this format when presenting Tarski's theory.
- Davidson is well aware that in his definition of truth/satisfaction Tarski presupposes the notion of meaning (translation). But the latter is precisely the concept that Davidson wants to clarify. For this reason, truth and not meaning is taken as a primitive notion (see our remarks on Frege earlier). Tarski's theory is thus turned upside down:

  > An acceptable theory of truth must entail, for every sentence s of the object language, a sentence of the form: s is true if and only if p, where "p" is replaced by any sentence that is true if and only if s is. Given this formulation, the theory is tested by evidence that T-sentences are simply true; we have given up the idea that we must also tell whether what replaces "p" translates s. (Davidson 1984, 134)

- Unlike Tarski, Davidson wants his theory of meaning to apply to natural languages. He disagrees with Tarski that natural languages are too unregimented to be amenable to formal treatment. Davidson is ready to accept into the object language all kinds of expressions that Tarski had previously banned, such as proper names, demonstratives, indexicals, and so on. These are context-dependent expressions known to introduce complications into the T-schema. For this reason, Davidson takes the truth-predicate to apply, not to sentences, but to utterances restricted to a given speaker and moment of time:

  the sentence s as uttered by x at moment t is true if and only if p,

  of which a typical instance would be:

  "Socrates is flying" is true at time t
  if and only if Socrates flies at time t.

The architecture of Davidson's theory should be clear by now. What the interpreter knows when he is able to interpret a language (foreign or his own) is a Tarski-type theory of truth. The theory is empirically adequate when its empirical consequences, the T-sentences, pass the empirical tests. (Notice that the truth-predicate is now primitively understood.) If that is the case, then the interpreter knows for each sentence, its truth-conditions, and knowing those amounts to knowing the meaning of the sentence (Davidson 1984, 24).

Davidson's idea was that if a basically Tarskian truth-definition were carried out through all natural language and extensional truth-conditions were given for all sentences, that would even constitute or produce a theory of meaning. As we saw, here he appealed to strict logical arguments, which have caused considerable debate. But the greatest discussion about his program has been among nonlogical philosophers, because the Davidsonian interpretation of meaning would have far-reaching applications in many philosophical questions, and the feasibility of its assumptions became the subject for wide debate.

Davidson was not the only author in the 1960s and 1970s who sought to press more content out of the formal definition of correspondence. In his influential writings, Popper assumed that the semantic truth theory had already almost achieved a full grasp of factual correspondence. (See, e.g., Popper 1972, ch. 9.) On the contrary, Hartry Field argued that the Tarskian definition is a good start, but it should be supplemented with a theory of reference. According to Field, Tarski succeeded in reducing all truth to satisfaction for atomic sentences but at that stage his definition becomes noninformative or trivial. Therefore his truth theory does not have explanatory force; Field really thinks of explanation in the sense of natural sciences. What is needed is a theory of reference that would develop the semantic theory to a strong correspondence theory. (See Field 1972.)

12. Paradoxes

In his book *The Foundations of Mathematics* (1925) Ramsey introduced a distinction between semantic and set-theoretic paradoxes. Examples of the former include the Liar paradox, which will be discussed next, but also other paradoxes which involve the notions of meaning, definition, and so on. Here are a few of them:

- Grelling's paradox. Let us call an adjective *heterological* if it does not correctly apply to itself, and *autological* if it does. Thus *smart* is heterological while *English* is autological. It is natural to ask: Is *heterological* itself heterological? If the answer is yes, then it does not apply to itself; but then it is not heterological. Is it autological? If the answer is yes, then it applies to itself and thereby it is heterological.
- Berry's paradox. With each number we associate an expression. Consider the least number not specifiable in fewer than 20 syllables. But we just specified it in 19 syllables: "the least number not specifiable in fewer than 20 syllables."
- Richard's paradox. We associate with each decimal its definition with a finite number of words, if possible. Let the class of such decimals be E. Let us well-order E. Now we define a number N to be such that if the nth figure in the nth decimal in E is m, then the nth figure in N is $m+1$, or 0 if $m=9$. Then N is different from every member of E, and yet has been defined in a finite number of words.

The set-theoretical paradoxes involve set-theoretical notions such as set, ordinal number, membership, and so on.

- Russell's paradox. Let's group sets into those that are members of themselves and those that are not. Consider the set of sets that are not members of themselves. It is easy to see that it is a member of itself if and only if it is not.
- Burali-Forti's paradox. Take the series of all ordinal numbers. This series itself has an ordinal number O. The series of ordinal numbers up to and including any particular ordinal number exceeds that ordinal by one; thus, the series of ordinals up to and including O has the ordinal number $O+1$.

The reader is referred to Haack (1978) for an extensive discussion of the paradoxes. In this section we focus on the Liar paradox, and analyze some of the solutions given to it, with an eye on the theories of truth we have discussed so far.

12.1. Tarski's Solution

In the first section of his famous article, Tarski showed how the requirement of material adequacy for a theory of truth formulated as the derivability of all the relevant instances of his T-scheme leads to inconsistency in combination with self-referential sentences like the Liar.

Let us stipulate that c denotes the sentence:

$$c \text{ is false.}$$

The relevant T-biconditional now yields

$$\text{“}c \text{ is false” is true if and only if } c \text{ is false,}$$

which together with the previous stipulation $c =$ "c is false" entails

$$c \text{ is true if and only if } c \text{ is false.}$$

Combined with the principle of bivalence, the last statement leads to a contradiction,

$$c \text{ is true and } c \text{ is false.}$$

The general mechanism involved should be clear. On one side, the language contains expressions α ("c" in our example) which denote the sentence $\neg\, Tr(\alpha)$. In addition, the language has standard names a ("c is false" in our example). The relevant instance of Tarski's T-schema is

$$Tr(a) \leftrightarrow \neg\, Tr(\alpha),$$

which together with $a = \alpha$ and the principle of bivalence, implies a contradiction.

Tarski's solution to the paradoxes is to ban them altogether from the language: In his hierarchy of levels described earlier, the Liar-like sentences never arise. This solution has been found unsatisfactory for several reasons mentioned before. (See Soames 1999, especially chaps. 3 and 4 for a discussion of the problems.) In the sequel we shall discuss two other types of solutions to the Liar-type sentences: truth value gaps and contextualist approaches.

12.2. Truth Value Gaps

One of the the first serious attempts to break with the Tarskian approach was that of Kripke (1975) and Martin and Woodruff (1975). Recall Kripke's fixed point construction described earlier and consider the Liar sentence L, which says of itself that it is false (i.e., it is equivalent to $\neg\, Tr(\ulcorner L \urcorner)$). For it to find its place into the fixed point $\langle E^+, E^- \rangle$, it must appear in the extension or anti-extension of the truth-predicate at some of the stages S_i. It is obvious that L cannot appear at stage 0 since nothing does. But L cannot appear at stage 1 either, for if it appeared in the extension of the truth-predicate at that stage, $\neg\, Tr(\ulcorner L \urcorner)$ had to appear there, too, and that is possible only if $\neg\, Tr(\ulcorner L \urcorner)$ was true at stage 0. But we already pointed out that there is no sentence of the form $\neg\, Tr(\ulcorner L \urcorner)$ which is true at stage 0. By the same reasoning, it follows that L cannot appear in the antiextension of the truth-predicate at stage 1.

Similar arguments show that L cannot appear at any stage. Thus L is neither true nor false.

The problem with this solution is well known. If "L" is neither true nor false, then it is not true. But we cannot express this fact consistently in the object language. For suppose that "L" is not true. But then "$Tr(\ulcorner L \urcorner)$," which asserts that "$L$" is true, is false. Hence "$\neg\ Tr(\ulcorner L \urcorner)$" is true. But "$\neg\ Tr(\ulcorner L \urcorner)$" is the Liar, whence "$L$" is true. Thus from the premise that "L" is not true we ended up with the conclusion that "L" is true. Thus Kripke's Strong Kleene proposal according to which the Liar is neither truth nor false is inadequate, for it does not allow one to state this fact in one's own object language.

Another problem with the truth value gaps approach has been discussed in connection with partial interpretations: Given the definition of $\varphi \rightarrow \varphi$ by $\neg\varphi \vee \varphi$, not all the implications $\varphi \rightarrow \varphi$ are valid in the Strong Kleene semantics, and thereby not all T-biconditionals are valid either. The fact that $\neg\varphi \vee \varphi$ does not hold may not be philosophically so damaging, but the nonvalidity of $\varphi \rightarrow \varphi$ is disturbing. We know there is no easy way to fix this problem by introducing the "right" kind of conditional. (See Beall 2005, section 4.1.3.)

In a recent series of papers, Field attempts to build up a language that has a notion of implication that renders all T-conditionals valid, and also allows one to assert consistently that the Liars and their negations are not true. This last claim has to be qualified, however, in the sense that the Liars and their negations turn out not to be true in a stronger sense of truth distinct from Tarski's material truth. (See, e.g., Field 2003, and Beall 2005, section 4.2.) Attempts to block inconsistencies arising from the presence of the Liar sentences in the language by introducing a stronger notion of truth (definite truth) than Tarski's truth are also typical for the supervaluationist approach. McGee (1991) shows how Tarski's piece of reasoning leading to inconsistency (see above) cannot be carried out with his notion of definite truth.

12.3. Contextualist Solutions

We have selected two proposals, one originating with Parsons and developed by Burge and Glanzberg; the other, due to Barwise and Etchemendy is developed within the framework of *Situation Semantics*.

12.3.1. Quantifier Shift

One way to get around the problem discussed in connection with truth value gaps has been suggested by Charles Parsons (1983) and Tyler Burge (1979). Their idea is that the truth-predicate as it appears in the Liar applies to different entities from the truth-predicate which is used to classify the Liar sentences (i.e., to say of the Liar that is not true). There is one essential modification though with respect to Tarski's theory: It is propositions and not sentences which are the truth-bearers. Consider now the reformulation of the

Liar sentence in terms of propositions

(c) c expresses a false proposition.

To allow for the possibility of a sentence not expressing a proposition at all, Parsons replaces Tarski's T-schema by the weaker:

(1) $\forall x$(if x is a proposition and "p" expresses x, then x is true if and only if p).

Together with the assumption that propositions are bivalent, (1) entails

(2) $\forall x$(if x is a proposition and "p" expresses x, then x is false if and only if $\neg p$).

Applied to the two Liar sentences, (1) and (2) lead to the conclusion that they do not express a proposition at all. Here is the argument.

Suppose x is a proposition and c expresses x. Then by (1) we get

(3) x is true if and only if c expresses a false proposition.

Suppose x is not true. By existential generalization, we infer

(4) $\exists x(x \text{ is a proposition} \wedge \neg(x \text{ is true}) \wedge c \text{ expresses } x)$,

that is, c expresses a false proposition. But now, from (3) we get that x is true. Thus starting from an arbitrary proposition x that c expresses, we landed in the conclusion that x is true.

(5) $\forall x((x \text{ is a proposition} \wedge c \text{ expresses } x) \rightarrow x \text{ is true})$.

(5) is equivalent with

(6) $\neg\exists x(x \text{ is a proposition} \wedge \neg(x \text{ is true}) \wedge c \text{ expresses } x)$.

But then there is no proposition that c expresses, for if c expressed one, say y, then by (5) y would have to be a true proposition, and that together with (3) implies that c expresses a false proposition. But this is in contradiction with (6). A similar argument shows that the strengthened Liar (d is the sentence: d does not express a proposition) does not express a proposition either. Parsons avoids the contradiction by his shifting quantifier domain assumption. According to it, both (4) (c expresses a false proposition) and (6) (c does not express any proposition) are true, but the quantifiers range over different domains. The former is larger than the latter.

The quantifier-shift proposal has been found attractive for at least two reasons. As made clear in the recent development of the theory due to Glanzberg (2004), the shifting-domain account amounts to a contextual expansion of the background domain of truth-conditions, which gets bigger and bigger, an assumption that is common practice in linguistic theory.

The second advantage is the possibility it opens up for narrowing down the gap between set-theoretic and semantic paradoxes introduced by Ramsey. Here is Parsons's argument. Given a predicate "Fx," the fact that a is its extension is expressed by the condition

(7) $$\forall x(x \in a \leftrightarrow Fx).$$

By analogy with (1) we have

(8) $$\forall y(y \text{ is the extension of "}Fx\text{"} \rightarrow \forall x(x \in y \leftrightarrow Fx)).$$

If we now take "Fx" to be "$x \notin x$," we obtain

(9) $$\neg\exists y \forall x(x \in y \leftrightarrow x \notin x).$$

But (8) and (9) entail

(10) $$\neg\exists y(y \text{ is the extension of "}x \notin x\text{"}).$$

Parsons adopts here the same solution he proposed for the Liar, that is, he takes the two quantifiers in (8) and (10) to have different domain of discourse. (Parsons 1983, 231–232.)

For another "indexical" solution, the reader is referred to Keith Simmons (1987), who argues that the idea of treating "true" as an indexical term was anticipated by some mediaeval authors.

12.3.2. Situation Semantics

Situation semantics (*SS*) was originally conceived as an alternative to traditional truth-conditional semantics which emerged from Frege's work and found its technical expression in the works of Tarski and Davidson and the possible worlds semantics of Hintikka and Kripke. It has its starting point in Barwise's and Perry's work on perception reports. For a developed version, see Barwise and Perry (1983).

The proponents of *SS* set themselves the task to bring ontology back into semantics. Its basic elements are (a) situations and (b) the relation theory of meaning.

Situations are, roughly speaking, parcels of reality consisting of individuals who have various properties and stand in relations to each other at spatiotemporal locations. If one abstracts from the spatiotemporal location and considers only the individuals together with their properties and the relations holding between them, then one gets a *situation-type*. When a location is added to that, one obtains a situation.

The situation-type is technically represented by a partial function from n-ary relations and n individuals to the truth values 0 and 1. To take a standard example in the *SS* literature, the situation-type of Molly barking and Jackie not barking is identified with the partial function s,

$$s(\text{barks}, \text{Molly}) = 1$$
$$s(\text{barks}, \text{Jackie}) = 0,$$

that is sometimes represented in a more perspicuous notation by

$$\begin{aligned} \text{In } s\text{:} \quad & \langle\langle \text{barks}, \text{Molly}; 1 \rangle\rangle \\ & \langle\langle \text{barks}, \text{Jackie}; 0 \rangle\rangle. \end{aligned}$$

As pointed out earlier, situations (or courses of events) are obtained by adding the appropriate locations to situation-types and are technically identified with partial functions from locations to situation-types. In the technical format used above, the situation e that has Molly barking at location l and Jackie not barking at location l' is represented by

$$\begin{aligned} \text{In } e\text{:} \quad \text{at } l\text{:} \quad & \langle\langle \text{barks}, \text{Molly}; 1 \rangle\rangle \\ \text{at } l'\text{:} \quad & \langle\langle \text{barks}, \text{Jackie}; 0 \rangle\rangle. \end{aligned}$$

The *relational theory of meaning* construes the linguistic meaning of a sentence as a relation holding between two types of situations, *utterances*, and the *situations they describe*, that is, the contents of the utterances. Thus when the *sentence*

I am sitting

is uttered by A, it describes the situation that A is sitting, and when uttered by B it describes the situation that B is sitting. The situation described is typically identified with the *descriptive content* of the utterance.

The development of situation semantics in the eighties led to a broader notion of *propositional content* or *Austinian proposition* (Barwise and Etchemendy 1987). At the early eighties, the meaning of a sentence was a relation between circumstances (discourse situations, utterances) and (described) contents. But little by little, the circumstances of the utterance, that is, the discourse situation, become part of the broader notion of propositional content. The standard example from Barwise and Etchemendy (1987, 121–122) is a poker game in which Holmes utters, "Claire has the ace of hearts." Traditionally the circumstances s of the the utterance (the utterer's spatiotemporal position) are distinguished from the descriptive content of the utterance, that is, the state of affairs of Claire having the ace of hearts, represented by

$$\langle\langle \text{has; Claire, the ace of hearts, } 1 \rangle\rangle.$$

With the development of Austinian propositions, s itself become part of the larger notion of content represented by

$$s \models \langle\langle \text{has; Claire, the ace of hearts, } 1 \rangle\rangle,$$

which includes s as a constituent. The emergence of Austinian propositions allowed Barwise and Etchemendy to model the attitude reports arising from Quine's and Kripke's puzzles and to deal with strengthened Liar paradoxes like

(1) (1) is not true.

(2) (1) is not true.

The following reasoning takes place: (1) is true if and only if (1) is not true. Given the law of excluded middle (1) is not true, which is what (2) says, hence (2) is true.

In the new setting, the Austinian proposition corresponding to (1) is

$$s \models \langle\langle \text{true}; (1), 0 \rangle\rangle, \tag{+}$$

where s is the focus situation that includes the conditions of utterance of (1). For reasons that we cannot detail here, (+) is false because there are no situations of the appropriate type to classify s. But the Austinian proposition corresponding to (2) is

$$s' \models \langle\langle s \models \langle\langle \text{true}; (1), 0 \rangle\rangle \rangle\rangle, \tag{$*$}$$

where s' is a new situation that expands s and contains the fact of the falsity of (+). Because of this, ($*$) is true. The similarity with the previous approach is obvious. We achieve consistency by switching to a different situation.

13. Philosophical Readings of Semantic Theory

13.1. Against Correspondence

The view about truth that has recently caused perhaps most debate among formal philosophers of language is *deflationary*. Frege already suggested that no property of truth can be found (Frege 1918). For him, truth is not definable at all. As a conscious theory, deflationary truth-definition begins from Frank Ramsey, according to whom the predicate of truth is "redundant": "it is true that p" *means the same* as p. "True" and "false" "are phrases which we sometimes use for emphasis or for stylistic reasons, or to indicate the position occupied by the statement in our argument" (Ramsey 1927, 38). In this way, the truth-predicate and the distinction between language and metalanguage would be eliminated and also the notion of correspondence becomes obviously superfluous. Hence this approach is fascinatingly simple. However, it is not in any way obviously right, and there are numerous problems arising because the truth-predicate is used in many other contexts that do not allow such an elementary elimination—the most famous is "everything he said is true," and also generalizations like "Consequences of true sentences are true" ought to be explained.

Deflationary views were occasionally defended even before their explicit logical formulation. For example, Strawson argued at one stage that it is an error to see truth as a predicate, since calling a proposition true is just a signal of approval. Newer deflationism has a different policy: It admits that truth-sentences do make an assertion, but claims that this assertion is not really about truth (since there is no such property). Genuine discussion of these questions started in the 1970s, and the first painstaking proposal may have been by Williams in 1976. Essentially, he elaborated Ramsey's position to save the basic redundancy thesis (Williams 1976). Nowadays the deflationary strategy has divided to several branches which use different more or less radical methods of explaining problematic cases. Formulated in most cautious way, it leads to "minimalist" theories that do not want to eliminate the truth-predicate altogether but insist that an adequate analysis of truth is found in the necessary equivalence in all instances of "the proposition that p is true iff q." (Such a theory needs a different condition for each p and must hence be infinitely long; see Horwich 1990.)

A very different and philosophically interesting vision of semantics was developed by Michael Dummett in numerous works from 1959 on. Inspired both by later Wittgenstein and by intuitionism, he suggested that the meaning of a sentence or a proposition is to be seen, not in its truth conditions, but in conditions of justification. This led to lively discussion that is complicated among other things by the fact that the debated thesis has stronger and weaker forms. (See the succession of papers in Dummett 1978.) Dummett's thesis would mean that, when we speak about meaning in semantics, truth ought to be either *analyzed* or *replaced* with *proof* (in a mathematical sense or in a broad natural sense). The position is often called semantic antirealism. According to Dummett, knowledge of truth conditions has no real content besides knowledge of justification conditions. And some sentences are fundamentally undecidable. Therefore he also wants to give up the logical principle of bivalence. (See Wright 1987, ch. 10.) It is noteworthy that the antirealist view in philosophical semantics has great relevance to directly logical questions. Among other things, it advises us to understand logical results from a proof-theoretic perspective and to turn to use some grade of intuitionist systems. On the other hand, Dummett has debated with some logicians who have defended even more militant constructivism. (See, e.g., Sundholm 1994 and Dummett's answer; Sundholm supported the intuitionist proof-theory like that introduced by Martin-Löf.)

13.2. Possible Worlds

The relations of logic and ontology are highlighted in the theory of possible worlds. It has obviously a connection, first of all, to modalities and modal logic for which it was developed. However, the mere use of possible worlds apparatus—say, valuating sentences in a frame $\langle W, R\rangle$—does not yet carry any ontological weight. But if the members or indices in W are interpreted as *possible worlds*, then the full ontological problem will arise. Indeed, this seems to be inevitable if the formalism is wanted to be of some help in understanding modal discourse, in other words, if it is wanted to say something of what modalities and modal features are, and not only something about the interrelations of modal sentences. Moreover, the simple modal contexts are not the only field of application of possible worlds. Thus, they are often used to provide semantics for intensional entities, like propositions or concepts, or to model relations, like conditionals, between intensional entities.

It is obvious that the ontology of worlds touches the philosophy of logic. Thus, the old questions about, for example, the metaphysics of propositions reappear in a new format. In fact, many authors have wanted simply to identify propositions with classes of possible worlds even if this differs from the old senses of proposition. (This, however, leads to the classical difficulties about intensionally equivalent propositions: They would immediately get equated, though they differ in respect to propositional attitudes.) It would be essential to have a clear idea of what the worlds are. Unfortunately, the ontology of possible worlds is in a chaotic state because all inquiries use different ways of exposition. It is not easy even to find a decent classification of the alternatives.

Because of its extremeness, the clearest theory may be the ultra-realism developed by David Lewis, most fully in Lewis (1986). According to Lewis, worlds are real individuals, and possibility sentences ought to be understood as genuine existential statements quantifying over individuals. (Worlds are maximal spatiotemporally connected individuals, and each possible world is one of these existents.) Propositions are subsets of the set of worlds, properties are subsets of the set of individuals. That one world is actual has no logical importance; it is expressed simply by labeling one world with a special index. Thus Lewis thinks he can build full semantics by means of extreme realism, set-theory, and mereology.

Most philosophers have not accepted the Lewisian theory. The most widely supported theories have been "actualist"—actualist in the sense that possibilities are understood from the standpoint of what is actual. The task then is to explain clearly what the ontological status and constitution of the possible worlds is and what it means that something is true at a world. There has been great variety in the answers. According to one popular idea, advocated by Alvin Plantinga, the worlds are to been seen as maximal consistent states of affairs. A basically similar point of view is probably supported by those who think that we can quantify over "ways things could be." Certain philosophers assume that possible worlds are abstract entities, comparable to mathematical ones, while some say that they are conceptual constructions regarding actual things (Kripke). The "linguistic" interpretation, in its turn, would aim at reducing the possible worlds to basically logicolinguistic entities, to descriptions satisfying structural conditions (Hintikka). All the solutions are still very much in need of clarification, and the confused state is a hindrance to the philosophy of logic, especially modal logic. (See Forbes 1985; Divers 2002; Melia 2003.)

An extra problem arises because of the possible identity of individuals in different worlds. That is a crucial question for all modal logic with quantification *de re*, but opinions about its ontological basis still conflict. And how do individual names find their referents in different worlds? Lewis (like Leibniz) thought that one and the same individual could not exist in two worlds; the other worlds can have only "counterparts" of the individual. Most others think (like Scotus and Ockham thought, and as was common in the fourteenth century) that the *same* individual may exist in many worlds, but then it is urgent to have some idea of when the individual is the same. For example, how different may it be in other worlds? For some (Kripke), this just is a given fact demanding no explanation; for some, it is explained by the individual essence (*haecceitas*) that belongs indistinguishably to the individual and remains the same in various worlds. Essentialism is a feasible and classical position but nowadays not popular. Thus some authors have argued that no further definition of transworld identity is needed, whereas others say that the identification happens by means of basically nonlogical criteria of (sufficient) similarity in empirical features.

13.3. Analyticity

The ontology of worlds and propositions can easily lead us to what was presumably the most famous issue in the philosophy of logic during the century, that is, the controversy about *analyticity* and necessary truth. The concept of analyticity was important for the Vienna Circle because one of its main theses was that all propositions that can be known a priori are analytic. In these debates the concept was seen to be unclear, and attempts of explanation followed. One classical idea for that purpose was, and still is, to say that analytic truths are true "in virtue of meanings." Carnap (1947) tried to formalize this by replacing analyticity with *L-truth*. The *L*-true sentences hold in any state-description and are thus true because of the semantic rules of the language, as was remarked in section 7.2. But the doctrine of unique semantic rules was dubious. Moreover, the "paradox of analysis" was disturbing: How can analytic truths ever be nontrivial?

This was the background for Quine's attack, which made analyticity a subject of urgent debate. In the chapter about modal logic it was already noted how his criticism had far-reaching impact. His general pursuit was to get rid of all intensional entities and phrases in favor of a naturalist vocabulary; sets were the only nonphysical type of things he had to admit. The rejection of modal and attitudinal theories is hence a natural consequence. But his approach has even more profound implications in semantics. They got their classical expression in "Two Dogmas of Empiricism" (Quine 1951) and were later developed and somewhat modified in numerous works by Quine and his followers.

The criticism concerns the whole traditional division of truths into two strictly distinct classes. Philosophers used to think that some truths are analytic and some synthetic, some truths are necessary and some contingent. It had, of course, been an old subject of dispute how the divisions ought to be precisely delineated and if they coincided or not. But Quine argued that there simply is no such a dichotomy. According to him, the supposed status of analyticity would have to be explained by semantic rules, which are explained by synonymy, which in its turn is explained by the analyticity of substitutions. So, analyticity remains unexplained. The so-called analytic truths are nothing but unusually well-grounded truths. Later, Quine told that the dichotomy of analytic and synthetic ought to replaced by a scale: Some sentences are more deeply embedded in "the web of our beliefs" than others, but they all belong to a whole which could be amended at any point. According to this "confirmation holism," amendment of analytic or even logical truths would be expensive but possible in principle. (Quine, however, was long unwilling to allow any changes in the ordinary logic.)

Taken literally, the argument would be extremely important for logic: No truths would then be privileged *logical* truths, differing from others, and no inferences would be special logical inferences. Quine's criticism, however, has not been universally accepted. Already in the 1950s it was pointed out that our ordinary way of thinking is far from Quine (see Grice and Strawson 1956). Moreover, it is by no means clear that the circularity that the "Two Dogmas" discloses is in any way vicious. After the Chomskyan linguistics was created, it also became clearer what could be meant by linguistic rules and necessities. Even if it is admitted that it is often not clear which inferential connections contribute to meanings it does not follow that there is no difference between those that do and those that do not. The concepts "analytic" and "logical" are nowadays rather freely used again, though with greater care than before, but the battle is not over and it has become obvious that the whole idea of conceptual relations is in need of fundamental inquiry. (See, e.g., Peacocke 1992; Jackson 1998.)

13.4. Names and Presuppositions

The remarks so far have all been about complete sentences or propositions. Their components have received much less attention in philosophical semantics. One problem, however, has been widely debated, that of the references of individual names and name-like expressions. Frege and Russell found these references by replacing the names with suitable descriptive phrases which were sufficient for identifying the thing. Russell made this thought famous with his "On Denoting." (In fact it is questionable if they ever really thought that names were synonymous with the descriptions, though Russell sometimes hints at that.) Such a procedure was long taken for granted.

A new turn took place when Kripke argued that names must not be understood descriptively. Instead, they are demonstrative, given in the first place ostensively by some kind of baptizing that fixes the name on an object, and then transmitted to other users of language (Kripke 1972). They have no meaning: They only designate. It also follows that a name is a rigid designator which designates the same object in all possible worlds. This opinion, often called the "new theory of reference," very soon became dominant. However, the original arguments to support it turned out to be inconclusive. It is not clear that all names are rigid, and neither is it clear that this would imply a theory of purely causally delivered reference. The alternative descriptive view has not in any way been refuted if it is interpreted cleverly enough.

It is not possible here to dwell on any issues of modern linguistic semantics though some of them have great philosophical relevance. For example, the logical analysis of presupposition and implicature has given rise to plenty of literature. In fact, the whole linguistic inquiry of implicature started from H. P. Grice's philosophical papers. Implicature, perhaps, is a matter of pure pragmatics, but presupposition does concern genuine logical semantics: What other sentences must be true so that a certain statement can have a truth value at all? This question appeared in connection to Russell's definite descriptions, when Strawson argued that the use of an individual term does not assert the existence of the individual but presupposes it. Hence, Russell's favorite "The King of France is bald," would not be false, as the theory of definite descriptions declares, but neither true nor false. This is by no means the only case of presupposition: Many temporal expressions, factive verbs, counterfactuals, and so on, show the same feature. Linguists have studied these phenomena intensely since the 1970s. Logicians would be needed especially in telling how the presuppositions of simple sentences are connected to the presuppositions of complex sentences (see, e.g., van der Sandt 1988). However, precisely this seems to be an extremely obscure issue.

To sum up, one gets the general impression that in the first half of the twentieth century only few specialists were active with the philosophical implications of logical semantics. After "analytic philosophy" strengthened its position, and after the new basic results of logic became better known, semantics grew into an established philosophical discipline. Quine's provoking work must have been highly important here. A classical source of this stage is the influential treatise by Pap (1958). Next, the conception of semantics became wider, and the word was often used almost as a synonym for philosophical studies of language; according to one slogan, semantics was the meeting place of philosophy, psychology, and linguistics. (See the widely read anthology Steinberg and Jakobovits 1971.) From the 1980s on, the philosophical fashion has again changed, and even philosophical discussion about semantic theory has become more technical.

References

Austin, J. L. 1950. Truth. In *Philosophical Papers of J. L. Austin.* New York: Oxford University Press, 1961.

Auxier, Randall E., and Lewis Edwin Hahn, eds. 2006. *The Philosophy of Jaakko Hintikka.* Chicago: Open Court.

Bar-Hillel, Yehoshua. 1954. Indexical expressions. *Mind* 63: 359–379.

Barwise, Jon, and John Etchemendy. 1987. *The Liar: An Essay on Truth and Circularity.* Oxford: Oxford University Press.

Barwise, Jon, and John Perry. 1983. *Situations and Attitudes*. Cambridge, Mass.: MIT Press.

Beall, J. C. 2005. Logical and semantic paradoxes. In *Handbook of the Philosophy of Science*. Vol. 10: *Philosophy of Logic*, ed. Dale Jacquette. Amsterdam: North-Holland.

Burge, Tyler. 1979. Semantical paradox. *Journal of Philosophy* 76: 169–198. Reprinted in Martin (1984).

Burge, Tyler. 1986. Frege on truth. In *Frege Synthesized*, eds. Leila Haaparanta and Jaakko Hintikka, 97–154. Dordrecht: Reidel.

Carnap, Rudolf. 1934. *Logische Syntax der Sprache*. English translation: Carnap 1937. *The Logical Syntax of Language*. London: Kegan Paul.

Carnap, Rudolf. 1942. *Introduction to Semantics*. Cambridge, Mass.: Harvard University Press.

Carnap, Rudolf. 1947. *Meaning and Necessity: A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.

Chapuis, André, and Anil Gupta, eds. 2000. *Circularity, Definition and Truth*. New Delhi: Indian Council of Philosophical Research.

Church, Alonzo. 1943. Review of Carnap's *Introduction to Semantics*. *Philosophical Review* 52: 298–304.

Church, Alonzo. 1956. *Introduction to Mathematical Logic*. Princeton, N.J.: Princeton University Press.

Davidson, Donald. 1967. Truth and meaning. *Synthese* 17: 304–323. Reprinted in Davidson (1984).

Davidson, Donald. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

Davidson, Donald. 1990. The structure and content of truth. *Journal of Philosophy* 87: 279–328.

Divers, John. 2002. *Possible Worlds*. London: Routledge.

Dummett, Michael. 1978. *Truth and Other Enigmas*. London: Duckworth.

Feferman, Solomon. 1991. Reflecting on incompleteness. *Journal of Symbolic Logic* 56: 1–49.

Field, Hartry. 1972. Tarski's theory of truth. *Journal of Philosophy* 69: 347–375.

Field, Hartry. 2003. A revenge-immune solution to the semantic paradoxes. *Journal of Philosohical Logic* 32: 139–177.

Forbes, Graeme. 1985. *The Metaphysics of Modality*. Oxford: Oxford University Press.

Frege, Gottlob. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: Louis Nebert. English translation: Frege 1972. *Conceptual Notation*, translated by T. W. Bynum. Oxford: Clarendon Press.

Frege, Gottlob. 1884. *Die Grundlagen der Arithmetik*. English translation: Frege 1950. *The Foundations of Arithmetic*, translated by J. L. Austin. Oxford: Basil Blackwell.

Frege, Gottlob. 1891. Funktion und Begriff. In Frege (1967), 125–142. English translation: Function and concept. In Frege (1980).

Frege, Gottlob. 1892a. Über Sinn und Bedeutung. In Frege (1967), 143–162. English translation: On sense and meaning. In Frege (1980).

Frege, Gottlob. 1892b. Über Begriff und Gegenstand. In Frege (1967), 167–178. English translation: On concept and object. In Frege (1980).

Frege, Gottlob. 1893. *Grundgesetze der Arithmetik*, Band I. Jena: Verlag von H. Pohle. English translation: *The Basic Laws of Arithmetic*, translated by Montgomery Furth. Berkeley: University of California Press, 1964.

Frege, Gottlob. 1918. Der Gedanke: eine logische Untersuchung. In Frege (1967), 342–361. English translation: The thought: a logical inquiry. *Mind* 65: 289–311 (1956).

Frege, Gottlob. 1967. *Kleine Schriften.* Herausgegeben von Ignacio Angelelli. Hildesheim: Georg Olms.

Frege, Gottlob. 1980. *Translations from the Philosophical Writings of Gottlob Frege.* Edited by Peter Geach and Max Black. Third edition. Oxford: Basil Blackwell.

Glanzberg, Michael. 2004. A contextual-hierarchical approach to truth and the liar paradox. *Journal of Philosophical Logic* 33: 27–88.

Gödel, Kurt. 1944. Russell's mathematical work. Reprinted in *Gödel: Collected Works*, ed. Solomon Feferman et al. Volume II. Oxford: Oxford University Press, 1990.

Grice, H. P. and P. F. Strawson. 1956. In defence of a dogma. *Philosophical Review* 65: 141–158.

Gupta, Anil, and Nuel Belnap. 1993. *The Revision Theory of Truth.* Cambridge, Mass.: MIT Press.

Haack, Susan. 1978. *Philosophy of Logic.* Cambridge: Cambridge University Press.

Habermas, Jürgen. 1996. Richard Rorty's pragmatic turn. In Habermas, *On the Pragmatics of Communication.* Cambridge, MA: MIT Press.

Henkin, Leon. 1961. Some remarks on infinitely long formulas. In *Infinitistic Methods*, 167–183. Oxford: Pergamon Press, and Warsaw: PAN.

Hintikka, Jaakko. 1996. *The Principles of Mathematics Revisited.* Cambridge: Cambridge University Press.

Hintikka, Jaakko, and Gabriel Sandu. 1989. Informational independence as a semantical phenomenon. In *Logic, Methodology and Philosophy of Science VIII*, ed. Jens Erik Fenstad et al., 571–589. Amsterdam: North-Holland.

Hodges, Wilfrid. 1997. Compositional semantics for a language with imperfect information. *Journal of the IGPL* 5: 539–563.

Horwich, Paul. 1990. *Truth.* Second edition. Oxford: Oxford University Press.

Hyttinen, Tapani, and Gabriel Sandu. 2000. Henkin quantifiers and the definability of truth. *Journal of Philosophical Logic* 29: 507–527.

Jackson, Frank. 1998. *From Metaphysics to Ethics: A Defence of Conceptual Analysis.* Oxford: Oxford University Press.

Kirk, Robert. 1986. *Translation Determined.* Oxford: Oxford University Press.

Kirkham, Richard L. 1992. *Theories of Truth.* Cambridge, Mass.: MIT Press.

Kripke, Saul. 1972. *Naming and Necessity.* Cambridge, Mass.: Harvard University Press.

Kripke, Saul. 1975. Outline of a theory of truth. *Journal of Philosophy* 72: 690–716. Reprinted in Martin (1984), 53–82.

Lewis, David. 1986. *On the Plurality of Worlds.* Oxford: Blackwell.

Martin, Robert L., ed. 1984. *Recent Essays on Truth and the Liar Paradox.* Oxford: Oxford University Press.

Martin, Robert L., and Peter W. Woodruff. 1975. On representing "true-in-L" in L. *Philosophia* 5: 213–217. Reprinted in Martin (1984), 47–52.

McGee, Vann. 1991. *Truth, Vagueness, and Paradox.* Indianapolis: Hackett.

Melia, Joseph. 2003. *Modality.* London: Acumen.

Moore, G. E. 1899. The nature of judgment. In *G. E. Moore: Selected Writings*. London: Routledge, 1993.

Myhill, John. 1950. A system which can define its own truth. *Fundamenta Mathematicae* 37: 190–192.

Neale, Stephen. 1995. The philosophical significance of Gödel's slingshot. *Mind* 104: 761–825.

Neurath, Otto. 1933. Protokollsätze. *Erkenntnis* 3: 204–214.

Nuchelmans, Gabriel. 1989. Proposition. In *Historisches Wörterbuch der Philosophie*, Band 7, eds. J. Ritter and K. Gründer. Basel: Schwabe.

Oberdan, Thomas. 1990. Positivism and the pragmatic theory of observation. In *PSA 1990: Proceedings of the 1990 Biennial Meeting of the Philosophy of Science Association*, vol. 1, eds. Arthur Fine, Micky Forbes, and Linda Wessels, 25–37. East Lansing: Philosophy of Science Association.

Oberdan, Thomas. 1992. The concept of truth in Carnap's *Logical Syntax of Language*. *Synthese* 93: 239–260.

Olson, Kenneth Russell. 1987. *An Essay on Facts*. Lecture Notes 6. Chicago: CSLI Publications.

Pap, Arthur. 1958. *Semantics and Necessary Truth*. New Haven, Conn.: Yale University Press.

Parsons, Charles. 1983. *Mathematics in Philosophy: Selected Essays*. Ithaca, N.Y.: Cornell University Press.

Peacocke, Christopher. 1992. *A Study of Concepts*. Cambridge, Mass.: MIT Press.

Popper, Karl R. 1972. *Objective Knowledge*. Oxford: Oxford University Press.

Prior, A. N. 1961. On a family of paradoxes. *Notre Dame Journal of Formal Logic* 11: 16–32.

Quine, Willard Van Orman. 1951. Two dogmas of empiricism. In Quine (1953).

Quine, Willard Van Orman. 1953. *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press.

Quine, Willard Van Orman. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.

Ramsey, Frank Plumpton. 1927. Facts and propositions. Reprinted in *F. P. Ramsey: Philosophical Papers*, ed. D. H. Mellor, 34–51. Cambridge: Cambridge University Press, 1990.

Ramsey, Frank Plumpton. 1991. *On Truth*, ed. Nicholas Rescher and Ulrich Majer. Dordrecht: Kluwer.

Rescher, Nicholas. 1973. *The Coherence Theory of Truth*. Oxford: Oxford University Press.

Rojszczak, Artur. 2005. *From the Act of Judging to the Sentence: The Problem of Truth Bearers from Bolzano to Tarski*. Dordrecht: Kluwer.

Rouilhan, Philippe de, and Serge Bozon. 2006. The truth of IF: Has Hintikka really exorcized Tarski's curse? In Auxier and Hahn (2006).

Russell, Bertrand. 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press.

Russell, Bertrand. 1905. On denoting. *Mind* 14: 479–493. Reprinted in Russell (1956), 41–56.

Russell, Bertrand. 1912. *The Problems of Philosophy*. Oxford University Press.

Russell, Bertrand. 1918. *The Philosophy of Logical Atomism*. In *Monist* 28 and 29. Reprinted in Russell (1956), 178–281.

Russell, Bertrand. 1940. *An Inquiry into Meaning and Truth*. New York: Norton.

Russell, Bertrand. 1956. *Logic and Knowledge: Essays 1901–1950*. London: George Allen & Unwin.

Sandu, Gabriel. 1996. IF first-order logic, Kripke, and 3-valued logic. Appendix to Hintikka (1996), 254–270.

Sandu, Gabriel. 1998. IF-logic and truth-definition. *Journal of Philosophical Logic* 27: 143–164.

Simmons, Keith. 1987. On a medieval solution to the liar paradox. *History and Philosophy of Logic* 8: 121–140.

Smorynski, C. 1977. The incompleteness theorems. In *Handbook of Mathematical Logic*, ed. Jon Barwise, 821–865. Amsterdam: North-Holland.

Soames, Scott. 1999. *Understanding Truth*. Oxford: Oxford University Press.

Stainton, Robert. Metaphysics, substitution salva veritate and the slingshot argument. Forthcoming in *Encyclopedia of Language and Linguistics*, second edition.

Steinberg, Danny D., and Leon A. Jakobovits, eds. 1971. *Semantics*. Cambridge: Cambridge University Press.

Sundholm, Göran. 1994. Vestiges of realism. In *The Philosophy of Michael Dummett*, eds. B. McGuinness and G. Oliveri, 137–165. Dordrecht: Kluwer.

Tarski, Alfred. 1935. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica Commentarii Societatis philosophicae Polonorum* 1: 261–405. English translation: The concept of truth in formalized languages. In Tarski (1956), 152–278.

Tarski, Alfred. 1944. The semantic conception of truth. *Philosophy and Phenomenological Research* 4: 341–376. Reprinted in *Readings in Philosophical Analysis*, eds. Herbert Feigl and Wilfrid Sellars, 52–84. New York: Appleton-Century-Crofts, 1949.

Tarski, Alfred. 1956. *Logic, Semantics, Metamathematics*. London: Clarendon Press.

Van der Sandt, Rob A. 1988. *Context and Presupposition*. London: Croom Helm.

Williams, C. J. F. 1976. *What Is Truth?* Cambridge: Cambridge University Press.

Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul.

Wittgenstein, Ludwig. 1978. *Remarks on the Foundations of Mathematics*. Edited by G. H. von Wright, R. Rhees, and G. E. M. Anscombe. Oxford: Basil Blackwell.

Wittgenstein, Ludwig. 1980. *Culture and Value*. Oxford: Basil Blackwell.

Wright, Crispin. 1987. *Realism, Meaning and Truth*. Oxford: Blackwell.

14

The Philosophy of Alternative Logics

Andrew Aberdein and Stephen Read

1. What Are Alternative Logics?

> Several logicians have in the last fifty years been trying to find some simpler and better mode of ascertaining when arguments are good, but they have not yet agreed upon the subject. Until they do agree upon something better, we shall do well to learn the old rules, which are certainly both ingenious and useful. (Jevons 1876, 56f.)

So wrote Stanley Jevons shortly before Gottlob Frege's *Begriffsschrift* laid the foundations for an agreement among the majority of logicians that was to last well into the twentieth century. The focus of this agreement is the truth-functional propositional calculus, sometimes augmented by first-order quantifiers. This has become known as classical logic, or **K**.[1] Although **K** began as a purely mathematical formalism, it rapidly came to be applied to the assessment of natural argumentation, eventually achieving a near hegemony in this role. There have always been dissidents to disturb this appearance of unanimity, but in recent decades they have become especially conspicuous. Jevons's appraisal of the state of traditional logic a century and a quarter ago might as readily be applied to classical logic today.

Once the employment of an amended logic has been recognized as a legitimate response to a philosophical or scientific problem, two strategies are available. The choice is whether to introduce novel material specific to the problem while leaving the existing logical system intact, that is, to produce a conservative extension; or to amend what is already there, that is, to undertake a revision. The two strategies are essentially distinct; our concern in this chapter is with the latter.

1.1. What Is a Logical Theory?

Before we can explain how logics change we must clarify what we mean by "a logic." Most logics can be presented in many different ways: natural deduction presentations, sequent calculi, various axiom systems, and so forth. We may distinguish three basic types of presentation: logistic systems, which codify logical truths; consequence systems, which codify valid arguments; and deductive systems, which codify proofs (Corcoran 1969, 154ff.).[2] Our concern is with substantive divergence among logical systems intended for the formalization of rational argumentation. Although logistic systems may be adequate for some purposes, such as codifying the truths of arithmetic, they are too coarse-grained to capture all the differences with which we are concerned.[3] Conversely, deductive systems offer too fine-grained a classification: Differences that occur only at this level are outside the scope of our inquiry. Therefore our attention may be safely restricted to consequence systems.

However, the comparison of formal presentations of consequence systems is not enough to explain how such systems of logic develop. We must go beyond this, to provide a characterization of how they are deployed. The motivations for logical endeavor are many and various, but one difference among them is especially important here. On the one hand, research in logic can be pursued to improve understanding of reasoning in natural language (or some technical or scientific enrichment thereof): natural argumentation. On the other hand, logic can be a purely formal enterprise, manipulating symbols in accordance with explicit rules. We might characterize this as a distinction between "rough" and "smooth" logic (Goldstein 1992, 96).[4] We can readily identify clear examples of each: Purely formal results and applications to mathematics or computer science are obviously smooth; work on inductive logic or practical reason obviously rough. However, there is a continuum of work between these two extremes: Most interesting logical research has both smooth and rough aspects. When applied to whole systems of logic, the distinction should pick out those systems that could be advocated as improving our understanding of natural argumentation. It is **K**'s claim to be successful as a rough logic that is disputed by reformers; its success as a smooth logic is not in doubt, but then neither is that of many systems which could never be mistaken for rough logics.[5] No system of logic is maximally rough—the fit with natural argumentation can, indeed should, never be perfect—and it is an open question whether improvement on **K** is achievable.

We shall define a "logical theory" as the context in which a system of rough logic is deployed. Logical theories model the arguments of natural argumentation so as to explicate their rationality, in the same way that theories of natural science model phenomena in the natural world. Of course, to model the world a scientific theory needs not only a formal system but also a schema for identifying the features of this system with features of the world. Inevitably such translation schemata simplify and distort the world, hence some defense must be available to justify the special importance of the features focused on. Similarly, logical theories must also offer a schema for parsing the sentences of natural argumentation into propositions of the logical system. Moreover, the distortions of this parsing theory will require some theoretical defense, which the logical theory must also provide.

Thus a logical theory must contain more than just the underlying formal system. Michael Resnik (1996, 491; 1985, 225) characterizes a logical theory as "a quadruple consisting of a formal system, a semantics for it, the attendant metatheory, and a translation method for formalizing informal arguments." Logical theories can diverge by the revision of any of these four components. Changes to the first two elements are the principal subject matter of this chapter. Developments in the metatheory of a logical system are tangential to our concerns: Although congenial metatheoretic features, such as interpolation or the subformula property, have been proposed as reasons for preferring one system over another, such preference is generally too fine-grained to be considered here. For example, metatheoretic concerns may motivate the choice of one system of relevance logic over another, but not the choice of relevance over classical logic. Choices of the latter kind typically turn on the effectiveness with which competing logical theories meet a common purpose: representing natural argumentation. (We shall see in section 2.1.3 that this sort of comparison has been attempted where the purposes of the systems under comparison are incompatible.) The fourth component, the parsing theory, provides this representation.

That two logical theories may diverge by the revision of the parsing theory raises a number of special problems. The parsing theory plays a similar role in logic to that of observation theory in empirical science, and it inherits some of the same difficulties. For a scientific theory to offer an explanation of an observed phenomenon, the observation must be rendered in terms of the theory. This process is accomplished by the observation theory. It can profoundly affect the explanations or predictions offered by the scientific theory as a whole, and is itself conditioned by that theory.[6] For example, two biologists observing the same slide, but in the grip of diverging observation theories, may focus their attention on very different features, and thereby record very different observations. Even if their theories were otherwise in agreement, this difference of observation would lead them to differ sharply in their assessment of the slide. Thus two ostensibly similar theories may differ in their predictions solely on the basis of a difference at the observation level. Conversely, two fundamentally different scientific theories may coincide in predictions if their observation theories are constructed so that the differences are canceled out.

This confusion of scientific theory and observation theory may make the rational reconstruction of such theories more fugitive, but it raises few conceptual difficulties. For logical theories the situation is more confused. Whereas scientific observation theories are typically uncontroversial by comparison with the associated scientific theories, parsing theories have been understood as susceptible to more robust criticism. Hence advocacy of revision of the parsing theory over revision of the rest of the theory appears more methodologically respectable than in the scientific case. Why should this impression obtain? Formal systems of logic have historically been understood as much more normative than scientific theories. Thus, whereas an elegant and enduring scientific theory that required an elaborate and poorly motivated observation theory to cope with recalcitrant observations would be seen as standing in need of reform, the complicated parsing theory necessary to prop up some theoretically attractive logical system would be more readily tolerated. Reinforcing this point is a tacit presumption that logic is irrevisable. Once it has been accepted that logical systems can be revised, there is much less call for elaborate parsing theories.

However, this does not answer the fear that, because the parsing theory is unconstrained by the logical theory, it may always be stretched to accommodate the shortcomings of the formal system. In scientific development, it is an important methodological goal that observation theories should be as "transparent" as possible and that any substantive content within them should ultimately be incorporated into the theory proper. In logic the notion of a "transparent" parsing theory raises special difficulties, which might threaten this goal. One scientific observation theory is more transparent than another if less processing of the raw data is required. In logic the raw data are the utterances that make up natural language argumentation. Hence for any specific logical theory the most apparently transparent parsing theory is that which maximizes the preservation of the surface form of such utterances. But there is more to argumentation than surface form. However transparent the parsing theory, there must be some scope for latitude in the parsing of an utterance, because natural language, even in technical contexts, is inexact, elliptical, allusive, and also more expressive than any formal system. Moreover, the parsing theory is responsible not only for associating formal propositions with informal inferences; it must also assemble them into patterns of argument.[7]

The key question is how constrained a latitude should be afforded to the parsing theory. We have already seen that excessive latitude can license the retention of ad hoc logical theories. But the opposite pole, a perfectly transparent parsing theory capable of precisely capturing what is meant by any locution, must be an unattainable ideal. In particular, it would be unacceptable to Quineans, in so far as it depends on determinacy of translation, underpinned by realism about meanings (Resnik 1985, 229n5). The Quinean response is to understand formalization in terms of a cooperative feedback procedure, whereby prospective parsings are offered to the informal arguer for his approval. Eventually agreement will be achieved, or, if the arguer is sufficiently eccentric in what he is prepared to accept as representing his words, he will simply forfeit his inclusion in the discourse. Alternatively, we might observe that however sophisticated a logical system may be, it is inevitably, indeed deliberately, far less expressive than any natural language. Hence the parsing process is necessarily procrustean, and the scope for the divergence of translation that motivates Quine's indeterminacy thesis is limited. Moderate transparency of translation appears reasonable, at least as a regulative ideal.

A broader characterization of the content of a logical theory is offered by Paul Thagard. What he calls an "inferential system" is defined as "a matrix of four elements: normative principles, descriptions of inferential practice, inferential goals, background psychological and philosophical theories" (Thagard 1982, 37). The first two of these elements are present in Resnik's analysis: The syntax, semantics, and metatheory of a logical theory constitute its normative principles, and the parsing theory is a means by which descriptions of inferential practice may be given in terms of those principles. The second two elements introduce grounds for divergence between logics which we have not yet addressed. The inferential goals prescribe what the inferential practice is intended to achieve and what the valid inferences are expected to preserve. The preservation of truth and avoidance of falsehood are the most familiar examples and common to most deductive logics. Some systems qualify these goals further: for example, by requiring constraints of relevance on the preservation of truth, as many of its protagonists describe it—though problematically so (see Read 2003). Other systems differ more substantially: Paraconsistent logic is concerned to avoid triviality rather than falsehood; intuitionistic logic is motivated by the preservation of warrant, rather than truth *simpliciter* (to characterize the distinction from the classicist's point of view); for inductive logics the preservation of truth is no longer the highest goal.[8] Resnik (1985, 235) finds Thagard's concern that logic should aim at "furthering human inferential goals" unduly psychologistic. Although Thagard's conception of logic *is* psychologistic, and his presentation of this material may betray as much, an understanding of the goals which a logical program is intended to pursue is crucial to the assessment of the status of such a program. Resnik is right to observe that the historical motivation for logical development has been theoretical not practical, but disagreement about how theoretical goals should be pursued is the key to some disputes between protagonists of different formal systems.

Thagard's second novelty is his contention that logical theories are constrained by psychological and philosophical theories. In respect of philosophical theories this seems uncontroversial: For example, Michael Dummett's advocacy of intuitionistic logic is grounded in his adoption of an antirealist theory of meaning (see 2.1.1). As we have already observed, Thagard also wishes to defend psychologism about logic. Specifically he sees human cognitive limitations as imposing constraints on logic. If, as he suggests, logics should contain no principles that humans are cognitively incapable of satisfying, then their development must be informed by psychological theories of human cognitive capability. Such psychologism has been widely criticized; four brief points will suffice here. If the purpose of logic were purely the description of inferential practice, it would be under the same constraints as that practice. But logic works by modeling intuitions whose normativity transcends actual practice. Second, the principles that result never impose obligations to perform humanly impossible tasks, despite Thagard's concern; rather, they are hypothetical imperatives, concerning what should be done to ensure the validity of inferences, should these inferences be carried out (see Resnik 1985, 236.) Third, we would then need to build in finitistic constraints on length of wff, length of proof, size of countermodel, and so on. Finally, and decisively, Thagard's psychologism is itself a background philosophical theory, hence this whole issue can be subsumed under the requirement that philosophical theories are relevant to the assessment of logical theories. In general, because our concern is with the methodology of logical development, rather than its ontology, we should aim for as much neutrality as possible with respect to competing accounts of the nature of logic, such as Thagard's psychologism. However, where justified, such accounts can be included among the background theories.

To take stock, a logical theory is the means by which a formal system may be promoted. It comprises the system itself, appropriate semantics and metatheory, a parsing theory, and an inferential goal. Taken together, we may call these components the "foreground" of the theory, since we would expect the logical theory also to contain background theories providing philosophical motivation.

1.2. Revolutions in Logic

In distinguishing between revisionary and nonrevisionary changes to logic, our underlying concern is an instance of a much more general problem. We are attempting to articulate a difference between talking about old things in a new way and talking about new things (whether in an old or a new way). That is to say, between advancing a new theory which is intended to cover the same ground as its predecessor, and seeking to analyze a new item, either by adaptation of the existing theory or by the introduction of a replacement theory. The first move is necessarily revisionary, the second is not. Before proceeding further it would be useful to have a clearer account of this difference between revisionary and nonrevisionary theory change.

Attempts at such an account have been made in some of the literature discussing the nature of scientific "revolutions." The earliest accounts of revolutions in science presumed that such change always marked a radical discontinuity, in which key concepts of the old theory were abandoned. Subsequent commentators (for instance, Crowe 1967, 123f.; Gillies 1992, 5) have argued that although revolutions of this character do occur, there can also be revolutionary change in which all the concepts are retained, albeit with a transformed character. This distinction is a familiar one in political history, where the revolutionary metaphor originates. We may distinguish between the Russian Revolution of 1917, in which the whole constitution was abandoned and replaced by something radically different, with different constituent parts, and the Glorious Revolution of 1688 in Britain, in which all the principal constitutional constituents, the Crown, Parliament, and so forth, were retained, although their character and relative significance changed dramatically.

Hence we may distinguish four relevant situations. A *glorious* revolution occurs when the key components of a theory are preserved, despite changes in their character and relative significance. (We will refer to such preservation, constitutive of a glorious revolution, as *glory*.) An *inglorious* revolution occurs when some key component(s) are lost, and perhaps other novel material is introduced by way of replacement.[9] A *paraglorious* revolution occurs when all the key components are preserved, as in a glorious revolution, but new key components are also added. The recent addition of a parliament to the constitution of Scotland is a political example of a paraglorious revolution. Finally, a theory is in *stasis* (a *null* revolution, as it were) when none of its key components change at all. Static theories may nonetheless undergo quite substantial change, notably in conservative expansion by new nonkey components. Hence stasis has something of the character of Kuhn's "normal science," and by distinguishing it from revolutionary change we might be thought to be reopening the dispute over the distinction between normal and revolutionary science.[10] However, there is little more than rhetorical weight in our use of the term "revolution" to describe these conceptual shifts, and we assume with the later Kuhn that their structure is similar at the microscopic and macroscopic levels (Kuhn 1977, 462). Provided that changes of radically different scales are not directly compared, the classification should be independent of this debate.

However, the classification of revolutions raises several further issues. First, we have not yet made clear how "key" and "preserved" are to be understood. Theories in empirical science are open to markedly divergent rational reconstructions, thereby generating controversy as to which components are genuinely "key." In logic this sort of dispute is much narrower and more readily resolved. Although there are many different systems of presentation for logic, there is comparatively little disagreement about which concepts these systems should respect. For present purposes, the key components of a logical system are its logical constants and its consequence relation. However, the definition of "preservation" is still troublesome. Various different accounts have been proposed for the empirical sciences (e.g., Fine 1967). The want of a suitable account for logical concepts lies behind the recurring debates over whether a new logic is "still" logic, to which we shall return in section 2.

Second, we should note that glory need not be transitive: A sequence of glorious revolutions may amount to an inglorious revolution. This could happen if the relative significance of the key components changes sufficiently for some components to cease to be key, or if preservation is itself nontransitive. However, this is less likely in the logical than the empirical case, since the range of possible key components is more narrowly constrained. Of course, inglorious revolutions can cancel each other out, so that characterization is straightforwardly nontransitive.[11]

Third, how is this classification related to the distinction between replacement of a theory by a successor and replacement by a competitor? There is a conceptual difference between this distinction and our classification of revolutions, since it is historical rather than methodological in character. Moreover, the difference between successors and competitors is imprecise; indeed, if the terms are understood with sufficient latitude, any successor may be seen as a competitor, since its advocacy is in competition to die-hard defense of the old theory, and vice versa, since a successful competitor succeeds the old theory.[12] However, it has been claimed that we can identify glorious and paraglorious revolutions with successors and inglorious revolutions with competitors (Crowe 1992, 310).[13]

Fourth, we need to know how this classification of revolutions is related to the contrast between conservative and nonconservative revision of the formal system. The adoption of a logic that is a conservative expansion of the antecedent system (an extended logic) can only represent a revolution if the new material is of key significance. Hence, if the new constants of an extended logic formalize hitherto extralogical (and thereby nonkey) material, its adoption will be nonrevolutionary; but if they formalize material hitherto formalized by the existing constants, the new system will be paragloriously revolutionary. Note that the question of what a constant formalizes, and thereby the precise delimitation of paraglorious from static extensions is settled by the parsing theory, not by the formal system alone. For example, the modal system **S4** would be expected to be a static extension of **K** and the relevance system $\mathbf{R}^{\neg}$ a paraglorious extension of **K**, because the new vocabulary of **S4** usually formalizes the hitherto ignored issue of modality, whereas the new intensional constants of $\mathbf{R}^{\neg}$ usually formalize much material hitherto addressed by the existing constants.[14] Yet sufficiently nonstandard parsing theories could overturn these preconceptions.

The consequence relation is always at least apparently preserved because all logical systems have a conception of consequence. Yet the characterization of consequence could undergo inglorious revolution. It might seem that in contrast to the constants, any change of consequence relation must be glorious, since the new relation will still be a consequence relation. However, that is to forget how weak a descriptor "consequence relation" is; what makes a relation a consequence relation is just its function within a logic. Hence it works like "head of state" rather than "king" or "president"; we would not call the replacement of a monarchy with a republic glorious just because both systems included a head of state.

Most commentators have argued that inglorious revolutions are impossible in mathematics.[15] Because logic and mathematics are prima facie similar endeavors there would appear to be a tension here, but it can be resolved. The ground for denying that inglorious revolutions occur in mathematics is that the discipline is cumulative in a way that empirical science is not: Both disciplines discard old material, but mathematicians never really throw it away. Quaternions or conic sections may be of no greater interest to the modern mathematician than phlogiston or caloric are to the modern physicist, but their legitimacy is not disputed. However, the mistake here is to focus on the whole discipline: Within the context of individual research programs all of this material has been just as decisively rejected. With rough logic this is much clearer: Our concern is with a specific range of research programs concerned with the formalization of natural argumentation, which are situated within a vast hinterland of smooth logic results.[16] Much of the material in the hinterland has been discarded from such programs as insufficiently rough; it still has a place as smooth logic, but has lost its prime application. In this fashion inglorious revolutions are possible within a cumulative discipline.[17]

Finally, there are epistemological difficulties in establishing the character of a revolution, since the preservation of terminology is, in itself, clearly neither necessary nor sufficient for the preservation of the underlying concepts: All may not be as it seems (see Gray 1992, 227). Hence there are 16, rather than 4, possible situations:

	S	G	P	I
S	SS	GS	PS	IS
G	SG	GG	PG	IG
P	SP	GP	PP	IP
I	SI	GI	PI	II

(In this table S, G, P, and I refer to the original four situations; the horizontal axis indicates reality and the vertical axis appearance. Hence the ordered pairs are really as indicated by the first letter, but appear to be as indicated by the second. Reality and appearance coincide on the diagonal, hence these situations are how the original four situations were initially understood.) Much of the problem here is that where there is genuine confusion or disagreement about the status of a revolution, we tend to use the same term before and after the revolution, either to describe something that endures through the revolution, or to (mis)describe two distinct but similar things. Hence the dispute becomes one of how (and whether) the meaning of that term has changed.

1.3. A Methodology of Logical Research Programs

With a characterization of the content of logical theories in place, we now turn to their dynamics, which we will approach by an appeal to the parallel treatment of theory change in the philosophy of science. Imre Lakatos's "On the Methodology of Scientific Research Programmes," or MSRP, is an attractive candidate for the treatment of theory change in logic because much of it is particularly applicable to formal contexts. Lakatos inherited from Popper an account of objectivity in terms of the process of discovery, rather than the objects discovered; something of considerable utility in the formal (and social) sciences, in which the former is much more readily accessible than the latter.[18]

Instead of taking individual theories in isolation, MSRP appraises series of theories, distinguishing between progressive and degenerating series. A series of theories, or *research program*, is said to be *theoretically progressive* if each theory has greater *empirical content* than its predecessor—that is, if it makes novel predictions (Lakatos 1970, 33). It is said to be *empirically progressive* if some of the excess content is corroborated—if some of the predictions come true (ibid., 34). Research programs are *progressive* if both theoretically and empirically progressive, and *degenerating* otherwise (ibid.).

What is the logical analog of "corroborated excess empirical content," the hallmark of a progressive shift of theory within a research program? The force of "empirical" here is to exclude both nonfalsifiable, "metaphysical" propositions and paraphrases, and strict corollaries of existing content, focusing instead on the production of new facts (ibid., 35). In his application of MSRP to mathematics, Hallett here employs a remark of Hilbert's, that "the final test of every new mathematical theory is its success in answering pre-existent questions that the theory was not designed to answer," to make non–ad hoc problem solving the hallmark of progress (Hallett 1979, 6; Hilbert 1926, 200). If anything, it is easier to describe a logical analog for empirical content than a mathematical one, since, unlike mathematics, (rough) logic always has an application. Hence the empirical content of a logical theory is its formalization of inference patterns in natural argumentation (where the intuitive validity of these is sufficiently well entrenched to resist being overturned in favor of a simpler calculus). When a new theory offers a plausible formalization of patterns of inference hitherto ignored, or judged ill-formed, or unconvincingly paraphrased, it exhibits excess empirical content.

A research program endures through the sequence of theories of which it is composed as a continuous programmatic component. This consists of two sets of methodological rules: the *negative heuristic*, which counsels against certain lines of inquiry, and the *positive heuristic*, which advocates others (Lakatos 1970, 48 ff.). The chief task of the negative heuristic is to defend the *hard core* of the program, that is, those propositions fundamental to its character (ibid., 48). The hard core contains the key features of a theory which must be retained in any revision if the successor theory is to belong to the same program. Hence a revolutionary change of theory will be glorious iff the hard core is unchanged, paraglorious iff the hard core is monotonically (and conservatively) increased, and inglorious iff the hard core is contracted or revised. The negative heuristic protects the hard core by ensuring that inferences from contrary evidence are directed not at the hard core but at a *protective belt* of auxiliary hypotheses: initial conditions, observational assumptions, and the like (ibid.). The research program is deemed successful if these moves can be achieved progressively; unsuccessful if they involve degeneration. This assessment of success works to rationalize the conventionalist strategy of preserving some propositions from criticism. We are justified in doing so if the program thereby exhibits progress, but if we can only do so at the expense of degeneration we may be obliged to revise or abandon our hard core.

The other characteristic feature of a research program is its positive heuristic. This consists in aspirational metaphysical generalizations which inform amendments to the negotiable elements of the program, that is, the protective belt (ibid., 51). A research program without a positive heuristic would warrant the methodological anarchy recommended by the later Feyerabend (1975).[19] This "anything goes" strategy would ensure that, at least conceptually, no stone went unturned, but for practical ends we might hope for a means to target our resources more effectively. One particular strength of the positive heuristic is that it permits practitioners to postpone consideration of apparent refutations of a progressive program. Providing that progress is being made, the positive heuristic will make a more pressing call on researchers' time than any anomalies. Thus anomalies only command attention when the program is in infancy or degeneration. A good illustration of this is provided by the considerable success of the classical logic research program in the first half of the twentieth century, which was not significantly impeded by known anomalies such as the paradoxes of self-reference and of material implication (Priest 1989a, 134f.).

An issue that is especially pertinent to the rational reconstruction of the development of logic is what one might call the nesting of one research program within another. For logic not only develops within its own research programs, it is also assumed in the development of many other programs in other disciplines. We require a more detailed account of scientific development, distinguishing between the different scopes, or depths of focus, that a research program may have.[20] A research group working on the synthesis of alkaloid compounds may take a prevailing theory of organic chemistry for granted, thereby including it in the hard core of their program: They would not be interested in methods that presume a general revision of organic chemistry. However, they would also subscribe—albeit more loosely—to some general research program of the whole discipline of organic chemistry. If there are theoretical organic chemists within that program who entertain the prospect of more wholesale revision, the hard core of the program will be much smaller.

Two features of this picture are immediately striking. First, the hard core of the general program will be a proper subset of the hard core of the specialized program. Second, different attitudes may be taken toward the content of a program's hard core. On the official attitude it contains only material of which the program's adherents are completely certain. This should tend to limit the size of the hard core and permit wide-ranging speculation as to the direction of future research. For practical purposes, so that a program may be kept within manageable bounds, it is convenient to augment the hard core of a research program by additional, conventional assumptions. This strategy is permissible within the more specialized programs of subdisciplines and specific projects, but methodologically vicious if adopted with respect to the discipline as a whole, since it would rule out potentially progressive revision. Within specialized programs individual researchers may harmlessly differ over which aspects of the hard core are conventional.

We may conceive of a whole array of depths of research program partially ordered by set-theoretic inclusion on the contents of their hard cores.[21] The theoretical end points of this array would be an empty hard core and a complete hard core. The latter would represent an irrevisable finished science. As an official view, this would have attained a state presumably unattainable by mere mortals;[22] as a conventional view, it would represent the cessation of scientific curiosity. On a realist account of science this end point (as an official view) must be unique. A research program with an empty hard core would represent the conceptual starting point for science suggested by Cartesian skepticism. More practical research programs are situated between these extremes. Programmes with very small hard cores containing only the most general principles would resemble Foucauldian epistemes (Foucault 1966, xxii, discussed in Gutting 1989, 140ff.). As he suggests, such programs would have a very wide disciplinary range, and would permit extensive revision within the more specific programs developed under their aegis. The content of the hard core of an episteme would be contained within the hard core of all contemporaneous research programs, making it hard to characterize and especially hard to revise. Close to the other extreme are research programs concerned with fine-tuning a theory or developing a specific application. Here most of the content of the theory would be contained in the hard core, although much of this would be assumed by convention.

The array imposes a partial ordering, rather than a total ordering, on research programs, thereby accommodating incompatible programs at the same stage of development. For any given program in the array we can identify a cone of programs with hard cores that properly include the hard core of the initial program. Where the initial program has the right degree of generality, we shall call this cone a *research tradition.*[23] We may now further refine the account of revolutions: Glorious revolutions conserve both program and tradition; paraglorious revolutions initiate successive programs within a tradition; and inglorious revolutions either initiate a competing program within the same tradition if the hard core of the initial program is conserved, or initiate a competing tradition otherwise.

The overall development of logic is too broad to be assimilated into a single coherent tradition. For example, any starting point from which we could develop both Brouwerian intuitionism, in which certain principles of mathematical intuition are conceptually prior to logic, and classical logicism, in which classical logic is conceptually prior to all of mathematics, would have a hard core little larger than that of the prevailing episteme. Although it is important to acknowledge the assumptions that the two programs share, there is insufficient community of content for the cone of programs containing them both to be a research tradition.

Within a given logical research tradition we shall be concerned with research programs at several different depths, which may be outlined as follows. First, there is the initial program, which characterizes the whole tradition, since its hard core is contained within that of all programs within the tradition. We would expect the hard core of this program to contain an incomplete articulation of each of the four components of a logical theory. Thus it would contain:

1. some components of the formal system: certain very general details of the composition of logical systems, "basic principles of reason," if there are assumed to be any, and perhaps ultimate analyses of the constants;

2. some constraints on the methodology of the parsing theory, such as a characterization of transparency, although the natural place for the theory proper will always be the protective belt;
3. a reasonably precise, but refinable, inferential goal; and
4. some general background theories: very general methodological principles and deep-seated philosophical theses.

At this stage, the content of the protective belt may still be fairly confused. If the program is progressive, successive revisions will yield a more completely articulated logical theory. Much of this theory may then be placed in the hard core by convention, to facilitate fine-tuning the theory. When this has been attained, the whole logical theory will have earned at least a conventional place within the hard core of successor programs applying the logic to more specific disciplines. Where a system can be characterized as an extension of a more primitive system, this development will be more piecemeal. Hence, within the classical research program, the propositional and first-order systems are regarded as having attained an optimal fit with natural argumentation, and are placed in the hard core while work continues on issues that are still contentious, such as higher order quantifiers or modal extensions.

We can now diagnose the thesis that logic is irrevisable as a confusion between research programs of different depths within the same tradition. From the perspective of a more developed program, a specific system may be taken as irrevisable, but that program exists within a tradition in which logic may be revised, hence it will always be conceptually possible to revise the system by adopting an ingloriously revolutionary program within the tradition. We can now see that the research programs of a logical conservative and a logical reformer differ not so much in the content of their logical theories as in the partition of this content into hard core and protective belt. The conservative insists on placing the whole formal system within the hard core, and redirecting any apparently conflicting evidence at aspects of the parsing and background theories within the protective belt. As a conventional expedient this could be advantageous, but the conservative regards this as an official view. Thus the supposed irrevisability of logic is relativized to the research program of the logical conservative. Within that program, logic is immune to revision, but the program is not unique, and not guaranteed to succeed. In this sense, both Kant and Frege were justified in regarding logic as nonrevisable, despite having different logics.[24]

An example of the competition between reformers and conservatives can be found in the variety of responses to the problem of the unwelcome existential commitments of nondenoting singular terms. Russell's (1905) "misleading form" strategy and Smiley's (1960, 125ff.) advocacy of a nonbivalent logic are the respective products of logically conservative and reforming research programs. The "misleading form" strategy will be a progressive use of the negative heuristic in the conservative program, but a potentially degenerating use of the negative heuristic in the reform program. Conversely, a move from classical to nonbivalent logic would be outlawed by the negative heuristic of the conservative program, but advocated by that of some reform programs. Since both programs are progressing, we are not yet motivated to abandon either.

The move to an extended logic need not induce a change of research program: Since extended logics do not conflict with the rules of the logic from which they are derived, the syntactic component of the hard core of the research program of that logic may be preserved. Hence, an extension may be an admissible change of theory within a research program. Of course, this is not to say that such a move will always be welcome: The positive heuristic may point elsewhere or the extension may lead to a conflict with hard-core aspects of other areas, such as proof theory or semantics, or inferential goals or background theories. An example of the latter sort of objection is Quine's (1953) opposition to quantified modal logic, which is an extension of propositional modal logic, a system he accepts.[25] Quine's complaint is that if modality is understood *de dicto*, then extension by quantifiers is not conservative of the semantics; we could resolve this by a *de re* understanding of modality, but that would conflict with Quine's preferred background theory.

If the hard cores of logical research programs contained all the rules of inference of their formal systems, the adoption of a nonconservatively revisionary system would always require a change of program. However, at the stage of a research tradition at which logical reform is entertained, we have argued that the hard core should contain only a partial characterization of the system. Hence nonconservative revisions do not always initiate a new program.

An important requirement for this model of scientific change is an account of when research programs and traditions should be abandoned. In essence, the story is the same as that for change of theory within a research program: a program should only be replaced by a rival with greater heuristic or explanatory power, that is, if the rival can explain everything that the original program does, as well as some novelties. However, novelties may be obvious as such only in retrospect, particularly when they turn on the reinterpretation of elements of the original program or tradition. Moreover, a later theory within a defeated program or tradition may be able to make a comeback; only if no such reply is forthcoming should a program or tradition be abandoned. An eventually superior rival may be slow to draw level with and overtake a well-established program or tradition. The positive heuristic of a program need not have been exhausted for the program to be superseded by a more successful rival, although the explanatory potential of a moribund theory should not be overlooked. In practice, this is unlikely to be a problem as the development of a progressive program or tradition is likely to hasten the degeneration of its rivals, since its novel facts will represent anomalies for the rivals. Furthermore, it can be productive to work simultaneously on rival programs within a tradition, or even on rival traditions (Lakatos 1971, 112n3).

This account of theory change is slow, but sure. As in historical science, there are no decisive "crucial experiments," no "instant rationality," but the methodology does provide for the progressive sidelining and eventual elimination of unproductive research programs and traditions (Lakatos 1970, 86f.). Indeed it is crucial that this should happen, lest we fall into a skeptical relativism. Thus we are now in a position to answer a concern raised by a conventionalist account: In logic a research program or tradition may be able to defend itself against refutation indefinitely by repeated employment of a strong negative heuristic. However good its negative heuristic, a program or tradition cannot survive indefinitely in the face of a more explanatory rival. Yet where the negative heuristic is especially strong, as in logic, the transition may be very slow. This tardiness motivates a methodological commitment to scientific pluralism; science cannot advance without competition between programs. It is particularly important that no theory is permitted to achieve a position of hegemony that permits it to dispatch potential rivals before they have developed sufficiently to pose a threat. Some commentators, for example, Priest (1989a, 138ff.), have been keen to diagnose this condition in contemporary classical logic.

1.4. Classical Recapture

The recapture relationship is an important element to any understanding of the connection between different systems of logic.[26] Loosely speaking, one system of logic recaptures another if it is possible to specify a subsystem of the former system which exhibits the same patterns of inference as the latter system.[27] In particular if a relationship of this kind can be shown to exist between a nonclassical logic and **K**, the nonclassical system is said to exhibit classical recapture. This has been invoked by several proponents of nonclassical logics to argue that their system retains **K** as a limit case, and is therefore a methodologically progressive successor to **K**. In this section we advance and defend a new and more precise account of recapture and the character of its reception by the proponents of the recapturing system. We then indicate some of the applications of classical recapture that this account makes possible.

Our account of recapture builds on an account of the equivalence of consequence systems developed in Aberdein (2000). When $\mathbf{L}_1$ and $\mathbf{L}_2$ are equivalent, we write $\mathbf{L}_1 \cong \mathbf{L}_2$. The account of equivalence utilized a schematized representation of such systems, $\mathbf{L}_i$, as couples, $\langle W_i, V_i \rangle$, where W_i is the class of well-formed formulae of the language underpinning logic $\mathbf{L}_i$ and V_i is the class of valid inferences of $\mathbf{L}_i$ (a subclass of the class of sequents[28] defined on W_i). Equivalence consists in a one-to-one correspondence between equivalence classes of the wffs of the systems that preserves the partitions of the classes of inferences into valid and invalid subclasses.

Definition 1 $\mathbf{L}_1$ is a *proper reduct* of $\mathbf{L}_2$ iff $\mathbf{L}_1$ and $\mathbf{L}_2$ are inequivalent, W_1 is defined on a proper subset of the class of constants of $\mathbf{L}_2$ and V_1 contains precisely those elements of V_2 which contain only elements of W_1.

Hence, reduction is the inverse of conservative extension. Formally, we may say $\mathbf{L}_1$ extends $\mathbf{L}_2$ iff $\mathbf{L}_1$ and $\mathbf{L}_2$ are inequivalent and $\mathbf{L}_1$ is equivalent to a logic that has a proper reduct that is equivalent to $\mathbf{L}_2$. However, reducts are not the only sort of contractions that may be defined on formal systems; the definition may be generalized as follows.

Definition 2 $\mathbf{L}_1$ is a *proper subsystem* of $\mathbf{L}_2$ iff $\mathbf{L}_1$ and $\mathbf{L}_2$ are inequivalent, W_1 is a proper subset of W_2 and V_1 contains precisely those elements of V_2 which contain only elements of W_1.

The metaphors of strength, size, and inclusion which so often illustrate the mereology of logical systems suffer from an ambiguity: There is a tension between a deductive characterization, a measure of how much may be deduced from how little, and an expressive characterization, a measure of the subtlety of the distinctions that can be preserved.[29] An increase in one may represent a decrease in the other. Hence, "subsystem of **L**" has often been used to designate a system axiomatized by a subset of the axioms of **L**, or with a deducibility relation which is a subrelation of that of **L**. The definition of subsystem adopted above reverses this usage, making explicit the generalization of the definition of reduct, but rendering these latter "subsystems" supersystems, the inverse of subsystems. In short, reducts are exclusively generated by reducing the set of constants on which the class of wffs is based, but subsystems may also be generated by reducing the class of wffs in some other way. For example, **K** is a subsystem of intuitionistic logic, **J**. Only some of the formulae of **J** are decidable: those for which the law of the excluded middle (LEM) is valid (and double-negation elimination [DNE] is admissible). Restricting **J** to precisely these formulae, as could be achieved in the appropriate presentations by adding LEM to the axioms of **J**, or DNE to the definition of its deducibility relation, produces a subsystem, **K**. But this *sub*system has either an extra axiom or an extra rule of inference.

This apparatus provides the means for a formal account of recapture.

Definition 3 $\mathbf{L}_1$ *recaptures* $\mathbf{L}_2$ iff there is a proper subsystem of $\mathbf{L}_1$, $\mathbf{L}_1^*$, which is defined in terms of a constraint on W_1 finitely expressible in $\mathbf{L}_1$, and which is equivalent to $\mathbf{L}_2$. If $\mathbf{L}_2$ is **K**, then $\mathbf{L}_1$ is a *classical recapture logic*.

That is to say, if one system recaptures another, we may express within it some finite constraint by which a subsystem equivalent to the recaptured system may be generated. For example, we can see that **J** is a classical recapture logic, with the constraint of decidability. The relevance system **R** has also been claimed to recapture **K**, with the constraints of negation consistency and primality (see Mortensen 1983). Quantum logic also recaptures **K**, with the constraint of compatibility. Indeed, many nonclassical logics are classical recapture logics: exactly which will turn on which constraints are deemed expressible. It has even been suggested that the recapture of **K** is a necessary criterion of logicality, in which case all logics would be classical recapture logics.[30]

Different nonclassical logicians have different attitudes to classical recapture. Some attempt to reject it outright or deny its significance, others embrace it,

Radical left	"My system does not recapture **L**."
Center left	"My system does recapture **L**, but this is merely a technical curiosity."
Center right	"My system recaptures **L**, which shows that **L** is retained as a limit case."
Reactionary right	"My system recaptures **L**—and extends it, too."

Figure 14.1

while others see recapture results as motivating the reduction of the recapturing system to a conservative extension. Thus, before recapture can contribute to the understanding of how logical systems change, we must distinguish among the variety of responses that advocates of a system may make to the prospect of recapturing a prior system (typically **K**). We order these responses by analogy with a spectrum of political attitudes: radical left, center left, center right, and reactionary right. This is a formal not a sociological analogy: We do not intend to imply that views on logic may be correlated to political allegiance (pace some sociologists of scientific knowledge). The spectrum of attitudes to the recapture of the prior system **L** may be summarized in fig. 14.1.

The most extreme attitude is the radical left: formal repudiation of recapture status. Individuals of this tendency deny that their system recaptures the prior system, claiming that no suitable recapture constraint is expressible in the new system. If classical recapture were a criterion of logicality, then a radical-left response could only be embraced by quitting the discipline of logic. Yet such a criterion must be open to doubt, since some familiar programs include proponents from the radical left. For example, Nuel Belnap and Michael Dunn's argument that relevance logic does not recapture **K** places their relevantist in this camp (Anderson, Belnap, and Dunn 1992, §80.4.5, 505).[31] The subordination of logic to mathematics by some intuitionists may also be understood as preventing classical recapture.

The less radical center left acknowledge the formal satisfaction of recapture, but deny its significance. Proponents of this stance argue that the formal equivalence between a subsystem of their system and another system is irrelevant, since the other system cannot be understood as formalizing anything intelligible in terms of their theory. Hence some advocates of **J** regard the double-negation translation of **K** into their system as no more than a curiosity, since they reject the cogency of classical concepts.[32] Whereas the radical left presume a logical incompatibility between the recapture result and indispensable formal components of the research program, the center left claim an heuristic incompatibility with indispensable nonformal components of the research program. To defend a position on the center left, one must demonstrate that conceding more than a technical significance to recapture will induce an intolerable tension between successful problem solving within the program and the retention of its key nonformal components, such as the central aspects of its parsing theory. Thus, although a recapture constraint can be articulated, it does not correspond to any plausible feature of natural argumentation.

On the center right, recapture is embraced as evidence of the status of the new system as a methodologically progressive successor. The meaning invariance of all key terms is welcomed in this context, and recapture is understood as establishing the old system as a limit case of its successor. The center right hold with Einstein that "there could be no fairer destiny for any ... theory than that it should point the way to a more comprehensive theory in which it lives on, as a limiting case" (1916, 77). By contrast, left-wing recapture involves a far more comprehensive rejection of the old system, by which its intelligibility is denied, and it is ultimately to be dismissed as an incoherent wrong turning. This is much more plausible behavior in a competitor than a successor theory, and suggests left-wing recapture as a criterion for this tricky distinction. This is corroborated by the enthusiasm shown for classical recapture among systems typically promoted as succeeding **K**, and the opposition shown by its self-proclaimed competitors. Most nonclassical logics have been defended as successors to **K** by at least some of their advocates. For example, Hilary Putnam's quondam advocacy of quantum logic was of this character, as is Graham Priest's support for paraconsistent logic: Both logicians find classical recapture significant and take care to establish it for their systems (Putnam 1969, 184; Priest 1987, 146ff.). Conversely, the most credible left-wing stance is from proponents of **J**, and this system has the greatest claim to be a true competitor to **K**, rather than a would-be successor.

Least radical of all are the reactionary right, who argue that the subsystem of the new system equivalent to the old system is actually a proper reduct of the new system, that is, that the new system should be understood as extending the old system. Hence the *status quo* is maintained: The old system is still generally sound, but can be extended to cover special cases. In this case there is no rivalry between the systems (see Haack 1974, 2), because there is no disagreement within the common ground they share. Many ostensibly nonclassical programs have at some stage been promoted as conservative extensions of **K**: for example, Maria Luisa Dalla Chiara's (1986, 447) modal quantum logic $\mathbf{B}^{o}$ or Robert Meyer's (1986) classical relevance system $\mathbf{R}^{\neg}$. Modal logic may be understood as having successfully completed a move from the center right to the reactionary right: Although it is now understood as extending **K**, its early protagonists conceived it as a prospective successor system.[33]

Note that if $\mathbf{L}_1$ extends $\mathbf{L}_2$, then $\mathbf{L}_1$ recaptures $\mathbf{L}_2$ and in fact this is the *only* way in which $\mathbf{L}_1$ can recapture $\mathbf{L}_2$, if $\mathbf{L}_1$ extends $\mathbf{L}_2$. For, if $\mathbf{L}_1$ is an extension of $\mathbf{L}_2$ then $\mathbf{L}_2 \cong \mathbf{L}_3$, where $\mathbf{L}_3$ is a proper reduct of $\mathbf{L}_1$. But because $\mathbf{L}_1$ recaptures $\mathbf{L}_2$, $\mathbf{L}_2 \cong \mathbf{L}_1^*$, where $\mathbf{L}_1^*$ is a subsystem of $\mathbf{L}_1$. So by transitivity of equivalence, since $\mathbf{L}_1^* \cong \mathbf{L}_2 \cong \mathbf{L}_3$, $\mathbf{L}_1^* \cong \mathbf{L}_3$: The subsystem by which $\mathbf{L}_1$ recaptures $\mathbf{L}_2$ is equivalent to a proper reduct of $\mathbf{L}_1$. For example, the subsystem of $\mathbf{S4}_{\neg 3}$ equivalent to **K**, which establishes that $\mathbf{S4}_{\neg 3}$ is a classical

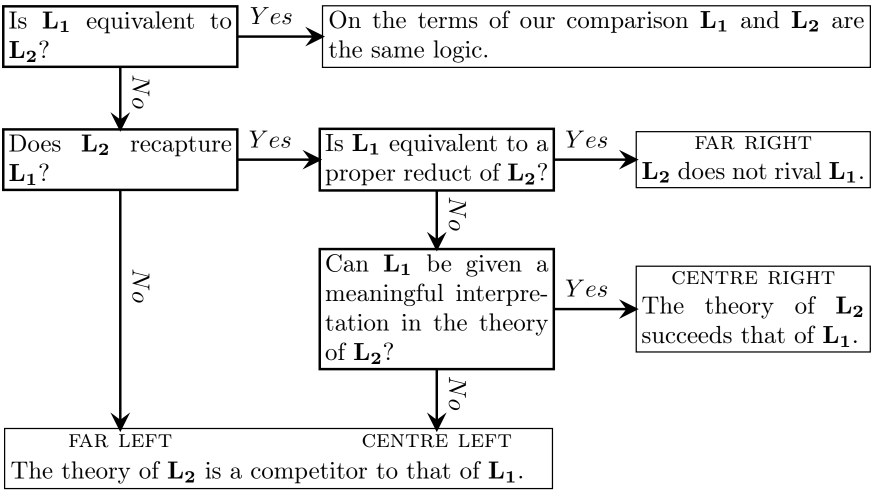


Figure 14.2

recapture logic is itself equivalent to the proper reduct of $\mathbf{S4}_{\neg 3}$ defined over that system's nonmodal constants. Thus, if the reactionary stance is technically feasible, it is the only plausible response to recapture. This represents a dualism with the radical stance, which is also mandated by properties of the chosen formal system.

Different logical research programs encompass different "political" complexions: Some are clearly associated with one stance, whether for technical or historical reasons, in others there is dispute as to which approach is appropriate. Two further points may serve to reinforce the political analogy: Programs appear to drift to the right as they grow older, and there is a strong community of interest between the two ends of the spectrum. The reactionary agrees with the left-wingers that the constants of the new system have different meanings from those of the old. The difference is that the left wing think that the new meanings must replace the old, whereas reactionaries believe that they can be assimilated into an augmented system through employment alongside the old meanings. The greater the difference between the new and the old constants, the more difficult it is to maintain a centrist position.

The full range of options may be seen more clearly as a flow chart, shown in fig. 14.2. This chart has been devised to display the consequences of a change of theory in which a specific formal system ($\mathbf{L}_2$) replaces another ($\mathbf{L}_1$). However, it should be stressed that in the practical development of logical research programs, a dialectic exists between the choice of formal system and the attitude taken to the recapture of the prior system. Hence, providing that enough of the formal system remains within the revisable part of a logical research program, there are always two alternatives: Embrace the consequences of the formal system, or change the system to resist them.

With this picture in place, we can begin to outline some of the uses to which it may be put. In the first place, we now have the resources to draw some fundamental distinctions between different sorts of theory change. An important feature of the flow chart is that its first three questions can be answered purely by comparison of the formal systems $\mathbf{L}_1$ and $\mathbf{L}_2$, but the fourth question—"Can $\mathbf{L}_1$ be given a meaningful interpretation in the theory of $\mathbf{L}_2$?"—requires an appeal to the theories by which the systems are advanced, and perhaps the research program behind that. Hence, while certain outcomes are necessitated by formal features, other outcomes are underdetermined by such data alone. Solely on formal data we can observe that rivalry must occur unless one system conservatively extends the other, and that competition must occur unless one system recaptures the other. However, broader consideration is required if more than these weak sufficiency conditions for the rival/non-rival and competitor/successor distinctions are sought. Indeed, logical *theories* can be rivals even when the embedded *systems* are related by conservative extension, or even equivalence: For example, $\mathbf{R}^{\neg}$ conservatively extends $\mathbf{K}$, but its promotion would presume a radically nonclassical parsing theory, and many systems of logic have more than one alternative semantics, promoted by rival theories.[34] Yet presuming that the remainder of the theory changes no more than necessary, a clear taxonomy of the consequences of different species of logical revision may be seen to emerge.

1.5. Heuristic Contexts

In 1.3 we saw how helpful MSRP could be in reconstructing the history of alternative logic. However, Lakatos's greatest contribution to the philosophical analysis of logical methodology is to be found in his *Proofs and Refutations*. Much of this work is spent in an attempt to articulate what he would come to call positive and negative heuristics for research programs in mathematics, a goal in which he was strongly influenced by the work of George Pólya.[35] At the center of Lakatos's idealized heuristics is a useful account of the variety of responses to anomaly and their significance for theoretical development, which may be applied to both formal and empirical subjects. He distinguishes four strategies of response: "monster-barring," "monster-adjusting," "exception-barring," and "monster-exploiting" (Lakatos 1976, 14 ff.).[36] *Monster-barring* is the strategy of excluding anomalous cases from consideration by constructing ever tighter definitions of the subject matter. "Using this method one can eliminate any counterexample to the original conjecture by a sometimes deft but always ad hoc redefinition of the [subject matter], of its defining terms, or of the defining terms of its defining terms" (ibid., 23). *Exception-barring* "plays for safety" by restricting the domain of the theory so that the anomalous area is no longer treated. Exception-barring coincides with "monster-barring in so far as [the latter] serves for finding the domain of validity of the original conjecture; [but] reject[s] it in so far as it functions as a linguistic trick for rescuing 'nice' theorems by restrictive concepts" (ibid., 26). In its most primitive form this amounts to seeking to acknowledge the anomalies without altering the theory (ibid., 36). *Monster-adjustment* redefines the purported counterexample into terms that no longer conflict with the theory. Finally, *monster-exploiting* is the employment of anomalies as motivation for theoretical innovation and development. Primitive exception-barring, monster-barring, and monster-adjustment are strategies from the negative heuristic: They represent increasingly sophisticated methods for resisting the pressure for change exerted by an anomaly. Exception-barring and monster-exploiting are positive heuristic strategies: They utilize anomalies to improve the "original conjecture," which is the antecedent content of the theory.

Lakatos illustrates these strategies through worked examples, the most substantial of which concerns the Euler conjecture, $V - E + F = 2$, which relates the numbers of vertices (V), edges (E), and faces (F) of polyhedra (ibid., 6ff.).[37] This relationship can be easily verified for the five Platonic solids (regular polyhedra whose sides are regular polygons). Further inquiry turns up apparent counterexamples to the Euler conjecture: concave and stellated polyhedra; hollow polyhedra; twin polyhedra, formed by joining pairs of polyhedra at a vertex or an edge; the cylinder; and the "picture frame." Lakatos traces the history of attempts to prove and improve the Euler conjecture, from its inception in the 1750s to the origins of modern topology more than a century later. He imaginatively reconstructs the dialectic implicit in the development of this area of mathematics as a classroom dialogue. The methods discussed are introduced in turn as increasingly sophisticated responses to the puzzle cases.

For example, all the counterexamples could be ruled out of consideration by the blatantly nonexplanatory move of making satisfaction of the Euler conjecture part of the definition of "polyhedron": primitive exception-barring. More productively, successive monster-barring definitions of "polyhedron" could be adopted to exclude various counterexamples. For instance, if polyhedra are defined to be surfaces rather than solids, then hollow solids no longer count as polyhedra.[38] Less ad hoc still is the exception-barring move of restricting the domain of the Euler conjecture to cases to which it has been established to apply, such as convex polyhedra, with a view to determining its precise domain of application. Alternatively, puzzle cases may be reconciled with the conjecture by monster-adjustment. In this way the small stellated dodecahedron may be seen to satisfy the Euler conjecture if its faces are counted as 60 triangles, but not if they are counted as 12 pentagrams (ibid., 31).[39] For a compelling application of this method, an explanation of why the helpful interpretation should be adopted is required. Finally, Lakatos's preferred method, monster-exploiting, can be seen in two further moves: lemma-incorporation, whereby hidden assumptions are made explicit within the conjecture, and the increasing of content by replacing lemmata by others of wider generality.

An illustration of the spirit behind this sequence of methods is provided by David Bloor (1978, 252ff.; 1983, 139ff.), who assimilates Lakatos's treatment of anomaly to Mary Douglas's (1975, 306f.) anthropological account

	Low Group	High Group
Low Grid	Primitive Exception-Barring.	Monster-Adjustment; Exception-Barring.
High Grid	Monster-Exploiting.	Monster-Barring.

Figure 14.3

of possible responses to strangers.[40] She classifies societies with respect to axes representing the degree of "grid" and "group." Grid measures the importance of internal boundaries of rank, status and so forth to a society. Group measures the strength of the boundary separating the society from the rest of the world. High grid, low group societies are preoccupied with internal divisions and indifferent to the actions of strangers. Low grid, high group societies have strong social cohesion but little internal order, and are inclined to be hostile to strangers. Such open hostility will not work in high grid, high group societies because an excluded stranger might be exploited by another subgroup. Hence individuals within these societies will seek either to justify overall exclusion of the stranger, or to assimilate him into their own subgroup. Low grid, low group societies are competitive and individualistic; strangers are welcomed for the advantage they may bring to individual competitors. This structure may be represented diagrammatically (Douglas 1970, 82ff.). A diagram of this kind (fig. 14.3) demonstrates how Lakatos's responses to anomaly are related to Douglas's responses to strangers (Bloor 1978, 258). Thus primitive exception-barring corresponds to indifference, monster-barring to fear and aggression, monster-adjustment to assimilation, exception-barring to well-motivated exclusion, and monster-exploiting to opportunistic exploitation. This picture assembles the different responses into an implicit hierarchy, from decadent primitive exception-barring, through isolationist monster-barring, aristocratic exception-barring, and whiggish monster-adjusting to free-market monster-exploiting.

So far we have followed Bloor (and diverged from Lakatos, for whom sociological factors are irrelevant to rational reconstruction) in the central assumption of the strong program in the sociology of scientific knowledge: that theories resemble the societies that produce them, thereby associating each strategy with a society in which it is expected to be typical. However, we can retain this picture as an account of the heuristic practices characteristic of different stages in the development of research programs, while abstaining on this sociological assumption. Abstracting from the sociological detail, in accordance with Lakatos's principles of rational reconstruction, we may thereby think of each quadrant of the diagram in fig. 14.3 as a *heuristic context*. It is difficult for Bloor to explain how the same societies, the same institutions, and even the same individuals can simultaneously contribute to multiple disciplines occupying different heuristic contexts. By decoupling sociological context from heuristic context, it becomes easier to see why each strategy will be hard to defend away from its home quadrant. For instance, Bloor's (1983, 146) contention that it would be impossible to sustain monster-barring in a low grid, low group society immediately invites empirical counterexample. The underlying point is more easily accepted: A methodological move that does little more than isolate anomalies will not be of much use in an heuristic context in which diversity and experimentation are encouraged.

1.6. A Hierarchy of Logical Reform

The hierarchy of heuristic contexts, when applied to a reform-minded logical research tradition, yields the following sequence of possible responses to the pressure for change of logical system.[41]

- I Indifference: primitive exception-barring;
- II Nonrevisionary responses:
  - a. Delimitation of the subject matter of logic:
    - i. monster-barring;
    - ii. exception-barring;
  - b. "Novel paraphrase": monster-adjustment;
  - c. "Semantic innovation": monster-adjustment;
- III Conservatively revisionary response: monster-exploiting;
- IV Nonconservatively revisionary responses:
  - a. Restriction of the logic: exception-barring;
  - b. Wholesale revision: monster-exploiting;
- V Change of subject matter: monster-exploiting.

In this section we explain and illustrate the levels of this hierarchy. At the first level is brute indifference to the problem: primitive exception-barring. We can find plenty of examples in logic of refusal to acknowledge the existence of a problem, particularly in the early stages of the development of a program. Responses to the paradoxes of implication in the early development of the classical program furnish several examples. For instance, Russell (1903, 34) is prepared to argue that material implication offers an adequate account of entailment, a view subsequently described by Moore (1919, 58) as "an enormous howler." Russell's obstinacy might have had some advantage in maintaining the forward momentum of the program in its earliest heuristic context; after the program attained more systematicity, it became less defensible.[42] A more defeatist than obstinate indifference is the counsel that we should just put up with the problem: "the paradoxes of Strict Implication ... are unavoidable consequences of indispensable rules of inference" (Lewis 1932, 76).

The next step up are responses that are not revisionary of the formal system. The first of these, delimitation of the subject matter, consists in ruling the puzzle cases to be inappropriate for logical formalization. This could be either monster-barring, or, if sufficiently systematic, exception-barring. The monster-barring variant is typical of contexts where the overwhelming concern is maintenance of the boundary of logicality. Saint Anselm's injunction that "the heretics of logic are to be hissed away," quoted with approval in Burgess's (1983, 41) critique of relevance logic, is the motto of this approach. Further examples include Strawson's treatment of sentences with nondenoting subject terms as "spurious," and thus unfit for logical formalization;[43] and Resnik's (1985, 228) response to apparent counterexamples to non–truth-functional logic that "prior to the discovery of truth-functional logic no one would have thought of them." The context of these moves is suggestive of a low grid/high group heuristic context: Strawson is defending a general account of logical formalization; Resnik a general account of logical normativity.

Where the emphasis is on describing the limitations of formalization, rather than merely maintaining them, more systematic, and thereby exception-barring, responses result. The exclusion of vagueness from Frege's (1879) highly programmatic attempt at a *calculus ratiocinator* in his *Begriffsschrift* exhibits this response, since the exclusion proceeds from his attempt to articulate a logically perfect language, and is not just an ad hoc stipulation.[44] By contrast, his proposal to exclude nondenoting terms, by providing referents for all definite descriptions by stipulation, is more naturally viewed as monster-barring. This assessment, and that of the *Begriffsschrift* as a contribution to a high grid, high group enterprise, is reinforced by the swift recognition by other researchers in the same program of the incompatibility of this proposal with the heuristic context then occupied by their program.[45] An example of a proposal from a slightly less systematic program, is Peirce's treatment of the paradoxes of material implication as benign because of the "somewhat special sense" of "if ... then" used in logical contexts (Peirce 1896, cited in Passmore 1957, 140).[46] This is closer to primitive exception-barring, plausibly enough, since we could make a case for Peirce's program being situated not quite so far along the group axis as Frege's, because of his development of logic against a broader semiotic background.

The next nonrevisionary response, the novel paraphrase strategy, is most familiar from Russell's (1905, 480ff.) misleading form treatment of nondenoting singular terms. Grice's (1975) attempt to reconcile classical logic with the idiosyncrasies of natural language by means of "conversational implicatures" seeks to develop this method into a comprehensive account of what a suitable parsing theory for rough classical logic should look like.[47] Carnap's proposal for replacing vague expressions by precisified, "scientific" paraphrase prior to formalization exhibits the same approach (1950, cited in Haack 1974, 120). An example from a nonclassical program is the relevantist proposal to interpret the occurrence of "or" in *prima facie* valid instances of disjunctive syllogism (which is not generally valid in systems such as **R**) as fission rather than disjunction (Anderson and Belnap 1975, §16, 166). This strategy sets out to reinterpret the anomaly to reconcile it with the formal system central to the research program and thus employs monster-adjustment.

At its most subtle, this species of monster-adjustment can take the form of an admonition to understand formalized propositions in a particular way, rather than explicit paraphrase. For example, Wittgenstein (1921, §5.25; §5.254) seeks to avoid intuitionistic problematization of double-negation elimination (DNE) by counseling that negation be understood as an operation taking a proposition to its contradictory, rather than a constituent of propositions.[48] Ramsey (1927, 161f.) sought to capitalize on this idea with the suggestion that negated propositions be written upside down, making skepticism about DNE formally inexpressible.[49] This move involves a revision of notation, although not of the underlying system, bringing it closer to the next sort of monster-adjusting move, semantic innovation, and, by principled exclusion of puzzle cases from formalization, shows affinities to exception-barring. That these three methods can be so closely related is further corroboration for the taxonomy, since they share a heuristic context.

Also employing monster-adjustment are the various proposals to preserve classical logic by a more complicated semantics. For example, Kripke's (1975) proposal to address paradoxes of self-reference by employment of three-valued matrices that permit semantic consideration of wffs that have not (yet) received a definite evaluation as true or false, or van Fraassen's (1966) "supervaluational" semantics.[50] The proposers of both of these schemes present them as augmenting an underlying classical semantics, a monster-adjusting step, rather than as introducing a novel system of logic with a nonclassical semantics, which would place them further down the hierarchy.[51] There is some scope for skepticism whether monster adjustment is sufficient for the success of these proposals, particularly in van Fraassen's case, because it might be argued that the retention of classical inference compromises the problem-solving efficacy of the semantic innovation (see Read 1995, 142). More extensive revision of the classical logical program may still be required. All the examples of either of the two monster-adjusting steps available to logicians occur in sophisticated and highly structured programs, generally in response to more radical competitor proposals: high grid, high group heuristic contexts.[52]

The next level of the hierarchy consists of conservatively revisionary logical responses. These typically take the form of a switch to an extended logic in which a satisfactory treatment of the anomalies may be developed. Numerous examples can be furnished by most logical research traditions, involving extension by various sorts of quantifiers, identity functions, set-membership operators, and alethic, deontic, temporal, doxastic, and other modal operators. This strategy is monster-exploiting—in a modest way—and potentially progressive, although not all anomalies will yield to this treatment. Most of the extensions listed have been accompanied by rearguard claims that the resulting system is no longer purely logical or even intelligible. Examples of both moves may be found in Quine: His claim (1970, 68) that higher order quantification is mathematics, not logic is of the former kind, whereas his opposition to quantified modal logic (Quine 1953) is an example of the latter kind. These moves correspond to monster-adjusting and exception-barring moves, respectively:[53] That they are so controversial suggests that extending a logic is a tactic from a different heuristic context. Indeed, it is a low grid, low group move—monster-exploiting—in the modest sense that it requires acknowledgment that the formal system is not set in stone.

This assessment of conservative extension is clearest where it is the most radical of all proposed responses to the anomaly. By contrast with a nonconservative proposal, extension seems more of a monster-adjustment strategy; this is the role that it has in the reactionary response to recapture.[54] The point is that adopting an extended logic involves adjusting the anomalous cases sufficiently for them to be treated in a logical theory which is conservative over the prior theory, but also requires augmentation of the prior theory, and is therefore monster-exploiting. A change of inferential goals not motivated by the adoption of incompatible background theories would yield a novel research program that was not really a competitor to the original, and therefore treated at this level of the hierarchy. Because the background theories of the two programs would be compatible, the goal of one system could be expressed satisfactorily within the context of the other, hence it would be possible to remove the conflict altogether by representing the former system within an extension of the latter.[55] Accomplishing this nonrevisionary logical response, the aim of the reactionary response to recapture, would be an impressively progressive achievement for the program producing the extended logic, since it would acquire all the additional content of the other system.

In Kuhnian terms, the first three levels of the hierarchy represent the "normal science" of a logical research program. The heuristic contexts of indifference: "new sorts of phenomena ... are often not seen at all" (Kuhn 1962, 24); assimilation: "matching of facts with theory" (ibid., 34); and the limited enthusiasm of applying an existing method to a new area: "manipulations of theory undertaken ... to display a new application" (ibid., 30) are all suggested by Kuhn as typical activities of the normal scientist. However, there are two significant contrasts between Kuhn's position and that adopted here. First, Kuhn distinguishes only two heuristic contexts: normal science and crisis. Second, normal science is taken by Kuhn to be constitutive of, and dominant within, a whole discipline, not just of a research program or tradition within a discipline. Each of these contrasts serves to blunt Kuhn's controversially sharp dividing line between normal and revolutionary science. For Lakatos (1970, 69), criticism and competition are healthy, and hegemony is pathological: This is the reverse of Kuhn's evaluation.

In the fourth level of the hierarchy we find the responses employing a nonconservative revision of logic. The first of these is restriction of the logic:

avoidance of the anomaly by moving to a logic that lacks previously valid inferences and theorems. This exclusion of the puzzle cases from treatment is systematic, and thereby exception-barring, provided that the calculus resulting from the restriction has a finite, well-behaved presentation (without which the restriction would be blatantly degenerating). As the revision involved cuts deep, solely exception-barring uses of restriction are out of tune with the heuristic context necessary for their deployment, and are seldom encountered as serious reform proposals. Some logics, such as Birkhoff and von Neumann's (1936) nondistributive quantum logic, begin life as solely restrictive steps, and subsequently form the basis of progressive research programs, but only by additional monster-exploiting moves.[56] This is possible because eventually successful programs can survive occasional periods of degeneration, and conflicts between programs are not settled at the first contest (Lakatos 1970, 71).

The heuristic context sufficient for restriction characteristically results in a more substantial revision. This is the second sort of nonconservative revisionary response: wholesale revision, in which elements of the logical theory beyond the formal calculus are exposed to criticism, and reformulated in response. These elements, which include metalogical concepts, such as that of consequence, background theories and the inferential goal, are predominantly situated within the hard core of mature programs. So, except in the infancy of a program, when almost all of its content is still fluid, wholesale revision will initiate a new program, although not necessarily a new tradition. In the case of quantum logic, this stage occurred after a hiatus of some 30 years, in which the formal system was better known as a contribution to pure mathematics or as an interpretation of the foundations of physics—roles in which it has continued to progress, despite the degeneration of Putnam's quantum logic program (Coecke, Moore, and Wilce 2000, 6; Foulis 1997). The formal calculi associated with the intuitionistic, relevance, and paraconsistent programs explored in section 2 are also restrictions of that of classical logic, but all of these were developed in parallel with, or subsequent to, more radical moves.

How does wholesale revision work? Judicious restriction can permit clarification, precisification, and disambiguation of previously confused concepts. For example, as we will show in 2.3.1, the adoption of relevance logic permitted the articulation of the contrast between intensional and extensional constants, obscured in classical logic, and a more sensitive restatement of the consequence relation. Hence, in Lakatosian terms, the search for motivation for exception-barring steps can lead to a revision through proof analysis of the primitive conjecture (here the claim that a given logic is adequate for the formalization of natural argumentation), and thus constitute monster-exploiting. Lakatos (1976, 50, 136) quotes with particular approval the methodological injunction (from Seidel 1848) that "if you have a global counterexample [a counterexample to the main conjecture] discard your conjecture, add to your proof analysis a suitable lemma that will be refuted by the counterexample, and replace the discarded conjecture by an improved one that incorporates the lemma as a condition." For Lakatos this insight was crucial to the history of nineteenth-century mathematics, since it initiated the "method of proofs and refutations"—that is, monster-exploiting.[57] Bloor (1978, 263ff.) argues that this innovation was made possible by the changed social structure of German universities that resulted from earlier government reform proposals. What it undoubtedly shows is the adoption of a heuristic context in which more radical methods than had previously been deemed legitimate could be entertained.

Finally, we come to a strategy more radical than any yet addressed: change of subject matter (see Haack 1978, 155; Beall and Restall 2000, 490). We saw before that a change of inferential goal in which the background theories are preserved can occur at the conservatively revisionary level of the hierarchy. But changes of goal can also be precipitated by a nonconservative revision of the background theories. Typically this will alter the motivation of the whole logical enterprise, move the problem into a different area, and change the subject matter of logic. In so far as goals and the background theories that justify them are deep within the hard core of a program, their nonconservative revision must initiate a change of research program, and probably of research tradition. Thereafter the question of which program should be pursued, of which logic should be employed, can no longer be addressed directly. It is superseded by the question of which background theories obtain, and thereby of which goal is being pursued.

The proper place for settling disputes of this sort is at the level at which the background theories conflict, not at the level of the different calculi. Any divergence at the latter level is understandable but derivative: They have been designed to meet different specifications. Therefore the dispute is no longer in the discipline of logic, but rather in whatever discipline threw up the conflicting background theories. However, it is not impossible for goals and background theories to be revised without a change of program (or tradition), if the positive heuristic is specified in sufficiently general terms. Hence there is a crucial difference between responding to a problem with a novel positive heuristic whereby the goal and background theories are radically changed, and gradually adjusting the goal and background theories, in coevolution with other aspects of a logical research tradition, while preserving the positive heuristic. The latter move may be understood as wholesale revision, the previous level of the hierarchy, but the former is more profound, and can only be represented as a change in the subject matter of logic, the final level of the hierarchy.

Among proposals of this character are accounts of logic as the science of information flow;[58] systematic approaches to informal logic;[59] and perhaps some attempts at a "feminist" logic.[60] One of our goals in section 2 will be to argue that, while the relevance, quantum, and paraconsistent programs may be understood as wholesale revisions, intuitionism goes further and involves a change of subject matter. It is important to observe that the nonconservative revision of background theories involved in a change of subject matter need not entail an inglorious revolution in the formal system.[61] We explore its more positive applications in the conclusion.

2. What Alternative Logics Are There?

Each of the four subsections of section 2 is a case study applying the methods developed in section 1 to a specific reform proposal. Many different nonclassical systems have been promoted, particularly in recent years. One might mention: modal and multimodal systems, including alethic, temporal, deontic, epistemic, and doxastic modalities; paracomplete[62] and many-valued logics; free logic; fuzzy logic; second-order logic; nonmonotonic and dynamic logics; resource-sensitive and linear logics; and many other systems. To stay within a manageable length, and to retain some unity of focus, we have restricted our case studies to a much smaller range. We have concentrated on systems which have been seriously proposed as rival organons to propositional **K**. The focus on the propositional case is because it is where the classical program is at its strongest, and because the choice of quantifiers is seldom as fundamental as that of propositional constants.

Within these constraints, we have chosen a range of systems, all of which are independently interesting and each of which illustrates particular aspects of our discussion of logical revisionism in section 1. The first case study is of intuitionistic logic, **K**'s oldest and most familiar rival. In the second case study we turn to quantum logic, a system proposed on empirical grounds as a resolution of the antinomies of quantum mechanics. The third case study is concerned with systems of relevance logic, which have been the subject of an especially detailed reform program. Finally, the fourth case study is paraconsistent logic, perhaps the most controversial of serious proposals.

2.1. Intuitionistic Logic

The earliest and most enduring alternative to classical logic is intuitionistic logic, which has provided the formal component of several distinct programs. We shall begin by setting out the distinguishing features of the formal system, and of the two most important programs: mathematical constructivism and semantic antirealism. More detailed exegesis exploring the differences and important similarities of these programs follow.

2.1.1. What Is Intuitionistic Logic?

The origins of intuitionistic logic lie in constructivist philosophy of mathematics. Like much contemporary philosophy of mathematics, constructivism originated as a response to the crisis in the foundations of mathematics caused by the discovery of set-theoretic paradoxes induced by the unrestricted application of infinitistic methods. In common with several other approaches, such as Hilbert's formalism, constructivism sought to address this crisis by concentrating on a nonparadoxical domain of mathematics. Several different schools of constructivism may be identified, but they all achieve this narrowing of focus by arguing that the statements of mathematics should be understood in terms of proof rather than (classical) truth. This makes asserting the existence of mathematical objects illegitimate unless there are proofs of the existence of specific examples of each such object, that is to say, a means of constructing the object in finitely many steps. There is a sharp divide between most constructivists and mainstream philosophy of mathematics since constructivism is generally revisionary of mathematics, claiming that certain hitherto acceptable areas of mathematics should be discarded.[63]

It is possible to reconcile this attitude to mathematics with the retention of classical logic.[64] However, from the characteristic intuitionistic stance, mathematics is foundational, and logic is an anthology of a posteriori rules which mathematics has been found to obey. Hence it would be begging the question against the intuitionist to regard the existence of classically grounded constructivist programs as an argument against intuitionism: So to argue would be to presume the priority of (classical) logic, which the intuitionist specifically disputes (Haack 1974, 93). This intuitionistic stance originates with Brouwer, who defended his program as the recognition that "mathematics is an essentially languageless activity of the mind, having its origin in the perception of a move of time" (Brouwer 1952, 141). Logic is then no more than a formalization of the language used to describe this activity: If permitted to run unchecked, it risks outstripping the intuitions constitutive of mathematics. Subsequent intuitionists have placed less emphasis on Brouwer's Kantian approach to intuition; the key notion that remains is that because provability is the touchstone of good mathematics, mathematicians should cleave closely to it, and not rely on generalizations over "objects" for which no construction has been provided.

Adherence to these scruples requires the abandonment of certain familiar principles of classical logic, such as the law of the excluded middle (LEM), $A \vee \neg A$, and double-negation elimination (DNE), $\neg\neg A \vdash A$. For, if constructions are the only warrant for mathematical assertions, the occurrence, in any non-finite domain, of mathematical propositions for which we can construct neither a proof nor a refutation, conflicts with the unrestricted assertion of LEM. And the establishment of the lack of a construction establishing the lack of a construction of the proof of a proposition cannot be transformed into a proof for that proposition, contradicting DNE. Generalizing the interpretation of the constants behind the rejection of these principles yields the Brouwer–Heyting–Kolmogorov (BHK) interpretation:

i. c is a proof of $A \wedge B$ iff c is a pair (c_1, c_2) such that c_1 is a proof of A and c_2 is a proof of B;

ii. c is a proof of $A \vee B$ iff c is a pair (c_1, c_2) such that c_1 is a proof of A or c_2 is a proof of B;

iii. c is a proof of $A \rightarrow B$ iff c is a construction that converts each proof d of A into a proof $c(d)$ of B;

iv. nothing is a proof of $\bot$;

v. c is a proof of $\exists x A(x)$ iff c is a pair (c_1, c_2) such that c_1 is a proof of $A(c_2)$;

vi. c is a proof of $\forall x A(x)$ iff c is a construction such that for each natural number n, $c(n)$ is a proof of $A(n)$.[65]

$\neg A$ is introduced by definition as $A \rightarrow \bot$. Hence c is a proof of $\neg A$ iff c is a construction which would convert a proof of A into a proof of something known to be unprovable. So a proof of $\neg\neg A$ would show how any construction which purported to convert a proof of A into a proof of something unprovable could itself be converted into a proof of something unprovable. This amounts to saying that A cannot be shown to be unprovable, which is clearly too weak to establish that A is provable, hence the failure of DNE.

In accordance with his view of logic as a subordinate activity, Brouwer did not himself pursue the axiomatization of a system concordant with his program. The first complete axiomatization of a logic meeting the constraints of the BHK interpretation was developed by Heyting (1956, 101f., citing his 1930).[66] It is this calculus which has been subsequently designated "intuitionistic logic" (henceforth **J**). If we temporarily disregard the variant interpretations given to the constants and atomic propositions of the two systems, we can observe that **J** is a proper subcalculus of **K**: All theorems and valid inferences of the former hold in the latter, but not vice versa. Indeed, we can see by application of the BHK interpretation that all of the axioms of a Hilbert-style presentation of **K** are preserved, except those which yield LEM (or equivalently those giving DNE, in this context $\neg\neg A \rightarrow A$), as are all of the operational rules of a natural-deduction presentation of **K**, except DNE. One consequence is that the connectives and quantifiers may not be interdefined in **J** as they are in **K**. In the natural-deduction presentation of **J** an additional rule of absurdity elimination, $\bot \Rightarrow A$, is introduced. Although the consensus is to regard this as justified by the BHK interpretation, some constructivists have demurred. Hence Johansson (1936) omits this rule from his system, yielding minimal logic, a proper subcalculus of **J**, which also satisfies the constructivist constraints. Some superintuitionistic subcalculi of **K** have also been promoted as formalizing constructive reasoning, but none of these systems has attracted the same degree of support as **J** (see van Dalen 1986, 275ff.). In sequent-calculus presentation the similarities of the constants of **J** and **K** are even clearer, since the difference between the two systems may be restricted to the understanding of the deducibility relation, which is constrained in Gentzen's (1935, 82) calculus **LJ** such that there may be at most one formula to the right of the turnstile.[67]

The other principal intuitionistic program is semantic antirealism.[68] This program has the same roots as the mathematical program, but diverges crucially from Brouwer by defending **J** as appropriate to a respectable meaning theory for language, rather than to the prelinguistic content of mathematics (Prawitz 1977, 5). This alternative focus on knowability, rather than the narrower notion of provability, makes the program more readily applicable to nonmathematical discourse.[69] The central line of argument behind the semantic antirealist adoption of **J** is to dispute the intelligibility of the classical inferential goal, that is, epistemically unconstrained truth. As Michael Dummett (1991, 316) would have it, the classical conception of truth is "a piece of mythology, fashioned, like the centaur, by gluing together incompatible features of actual things. It has all the properties of explicit knowledge, save only that it is not explicit." A brief outline of the support advanced for this claim might run as follows.

For the classicist, all propositions have truth values, including propositions whose truth values we are not in a position to ascertain. These so-called verification-transcendent propositions must be either true or false, even though there are no means of determining which. The crux is a demonstration of the untenability of this position: the manifestation argument.[70] This proceeds from the observations that understanding a proposition requires knowledge of its meaning, and that such understanding must be publicly manifestable as the recognition of whatever is constitutive of meaning. But the truth conditions of verification-transcendent propositions cannot be fully stated. Hence, if meaning is truth-conditional, the meaning of these propositions cannot be fully manifested, thus the propositions cannot be properly understood. Yet such propositions are not unintelligible, so meaning cannot be expressed in terms of classical truth.

Instead, Dummett promotes an alternative theory that reduces the meaning of terms to the conditions for their warranted assertibility. This permits an antirealist account of verification-transcendent propositions which does not forfeit their meaningfulness. In particular, it motivates the adoption of **J**, since that calculus preserves warranted assertibility—by reasoning parallel to that of the BHK interpretation—and is the most natural result of linking the meanings of the logical constants to their assertibility conditions. Alternatively, but to the same effect, the semantic antirealist program can be conceived of as retaining a truth-conditional account of meaning, but with a radically revised account of truth. Hence the antirealist argues that all truths are in principle knowable, whether by replacing the notion of truth with that of warranted assertibility or by subjecting it to epistemic constraint.

This is not the place for a thorough critique of semantic antirealism, but we note certain immediate lines of response. One important point is that it is the loss of the principle of bivalence, that all propositions are true or false, which underpins the semantic antirealist's logical revisionism. However, the manifestation argument is a challenge not to this principle, but to the thesis that truth may transcend knowability. Hence the revisionist argument overlooks a conceptually possible position—called Gödelian Optimism by Neil Tennant (1997, 159ff.)—of accepting the manifestation argument as justifying epistemic constraint, while retaining bivalence, and thereby **K**.[71] The Gödelian Optimist holds that truth is both knowable and bivalent: that is, that there are no classical truthmakers which may *in principle* transcend our ability to come to know the truth of the propositions they make true (and likewise for falsehood). Of course, if "in principle" is interpreted at all strictly, then this position clearly becomes untenable. Yet the intuitionist must also be on his guard against an unduly conservative reading of "in principle knowable" that would reduce his position to an unwelcome extremism, such as strict finitism, or even the contingency of mathematics.[72]

Furthermore, even if accepting epistemic constraint imposes a revised logic, perhaps that revised logic need not be **J**. Dummett himself (1976b, 83f.) once tentatively proposed a meaning theory grounded in falsification rather than verification. The propositions of this theory would respect a logic which was neither **K** nor **J**, but rather dual to **J**: DNE would be admissible, but double-negation introduction would not be, and so forth. However, this proposal accepts the revisionary force of the manifestation argument; it merely channels it in an unexpected direction. Yet it could be argued that Dummett's requirements for an acceptable meaning theory could be met by a theory that was independent of the choice of logic.[73] Such a theory might proceed by giving equal significance in the constitution of meaning to the consequences of assertion, as well as the warrant for assertion. Some step of this kind may well be required anyway, to accommodate empirical discourse, which offers independent motivation for this meaning theory.[74] Finally, a variety of arguments have been advanced that turn on the alleged proof-theoretic superiority of **J** to **K**. We return to this strategy in 2.1.3.

The two historically substantive programs just outlined do not exhaust the possible applications of **J** as a rough logic.[75] As an alternative, one might propose an application of **J** in which the propositions received their classical interpretations. Since the deducibility relation of **J** is a proper subrelation of that of **K**, in such a program **J** would be sound with respect to classical semantics, although (perhaps tolerably) incomplete. Something of this kind has been suggested as a response to the sorites paradox (Putnam 1983, 285f.). Although there has been some subsequent discussion, no fully articulated program has yet emerged.[76] In particular, although it is clear that intuitionistic semantics would be inappropriate, it is not clear what should be employed instead (Read and Wright 1985, 58; Putnam 1985, 203). Sketchy as this program is—and it may well remain so—it still serves to demonstrate that the formalism of **J** does not *in itself* necessitate the sweeping revisions generally promoted on its behalf. Although this shows that **J** could in principle be promoted within a logical theory which was otherwise substantially classical, in practice its adoption has been advocated as resulting from dramatic revisions of classical background theories.

2.1.2. How Are Intuitionistic and Classical Logic Related?

The closest relationship that can obtain between two logics is equivalence, but **J** is inequivalent to **K**. The two systems may be formulated with the same atomic propositions, the same constants (at least typographically), and therefore equiform classes of wffs and of sequents. However, the two classes of sequents would be partitioned into valid and invalid subclasses in a different fashion, hence **J** would appear to be nonconservatively revisionary of **K**. The only difficulty with this assessment is that there are several well-known ways of embedding **K** into **J**. Each of these approaches is a variation on the double-negation translation, which maps classical wffs to intuitionistic wffs in such a way that the validity of sequents in which the wffs occur is preserved, in a sense to be made precise shortly. The first such translation is due to Kolmogorov (1925, 428):[77]

$$\begin{aligned} A^* &= \neg\neg A, \text{ for atomic } A; \\ (\neg A)^* &= \neg A^*; \\ (A \wedge B)^* &= \neg\neg(A^* \wedge B^*); \\ (A \vee B)^* &= \neg\neg(A^* \vee B^*); \\ (A \rightarrow B)^* &= \neg\neg(A^* \rightarrow B^*); \\ (\exists x A)^* &= \neg\neg\exists x A^*; \\ (\forall x A)^* &= \neg\neg\forall x A^*. \end{aligned}$$

Then $\vdash_{\mathbf{J}} A$ iff $\vdash_{\mathbf{K}} A^*$. Alternative versions were produced independently of Kolmogorov, and of each other, by Gödel (1933a) and Gentzen (1933, 60f.).[78] Gödel's version runs as follows:

$$\begin{aligned} A^* &= \neg\neg A, \text{ for atomic } A; \\ (\neg A)^* &= \neg A^*; \\ (A \wedge B)^* &= A^* \wedge B^*; \\ (A \vee B)^* &= \neg(\neg A^* \wedge \neg B^*); \\ (A \rightarrow B)^* &= \neg(A^* \wedge \neg B^*); \\ (\exists x A)^* &= \neg\forall x \neg A^*; \\ (\forall x A)^* &= \forall x A^*. \end{aligned}$$

Gentzen's translation is identical to Gödel's except that he translates $A \rightarrow B$ as $A^* \rightarrow B^*$. In a related fashion, Gödel established a similar theorem and anti-theorem preserving translation of (propositional) **J** into the modal system **S4**. That is $\vdash_{\mathbf{J}} A$ iff $\vdash_{\mathbf{S4}} A^*$, where A^* is recursively defined as follows:

$$\begin{aligned} A^* &= A, \text{ where } A \text{ is atomic}; \\ (\neg A)^* &= \neg\Box A^*; \\ (A \wedge B)^* &= A^* \wedge B^*; \\ (A \vee B)^* &= \Box A^* \vee \Box B^*; \\ (A \rightarrow B)^* &= \Box A^* \rightarrow \Box B^*, \end{aligned}$$

or alternatively, as follows (Gödel 1933b, 301):[79]

$$\begin{aligned} A^* &= A, \text{ where } A \text{ is atomic}; \\ (\neg A)^* &= \Box\neg\Box A^*; \end{aligned}$$

$$(A \wedge B)^* = \Box A^* \wedge \Box B^*;$$
$$(A \vee B)^* = \Box A^* \vee \Box B^*;$$
$$(A \rightarrow B)^* = \Box A^* \rightarrow \Box B^*,$$

McKinsey and Tarski (1948, 13) also established this result for a simpler translation, which has become the most familiar of the three. We shall refer to this as the GMT translation; it proceeds as follows:[80]

$$A^* = \Box A, \text{ where } A \text{ is atomic;}$$
$$(\neg A)^* = \Box\neg A^*;$$
$$(A \wedge B)^* = (A^* \wedge B^*);$$
$$(A \vee B)^* = (A^* \vee B^*);$$
$$(A \rightarrow B)^* = \Box(A^* \rightarrow B^*).$$

All three translations may be straightforwardly extended to intuitionistic predicate logic and a quantified extension of **S4**.[81]

Might these translations be used to show that **J** could be presented as an extension of, and therefore not a rival to, **K**? If this were so, it would be either because **K** was equivalent to a proper reduct of **J**, by the double-negation translation, or because **J** was equivalent to an established extension of **K**, by one of the translations into **S4**.[82] In assessing this challenge, note that equivalence as introduced in 1.4 is a relationship on wffs requiring the preservation of inferences as well as theorems and invalidity as well as validity. Gödel's **S4** translations preserve only theorem-hood and antitheorem-hood, and are therefore insufficient for our purposes. The GMT translation can be shown to preserve deducibility as well (Epstein 1995, 289, contra Haack 1974, 97). However, it is a translation *into* **S4**: There is no corresponding map from **S4** to **J**. Hence **J** has not been shown to be equivalent to **S4**.

The more serious proposal is that a double-negation translation might establish that **K** is equivalent to a proper reduct of **J**. It can be shown that if $\Gamma \vdash B_1 \vee \cdots \vee B_n$ is valid in **K**, then $\Box\Gamma^* \vdash \neg(\neg B_1^* \wedge \cdots \wedge \neg B_n^*)$ is valid in $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$ and vice versa, where * is defined by one of the double-negation translations given above, Γ^* is the result of applying * to each $A \in \Gamma$, and $\Box\Gamma^*$ is the result of prefixing every member of Γ^* with $\Box$ (Gallier 1991, 74). $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$ must be a proper reduct of **J** because all of the constants of **J** are primitive, precluding the introduction of $\vee$ or $\exists$ by definitional equivalence (McKinsey 1939, 156f.). Is this relationship between **K** and $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$ an equivalence relationship? It maps the valid inferences of **K** to valid inferences of $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$; it maps the valid inferences of $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$ to valid inferences of **K**; it maps the invalid inferences of **K** to invalid inferences of $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$; but it provides no means of mapping the invalid inferences of $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$ to invalid inferences of **K**. All four mappings are required for equivalence. An identity function from the wffs of $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$ to the wffs of **K** will preserve validity but not invalidity, because weak counterexamples such as Peirce's law or DNE, which are valid in

K but invalid in **J** (and a fortiori in $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$), will be translated into their valid counterparts in **K**. It seems unlikely, although conceivable, that *any* mapping sufficiently ingenious to preserve both validity and invalidity could be found. Moreover, it can be shown that the double-negation translations do not preserve any of the presently available semantics for **J**, so any such proposal would also require (perhaps unattainable) semantic innovation (Epstein 1995, 396). The underlying problem is that the double-negation translations define embeddings of **K** in $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$: The system into which **K** is translated is a proper subsystem of $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$. Establishing the equivalence of **K** to this subsystem would not show that **J** extended **K**, but (unremarkably) that **J** extended a system nonconservatively revisionary of **K**.[83] We must conclude that **J** is neither equivalent to **K** nor an extension of **K**, and therefore that it is nonconservatively revisionary of **K**.

The next major question about how **J** is related to **K** is whether **J** recaptures **K**. In formal terms this is easy to answer. The class of wffs generated from effectively decidable atomic formulae will behave classically under closure by the constants of **J** (Dummett 1959b, 167). That is, the class of wffs such that $\vdash A \vee \neg A$, for all atomic propositions A, and $\vdash A(t) \vee \neg A(t)$, for all atomic predicates A and terms t, form a system equivalent to **K**: **J** recaptures **K**. However, this does not address the "politics" of recapture. In 1.4 we identified a spectrum of responses to the possibility of classical recapture by a nonclassical logic with a spectrum of political positions. Superficially, the constructivist and semantic antirealist programs in which **J** is characteristically deployed are clear examples of the "left-wing" response: The possibility of recapture is denied or rejected as irrelevant. The radical-left strategy of ensuring that recapture does not work is unavailable without revising **J**, since **J** recaptures **K**, so both programs must be on the center left. However, this assessment is somewhat overhasty: It is possible to make more productive use of the recapture result. Proponents of both intuitionistic programs do sometimes describe **K** as unintelligible: For instance, Dummett remarks that "intuitionists ... deny that the [classical] use [of the logical constants] is coherent at all" (Dummett 1973c, 398).[84] This would suggest hostility to classical recapture. Yet although this hostility may be maintained by some intuitionists, in general the situation is more eirenic. In both programs it is generally conceded that there is a domain of propositions for which **K** is applicable (for example: Brouwer 1952, 141; Dummett 1959b, 167; 1973b, 238). Indeed, Dummett suggests that this "common ground" is sufficient for the intuitionist to gain an understanding of the classical meaning of other, disputed formulae which, although not "accepted as legitimate," is at least "not wholly opaque" (Dummett 1973b, 238). This would suggest that remarks inimical to recapture should be taken as hyperbole, leaving open the possibility of a center-right attitude.

There are several reasons why the intuitionist should welcome recapture, but are they enough for a center-right attitude? For a long time intuitionists were obliged to appeal to **K** to prove results in the metalogic of **J**, such as the completeness of the first-order system. Until intuitionistically acceptable proofs were produced (Veldman 1976; de Swart 1976), this provoked the classical criticism that the intuitionist was indulging in a practice that he wished to deny to others (Tennant 1997, 305f.).[85] Such criticism has sufficient rhetorical force to make the intuitionist's position appear exposed, but in principle he is on perfectly safe ground, providing that all of his employment of strictly classical inference occurs within a decidable domain. Even now that an intuitionistic metalogic is practicable, a case may be made that the intuitionist should retain a classical metalanguage, at least as an alternative to the intuitionistic version. For, as Dummett points out, insistence on the employment of the logic of a reform proposal throughout the metalanguage serves to insulate the proposal from criticism, and at the cost of handicapping its ability to persuade the practitioners of other systems of its merits (Dummett 1991, 55).

Here we should be careful to distinguish the practical claim, that the classicist will be more readily convinced by metalogical argument in classical terms, from the stronger methodological claim, that some specific system (perhaps **K**) must be employed in the metalogic for the constants to be properly interpreted.[86] The practical claim merely asserts the persuasive value in "preach[ing] to the Gentiles in their own tongue," as Meyer (1985, 1) describes the analogous enterprise in relevance logic. The Gentiles should not really need a translation in this instance, since the deducibility relation of **J** is a subrelation of that of **K**, which ensures that all intuitionistically valid proofs are classically valid too. Dummett (1991, 55) wishes to maintain the stronger claim, and argues that the metalogic should be as neutral as possible. (However, this eventually turns out to be a neutrality distinctly friendly to **J**, to paraphrase Dummett's [1973a, 603] characterization of a rather different claim to neutrality.) The crucial difference between the two claims, which Tennant (1997, 305) accuses Dummett of having missed, is that the former cannot ground the latter without opening **K**, as much as any other system, to the accusation that it is seeking to resist criticism through question-begging self-justification.

The foregoing argument is reprised in the analysis of the constants employed in the BHK interpretation: Unless they are understood classically, the interpretation cannot explain intuitionistic usage to the classicist (Makinson 1973, 77). Fortunately, the domain in which the interpretation is carried out is effectively decidable, and thereby recaptured in **J**. In addition, the Brouwerian account of logic as subordinate to mathematics should be seen as favorable toward classical recapture. If logic is merely the a posteriori codification of valid modes of mathematical reasoning, there can be no objection to some aspects of this reasoning fitting more than one codification (see Heyting 1956, 74). This, with the points made before, motivates the retention of **K** as a limit case of **J**, that is, center-right recapture. However, against this suggestion it should be recalled that center-right recapture would require the intelligibility of the inferential goal of **K**—epistemically unconstrained truth—within the theory of **J**. At least some proponents of **J** would regard this as unsustainable, relegating **J** to center-left classical recapture.

Conversely, it might be possible to move even further to the right, at least within the constructivist program. Most constructivists have followed Brouwer in holding that classical mathematical results remain unjustified until a constructive proof is forthcoming. However, there is an alternative tradition in which these results are regarded as having their own, weaker, sort of legitimacy. Hence Kolmogorov (1925, 431) argues that we should "retain the usual development" of what he calls "pseudomathematics" alongside the development of constructive mathematics, since he suggests that it is at least consistent intuitionistically.[87] Kolmogorov's approach has much closer affinities to the formalism of Hilbert's program than has Brouwer's: Whereas Brouwer seeks to partially license infinitistic material independently, both Hilbert and Kolmogorov seek fully to ground it in finite mathematics.[88] On Hilbert's account, "real" mathematics is restricted to finitistic results; the remainder, "ideal" mathematics, can still be a useful heuristic for finite results, providing that its relative consistency can be established. Kolmogorov (1925, 417, 431) argues that his program has a twofold advantage over Hilbert's: The finite basis is grounded in construction, not just consistency, thereby answering any charge of arbitrariness; and the existence of the double-negation translation of **K** into **J** offers a ready means for a relative consistency proof.

Following Kolmogorov's insight, one might regard a proper subsystem of $\mathbf{J}_{\neg,\rightarrow,\vee,\forall}$ as the logic of pseudomathematics—providing that such a system equivalent to **K** could be demonstrated, although it is not clear whether this is feasible. The logic of real mathematics would then be the stricter system resulting from an extension by independent, constructive, notions of disjunction and existential quantification: **J**. This hypothetical program would thus exhibit the "reactionary" response to recapture. However, it remains strictly hypothetical: Not only does it rely on an equivalence relation that we have no reason to believe obtains, it would also require an argument that disjunction and existential quantification are not intersystemically invariant (that is, cannot be identified) between **K** and **J**. There may be some justification for the latter point: Because the focus to the constructivist's challenge to classical mathematics is existence, it is understandable that he might have objections to the elimination rules for disjunction and existential quantification. However, we saw that these rules are retained in their classical form in axiomatic and natural deduction presentations of **J**: in both cases the revision appears to be of negation, which on this hypothetical program would be untouched. Furthermore, intuitionistic criticism of the elimination rule for disjunction would seem readily to generalize to *reductio ad absurdum*, even of the intuitionistically acceptable variety.[89] Finally, one might abandon **J** as such, and pursue issues in constructive mathematics in a version of **K**, extended either by a modal constant, or by additional constants for constructive disjunction and existential quantification.[90] This would be a clear-cut case of reactionary recapture, in which the priority of **K** would be wholly unchallenged.

2.1.3. The Significance of Proof Theory

If a formal system is to be promoted as a rough logical theory, and thereby as an organon, it must be provided with a suitable semantics and proof theory. This makes these aspects of the theory targets for critics of the enterprise, since if they are unequal to their task the theory will be blocked. Conversely, the advocate of a nonclassical program has much to gain by finding fault with classical semantics or proof theory. We briefly addressed the significance of the semantic interpretation of **J** in the last section; we return to this line of argument in discussing relevance logic, in 2.3, where the issue has been a much greater focus of contention. However, in the advocacy of **J** rather more attention has been paid to the role of proof theory. Whereas in semantics, a formal system either has or does not have a plausible interpretation, without which it cannot be readily promoted as an organon, in proof theory a wide variety of desiderata have been canvassed as hallmarks of good logical practice, engendering considerable complication. In particular, we must be careful to distinguish between those proof-theoretic properties which serve a practical, but dispensable, purpose—such as enhancing the ease of use of the system, or permitting a greater faithfulness to natural argumentation—and those properties which are claimed to be indispensable to the employment of any coherent system.

Many different proof-theoretic properties have been suggested as important for either or both of these purposes: Tennant (1996, 354f.) lists 14 different suggestions, without exhausting all possibilities.[91] Some of these serve only the former, practical purpose, such as the requirement that proofs have a "nice mereology." Others, such as "preservation of preferred species of truth" and "relevance by restricted transitivity of deduction," respectively, are either clearly satisfied by **K**,[92] or clearly not satisfied by **J**. Either way these properties do not discriminate in favor of **J**. Of the potentially decisive properties, the most frequently invoked are separability, inversion, normalizability, and harmony.

A system is *separable* if the operational rules for each constant contain no other constants, and every wff is derivable iff it is also derivable in a system in which the only operational rules are those for the constants contained by that wff (Ungar 1992, 7 n8). Hence, in the terminology of 1.4, a system will be separable if each of its proper reducts is equivalent to the system generated by the rules expressible in that reduct. The *inversion* principle requires that each elimination rule relates to the corresponding introduction rule as the inversion of a function relates to that function, "in the sense that a proof of the conclusion of an elimination is, roughly speaking, already available if the premiss of the elimination is inferred by an introduction" (Prawitz 1981, 242). So if the inversion principle applies, whenever the premisses of an elimination rule are obtained by application of the corresponding introduction rule, the conclusion of the elimination rule could have been obtained at an earlier stage in the proof. This gives rise to *reduction procedures* for the constants, whereby a passage of a proof in which a wff occurs as both the conclusion of an application of the introduction rule and a premiss of an application of the elimination rule may be eliminated. If no such passages occur in a proof then it is in *normal form*. *Normalizability* requires that all proofs can be placed in normal form. On certain additional assumptions the reduction procedures will then serve as an equivalence relation on proofs, whereby two proofs which reduce to the same normal form are equivalent (Ungar 1992, 155f.). Finally, a constant is in *harmony* if (1) the conclusion of its introduction rule is the strongest wff so derivable which may be eliminated by the elimination rules (where one wff is stronger than another if the latter may be derived from the former); (2) the major premiss of its elimination rule is the weakest wff licensing the derivation which may be introduced by the introduction rules; and (3) (1) and (2) can be established using precisely the constant's elimination and introduction rules, respectively (Tennant 1997, 321, simplifying somewhat).

The practical utility of these properties is not in doubt. Separability permits constants to be studied in isolation; normalizability assembles proofs into equivalence classes, and so forth. But this does not show that a system lacking these properties would be incoherent, and not just inconvenient. Harmony will be required by any proof-theoretic theory of meaning, to ensure that the warrant granted by the assertion of a wff does not exceed the warrant for that assertion. But relativizing the requirement to such a theory of meaning would be to beg the question; once again this is to shift the debate onto the choice of inferential goal. Conversely, one might imagine that separability should be inimical to any sufficiently holistic theory of meaning. Harmony may be employed to block the admission of mischievous constants, such as Prior's (1960) tonk, but it is not the only way this may be achieved.[93] Nevertheless, some requirements along these lines would seem reasonable constraints on any plausible proof theory. However, we have not yet seen that **J** is better placed than **K**. Each of the four properties of separability, inversion, normalizability and harmony is a necessary but insufficient requirement for the next on the list (Tennant 1996, 358; 1997, 314). So if the intuitionist could show that **K** is not separable he would have a powerful argument against its cogency as an organon; conversely, if the classicist can establish this property he is well placed to begin recovering the others.

It is well known that separability fails for the usual natural-deduction presentations of **K**. Peirce's law, $((A \to B) \to A) \to A$, is a theorem of **K** but cannot be proved solely from the natural deduction rules for $\to$. However, it is also well known that separability holds for most other presentations of **K**, notably the multiple-conclusion sequent calculus (Read 1995, 229). The intuitionistic response to this move is that multiple-conclusion systems are unacceptably classical because they involve sequents that cannot be given a sufficiently constructive interpretation (see, for example, Tennant 1997, 320). The classical understanding of $\Gamma \vdash B_1, \ldots, B_n$ is that the commas to the right of the turnstile function as implicit disjunctions. But for the derivation of such a disjunction from Γ to satisfy the BHK interpretation (at least in cases where Γ contains only nondisjunctive propositions), a derivation of a specific disjunct from Γ must exist. This need not be the case here: The multiple-conclusion sequent calculus for **K** validates inferences which do not meet this constraint. There are two natural responses to this argument. First, intuitionistic squeamishness about multiple conclusions seems misplaced, since although Gentzen characterized the difference between the sequent calculi for **J** and **K** as a restriction of the former to single conclusions, the minimum necessary constraint on the multiple-conclusion presentation of **K** required to yield a presentation of **J** is much more modest. All that is required is that applications of the right-hand introduction rules for $\rightarrow$ and $\forall$ (and $\neg$, if negation is taken as primitive) be restricted to situations in which there is only one wff on the right-hand side of the concluding sequent.[94] Thus there is no proof-theoretic objection to multiple-conclusion presentations of **J**. Indeed, there are such systems,[95] and they can be shown to be sound and complete with respect to the standard Kripke semantics for **J**, so it cannot readily be argued that they lack an interpretation.

Second, and more important, this intuitionistic complaint misses the point. The original claim was that separability was a general proof-theoretic property, exhibited by any reasonable system, but failing for **K**. We have seen that **K** has this property in multiple-conclusion presentation. Even if the presentation was intuitionistically unacceptable, the most that would be established is that separability fails for **K**, *if intuitionism is right*. How could the classicist be moved by such a conclusion? Less polemically, the intuitionist's argument rests on the BHK interpretation of disjunction, and thereby on a constructive account of truth. Once more the debate has been shifted to the choice of inferential goal.

What of the other proof-theoretic desiderata? In their standard formulations, inversion, normalizability, and harmony all fail for **K**. However, in a similar vein to the defense of classical separability, arguments have been produced to show that intuitively plausible analogs hold for some presentations of **K** (and indeed sometimes fail for **J**).[96] In each case a similar intuitionistic retort could be made, that nonconstructivist principles have been invoked.[97] But by the same token this would be question-begging unless buttressed by independent argument for the adoption of the constructive account of truth. Again the focus of the argument would be shifted from comparison of the formal systems to choice of inferential goal.

So, in practice, considerations of proof theory fail to shift the debate from a conflict within the background theories as to the inferential goal best fitted to the understanding of natural argumentation to a conflict between formal systems over the formalization of that argumentation. This is the character that one would expect revisionism to exhibit in an heuristic context focused on the subject matter of logic.

2.1.4. The Character of Intuitionistic Revisionism

So far we have primarily been concerned with formal aspects of the advocacy of **J**: syntax, semantics, and proof theory. However, we saw in 1.1 that research programs for rough logics must contain additional features: a parsing theory, an inferential goal, and background theories. As we shall see in the remainder of this chapter, the advocates of most nonclassical logics wish to retain broadly classical background theories. Hence they seek to modify the inferential goal as little as possible, and to revise the formal system in such a way as to permit a more natural and transparent parsing theory. We have shown that the advocacy of **J** is a very different enterprise. Both the mathematical constructivist's and the semantic antirealist's programs are motivated by a substantial revision of the background theory, which in both cases induces a strongly nonclassical inferential goal. Hence the former wishes to stipulate in his background theory that mathematics be constructive rather than classical, and therefore requires a logic that pursues proof rather than truth; and the latter insists in his background theory that the antirealist theory of meaning is the only coherent option, and therefore requires a logic that pursues warranted assertibility rather than epistemically unconstrained truth. In both cases the change of inferential goal can be represented as substituting something else for (classical) truth, or as offering a nonclassical account of truth, but this is an essentially terminological distinction: Either way, the inferential goal has been substantially revised. Such fundamental revisions will in turn affect the choice of parsing theory—if the formal system is designed to respect a different principle, natural argumentation will have to be cashed out in different terms. However, in contrast with other nonclassical programs, this change is of no special importance to the overall revision, and is not intended to achieve any particular gain of transparency or simplicity.

In 2.1.1 we demonstrated that the standard arguments for intuitionistic revisionism strongly conform with this picture. In both cases the argument originates outside the domain of logic: The constructivist wishes to challenge classical mathematics; the antirealist wishes to challenge the realist theory of meaning. Hence the revision can be placed in the final level of the hierarchy of revision sketched in 1.6: "change of subject matter."[98] A characteristic feature of this species of revisionism is that the positive heuristic, which dictates the methodology of the ongoing logical research program, is focused more specifically on a revision of the background, and less on the details of the preferred system, than is the case with more modest revisions.

In 1.3 we stressed the importance of distinguishing between differently focused programs, or different stages in the development of a program. Our concern here is with the intuitionistic program at the point of its divergence from the classical: an ongoing schema for logical development, rather than the sort of completed organon by which the salient motivating background theory might be furthered—if that is even attainable. This schema can be conceived of either historically, as (close to) the earliest stage of the intuitionistic program at which it is properly distinguishable from the classical program; or conceptually as (close to) the initial revision of the latter-day, more compelling, classical program.

Several points can be advanced in favor of this analysis of the intuitionistic programs. Within the constructivist program we have seen that there has been considerable promotion of a conception of logic as subordinate to mathematics. This has resulted in toleration of disputes as to which logic is most appropriate for the success of the program (Heyting 1956, 74, and 2.1.1). Within the antirealist program it has been argued that the program could be conducted without the adoption of nonclassical logic.[99] This implies that the adoption of **J** is not required for continuation of the antirealist program, and thereby that the choice of logic is not part of the indispensable hard core of that program. Incidentally, this version of the antirealist program, and the dual suggestion at the end of 2.1.1, which combined **J** with a classical background, *would* confront **J** directly with **K**. However, this direct dispute between the formal systems would be fomented only by the counterfactual expedient of employing one or other system in an unfamiliar program. Finally, arguments have been advanced that attempt to concentrate the dispute between **J** and **K** within the domain of logic. However, we saw in the last section that these arguments invariably require the invocation of assumptions from the background theory to have any prospect of success. Try as we might, the dispute between **J** and **K** keeps returning to the choice of inferential goal, and thereby to the content of the background theory. This would be surprising if the two systems were rival formalizations of a common inferential practice, as many other disputes might be characterized. In this case it serves to reinforce an analysis of the dispute as intrinsically extralogical.

Where two logical research programs differ in inferential goal it is reasonable to ask whether either goal might be represented within the other system. We have seen how this might be achieved for **K** and **J**, through extension by a modal constant of provability (or "ancillary" use of constructive constants) and by classical recapture, respectively. If the difference of goal was the most fundamental difference between these two programs, such a strategy would be sufficient to effect a reconciliation. If systems from both programs could wholeheartedly reproduce the inferential practices of the other program, it would be straightforward to find bridge laws between the two salient systems, making the choice of program little more than conventional. However, we have had little success in pursuit of this aim. We showed in 2.1.2 that $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$ cannot be equivalent to **K**, despite an initial impression to the contrary. Conversely, it is highly unlikely that an extension of **K** would be intuitionistically acceptable. If the relationship between the two programs was asymmetric, such that one program could reproduce the inferential practice of the other, but not vice versa, this could be regarded as an impressive feat of Lakatosian monster-exploiting by the more successful program. It could be argued that the GMT translation of **J** into **S4** shows that the classical program, of which **S4** is a part, has achieved this feat. However, although this move makes the intuitionistic program intelligible to the classicist, it is difficult to see it as doing justice to that program. In particular it would ignore the intuitionist's criticism of the classical principles which underpin **S4** as much as they do **K**.

The underlying obstacle to both of these attempts to defuse the dispute through a reductive analysis of the intuitionistic program is that they do not take account of the change of background theory which is intrinsic to the conflict. Any viable attempt at reconciling the classical and intuitionistic programs must also reconcile their background theories. This is not facilitated by the presence of flat contradictions of familiar aspects of the classical background within the hard core of both intuitionistic programs. There is still some scope for maneuver since, unlike the intuitionistic programs, the classical program need not be construed as placing its background theories within the irrevisable positive heuristic. Conversely, the intuitionistic programs, unlike their classical counterpart, do not require that the formal system be irrevisable. Hence it may be possible to retain the irrevisable components of both programs, by pursuing **K** within an antirealist program (Wright 1982, 468ff.). This effects a reconciliation, but at the expense of abandoning **J** altogether.

2.2. Quantum Logic

The promise held out by the quantum-logical program is that by employing a novel logic derived from the mathematics of quantum mechanics (QM) we may resist the counterintuitive metaphysical consequences normally associated with the adoption of this physical theory. Our chief concern in discussing the program is not so much its success or failure as its conceptual viability. Is the proposed move a true revision of logic or not? First we discuss the background to quantum logic, and introduce a specific formal system, **QL**.[100]

2.2.1. What Is Quantum Logic?

The logical system we shall be concerned with was first proposed by John von Neumann in 1932.[101] In collaboration with Garrett Birkhoff in 1936 he showed how a formal system could be derived from the mathematics of QM, but this work was not pursued further for at least another 20 years. QM is concerned with certain measurable properties—observables—such as position, momentum, and spin, which can be given a numerical value by experiment. A quantum mechanical system, $\mathfrak{S}$, consisting of one or more particles, has a full description in its state, which is given by a wave function $\Psi(\mathbf{r}_i, t)$ where $\mathbf{r}_i$ are the positions of the particles and t is the time. The solution space of the wave function is the Hilbert space $H(\mathfrak{S})$.[102] Although the wave function itself is unobservable, observables are represented by self-adjoint operators on the wave function. The range of each of these operators is a subspace of $H(\mathfrak{S})$, that is, a topologically closed set of the vectors of $H(\mathfrak{S})$, one which is closed under addition of vectors and multiplication by scalars. Hence these subspaces yield $H(\mathfrak{S})$ when taken together.

Thus Birkhoff and von Neumann were able to observe that there is a one-to-one correspondence between (true) elementary propositions of $\mathfrak{S}$, $U(m, r, t)$, which attribute the value r to some measurable physical magnitude m at time t, and these subspaces of $H(\mathfrak{S})$. Hence, U is true iff the subspace to which it corresponds, $h(U)$, is a subspace of $H(\mathfrak{S})$; $\models U$ is valid iff $h(U) = H(\mathfrak{S})$, and U semantically entails V iff $h(U) \subseteq h(V)$. Birkhoff and von Neumann proceed to show that the subspaces of $H(\mathfrak{S})$ may be arranged in a lattice, $L(\mathfrak{S})$, by employment of set-theoretic operations.[103] Set-theoretic inclusion, $\subseteq$, is reflexive, transitive, and antisymmetric, and may therefore serve as a partial ordering, $\leq$, on $H(\mathfrak{S})$. The intersection of two subspaces, $h(U) \cap h(V)$, is itself a subspace, and represents their greatest lower bound. Although the union of two subspaces is not a subspace, we can use a similar operation, the linear union of two subspaces, $h(U) \oplus h(V)$, which results in the space spanned by the union set of both their basis vectors. This is the smallest subspace of $H(\mathfrak{S})$ containing both $h(U)$ and $h(V)$, and therefore their least upper bound. Hence $L(\mathfrak{S})$ is a lattice.

In addition, Birkhoff and von Neumann demonstrate that $L(\mathfrak{S})$ is orthocomplemented. Orthocomplemented lattices have a greatest or unit element, $\top$, a least or zero element, $\bot$, and every element a has an orthocomplement $a^{\bot}$, such that $a^{\bot\bot} = a$; the least upper bound of a and $a^{\bot}$ is $\top$ and their greatest lower bound is $\bot$. $H(\mathfrak{S})$ itself contains all its subspaces (obviously) and thus corresponds to $\top$. The null-space $\mathbf{0}$, which contains only the null-vector, is a subspace of all Hilbert (sub)spaces and may therefore serve as $\bot$. The set-theoretic complement of a subspace is not itself a subspace, but again we can use an analogous operation: The orthogonal complement of a subspace, $h(U)^{\bot}$, is the subspace consisting of the vectors orthogonal to the elements of $h(U)$. (Two vectors are orthogonal if their inner product is the null vector.) Hence $L(\mathfrak{S})$ is an orthocomplemented lattice, or ortholattice. However, it is with the substitution of orthocomplementation for set-theoretic complementation that we have made our greatest departure yet from the orthodoxies of classical set theory, and indirectly, classical mechanics and classical logic. Not only do complementation and orthocomplementation diverge sharply in their results, but in orthogonality we have introduced an element alien to set theory.

The ortholattice $L(\mathfrak{S})$ may be taken as the algebraic presentation of a logic, **QL**. Hence the correspondence between the propositions of $\mathfrak{S}$, U, and the subspaces of $H(\mathfrak{S})$, $h(U)$, can be extended by identifying logical constants with features of the lattice of subspaces as follows: $\neg U$ is defined as the proposition V such that $h(V) = h(U)^{\bot}$; $U \wedge V$ is defined as W such that $h(W) = h(U) \cap h(V)$; $U \vee V$ is defined as W such that $h(W) = h(U) \oplus h(V)$; $U \supset V \equiv_{\text{def}} \neg U \vee (U \wedge V)$;[104] quantifiers are introduced by analogy with $\wedge$ and $\vee$. The logic thus defined diverges from **K**, most notably in disjunction. Pertinently, the distributive law, $A \wedge (B \vee C) \dashv\vdash (A \wedge B) \vee (A \wedge C)$, fails where the dimension of $H(\mathfrak{S})$ is greater than 1, as it is in all practical cases. (More fundamentally, whereas the Lindenbaum algebra of **K** is Boolean, that of **QL** is a partial Boolean algebra (Bub 1991, 27)—that is, a system of Boolean algebrae overlapping in a certain way—and is not embeddable into any Boolean algebra (Kochen and Specker 1967). Failure of distributivity is not necessarily the most acute account of the divergence of **QL** from **K**. Indeed, on a radical interpretation (Bub 1989, 202), defining validity over partial Boolean algebrae rather than Boolean algebrae, distributivity would be valid in **QL**.)

Birkhoff and von Neumann do not propose **QL** as an explicit revision of **K**, let alone as a resolution of the anomalies of QM. An argument for the latter position was subsequently advanced by Finkelstein (1969, 204ff.), and used by Hilary Putnam to motivate his revisionist program.[105] This approach is the most philosophically conspicuous defense of **QL**, but it has been shown to be substantially flawed (see Gardner 1971; Gibbins and Pearson 1981; and Redhead 1994, 167f. for one important line of criticism, or Gibbins 1987, 148f.; Sklar 1992, 200 for another). However, the success of this application of **QL** is independent of the program's philosophical viability. It is this viability, not whether **QL** may serve to resolve the anomalies of QM, with which we are principally concerned. It is plausible to suppose that these are also Putnam's priorities, since his real agenda is to show that his rejection of a priori knowledge extends to logic (see Putnam 1975, x). This explains why **QL** is ignored in (Putnam 1965), a paper on QM, despite being discussed in (Putnam 1962), a paper on epistemology, and why Putnam has been such a fair-weather friend to the quantum logical program: He needs to show the revisability of logic on empirical grounds, he does not need it to be actually revised. Hence the philosophical viability of the quantum logical program is enough to achieve Putnam's purposes, even if the program does not succeed on its own terms.

However, historically, it has been the promise of a realist understanding of QM that has made **QL** most attractive, not the promise that the paradoxes of quantum mechanics would dissipate if addressed "quantum-logically." And this approach still holds the most promise for the future of the program (see Dickson 2001 for an up-to-date defense). It is a notorious feature of QM that some propositions are complementary, or incompatible with each other. For instance, it may be possible to fully determine either the position or the momentum of a particle, but they cannot be determined simultaneously. Employment of **QL** maintains this feature because the subspace which represents the conjunction of a proposition stating the position of a particle with a proposition stating the momentum of that particle is zero dimensional, hence the conjunction is logically false. Thus either proposition may be true, but their conjunction must be false, as we would expect, since it corresponds to an observation we cannot perform.

2.2.2. Is Quantum Logic Compatible with Realism?

Is the combination of the **QL** program with a realist metaphysics tenable? To what extent is the quantum logician committed to this combination? For example, Putnam (1994, 276) renders the true proposition that an observable has a value by the disjunction $q_1 \vee q_2 \vee \cdots \vee q_N$, where each q_i attributes a different value to the observable, ensuring that there is some j for which q_j is true. However which j may only be clear *sub specie æternitatis*. So far, so nonclassical: The tension with realism arises when we attempt to provide the nondistributive calculus of **QL** with a semantics. The Kochen–Specker (1967) argument shows that no such semantics can satisfy the realist "admissibility criterion" that a truth valuation will only be admissible if it maps propositions onto the two element Boolean algebra of 0 and 1. Because **QL** is explicitly characterized by its non-Boolean structure, this "criterion" may look like an instance of the reprehensible strategy of attempting to discredit a revisionist proposal by assuming a contested principle in the metalanguage.

However, as Putnam has more recently argued, following a suggestion from Dummett, the admissibility criterion is necessary if we are to be able—even if only in our imagination—fully to visualize the quantum state of affairs.

> [As a metaphysical realist] whenever I guess that a disjunction is true, I must guess that a disjunct—a *specified* disjunct—is true. Whenever I guess that a statement is true, I must guess that its negation is false. If I guess that a conjunction is true, I must guess that every conjunct is true, and if I guess that two compatible propositions are true, I must guess that their conjunction is true. And, since $S \vee \neg S$ is a tautology in quantum logic, I must guess that one of each pair of propositions of the form S, $\neg S$ is true. But now, even if the world somehow does not obey Boolean logic, my guesses will certainly do so. (Putnam 1994, 279)[106]

Dummett's point is that the realist stance obliges us to adopt a Boolean algebra at least for our "guesses" about the truth values of propositions. For, if we believe, as realists, that every proposition of QM has a determinate (if perhaps unverifiable) truth value, then it should not be impossible, but merely staggeringly unlikely, that we should correctly *guess* the truth value of every such proposition. However our realism would constrain these guesses. Hence, if we guess that some disjunction is true, for instance, we must also guess that at least one *specific* disjunct is true, to maintain our hypothesis of the determinate truth value of QM propositions. But this means that realism would entail that our guesses formed a two-element Boolean algebra. So our guesses would comprise a mapping from **QL** to such an algebra, which is impossible. Hence the combination of a realist metaphysics with a non-Boolean metalanguage would oblige us to deny that we could even imaginatively fully visualize the world our metaphysics hypothesized. This would render us Boolean creatures in a (to us necessarily ineffable) non-Boolean world. (The later) Putnam takes this to be a *reductio* of the **QL** program.

In outline, the Kochen–Specker argument shows that **QL** cannot be given a Boolean semantics, and the "guessing" argument shows that this makes **QL** incompatible with realism. Several lines of response to this impasse have been advanced. The Kochen–Specker argument depends on a constraint on value assignments, the functional composition principle, which states that the operators of QM and the values possessed by the corresponding observables have a common algebraic structure. This principle depends on three assumptions (Redhead 1987, 133): (1) the so-called realist assumption that all observables have sharp values in all states; (2) a one-one correspondence between operators and observables; and (3) the existence of an observable possessing and measured by a given value for every operator yielding that value. If any of these assumptions is abandoned, then the Kochen–Specker argument will be blocked.

If the "realist" assumption (1) is dropped, the Kochen–Specker argument is blocked by relating the value of the observable to the context in which it is measured. This leads in the direction of the Copenhagen interpretation of QM, and away from the chief selling point of **QL**, the retention of our "common sense" metaphysical intuitions. If **QL** is to be promoted as a revisionary program, rather than a practically convenient calculus, any response to the Kochen–Specker argument which yields the Copenhagen interpretation must be rejected. However, that is not to say that dropping assumption (1) is in itself irreconcilable with our intuitions.

Dropping assumption (2) has been suggested by Bas van Fraassen (1973, cited in Redhead 1987, 134f.). This results in many different observables corresponding to each nonmaximal operator. (An operator is maximal if it corresponds to a complete set of commuting observables. Thus an operator yielding both the magnitude of the momentum of a particle and one of the momentum's Cartesian components is maximal, whereas an operator yielding only one of these values is nonmaximal.) Each of these "split" observables is identified by its relationship to a different maximal operator. Since the Kochen–Specker argument cannot be derived from consideration of maximal operators alone, it must be blocked by this splitting of observables (Redhead 1987, 134, citing Maczynski 1971). As a cautionary consideration, it has been demonstrated that this position entails accepting some form of nonlocality, and thereby perhaps sacrificing one of our common sense intuitions (Redhead 1987, 139ff., citing Heywood and Redhead 1983). Yet this falls far short of a demonstration that the main freight of these intuitions is incompatible with **QL**.

Arthur Fine (1974, 264) proposes that we drop assumption (3), in which case there would be a unique observable corresponding to every nonmaximal operator, but the measurement procedure associated with that operator would not necessarily yield the correct value of the observable. Redhead (1987, 135f.) complains that this scheme does not offer any explanation of which measurements do in fact yield values obtaining in the world. However, this would seem to misread Fine's strategy, which is to deny the need to talk in terms of "real," "possessed" values.

The suspicion addressed in this section was that the quantum logical program may be fundamentally incoherent, because inescapable features of **QL** were incompatible with the assumption of realism in the hard core of its philosophical background. However, we have shown that there are at least two promising strategies for defusing the Kochen–Specker argument without abandoning realism. This blocks the conclusion of the "guessing" argument, that for a realist the shift to the quantum logical program would render the reality of the world ineffable. These methods may have difficulties of their own, but the combination of realism and **QL** is clearly not inherently unstable.

An alternative would be to concede the ineffability of the world, disputing whether this is untenable, and whether it is incompatible with realism. Properly understood, the assumption of realism in the philosophical background of a quantum logical theory does not make **K** the only acceptable calculus, which suggests that the metaphysical indebtedness of **QL** is not as great as suspected.

2.2.3. (2*b*) or Not (2*b*)?

We shall now turn to a more familiar critical strategy. In his criticism of Putnam's advocacy of quantum logic, Dummett (1976a, 285) characterizes the possibilities for logical revision as follows:

> Let us assume ... a revision from classical to some non-standard logic: let us call their advocates C and N. Then there are four possible cases according to which of the following two pairs of alternatives hold. (1) N rejects the classical meanings of the logical constants and proposes modified ones; or (2) N admits the classical meanings as intelligible, but proposes modified ones as more, or at least equally, interesting. And (a) C rejects N's modified meanings as illegitimate or unintelligible; or (b) he admits them as intelligible, alongside the unmodified classical meanings. If cases (2) and (b) both hold, then we are in effect in a position in which only relabeling is involved.

"Relabeling" is defined by Dummett (ibid.) as a merely terminological change, such that although we may relinquish some *sentences*, or accept other, previously rejected sentences, we do not change our attitude to any *propositions*. Such a change would be on a par with translation; we wouldn't expect the German edition of a logic textbook to describe different systems of logic from its English counterpart—although the sentences would be different—because we would hope that the same propositions were expressed.

As Dummett notes, intuitionistic logic satisfies (1), since its proponents affect to find **K** unintelligible. (Interestingly, he doesn't ask whether it falls into [1a] or [1b]. We saw in 2.1.2 that there is a well-known translation of **J** into a modal extension of **K**, so [1b] would appear the more appropriate. A simplistic analysis might then suggest that such a facility of one logic to encompass another is strong evidence for its superiority. That this analysis is mistaken [as discussed in 2.1.4] is in itself suggestive that unintelligibility is not a necessary condition for significant dissent.) Quantum logic, however, Dummett argues to be an example of (2b), and thus of no more than heuristic usefulness.

Dummett argues that the quantum logical program must be tolerant of the introduction of the classical constants since it is committed to a realist understanding of atomic propositions (that is, propositions attributing some determinate value to a physical quantity of a system at a certain time).[107] Of course, Putnam denies this imputation (see his 1974 in particular); but has he failed to recognize to how much he is committed to **K**? On Dummett's account, although **QL** precludes the conjunction of propositions representing the simultaneous measurement of incommensurable values of a system, nevertheless the values that such measurements would yield, were they possible, are a matter of fact. If we measure the momentum of a particle, we are necessarily ignorant of the position that it had at the time of the measurement; but of all the propositions attributing a position to it at that time, one and only one is true. Dummett (1976a, 272) argues that such epistemological realism ensures that this epistemically unconstrained truth must be preserved by a classical logic. Thus the actual logic of the envisaged situation is classical and the **QL** calculus merely an addendum, tracking our (necessarily incomplete) knowledge of that situation. Crucially, the (realist) quantum logician must recognize **K** as intelligible, if that is the logic of how things really are. Conversely, the classicist should have no objection to the employment of the **QL** constants as supplementary to his own, providing the two are not confused. Hence **QL** is (2*b*).

We discuss Dummett's argument in two stages: first, by questioning whether **QL** really is (2*b*); second by disputing whether this assessment is as damaging as he suggests. In the last section we addressed an argument similar to the first of these stages: that, on realist assumptions, **QL** collapses into **K**. Here Dummett only endorses a weaker result: that the proponent of **QL** must concede the intelligibility of **K**. An uncompromising response to both arguments would be to accept the Kochen–Specker argument and the conclusion of the "guessing" argument, and thereby concede our inability fully to describe the world. On this understanding the ultimate structure of the world would be non-Boolean, committing us to the rejection of one formulation of a realist stance. However, many of our common-sense intuitions would be preserved: Sharp values would be ascribed to all observables in all states, measurement would be noncontextual and there would not need to be any action at a distance. Such an approach would make **QL** self-sufficient, in that all levels of reality would be described by the same system. This might be seen as exhibiting a confidence missing from an account on which the most fundamental level was Boolean, and therefore described by a different logic.[108]

However, the "damaging" concession of **K**'s intelligibility might still seem to be inevitable, since the conceptual resources of **K** are immediately available to **QL**: **K** is recaptured as the system generated by compatible propositions of **QL** (see Delmas-Rigoutsos 1997, 65f. for a proof of this result). But as a purely formal result this need not undermine the integrity of the quantum logical program any more than the recapture of **K** in **J** undermines that of the intuitionist program. Formal equivalence to a proper subsystem is not sufficient for intelligibility. This is why the "center left" response to the recapture result—accepting the formal connection, while flatly denying mutual intelligibility (see 1.4)—is available in both programs. Hence recapture does not entail that (2*b*) is satisfied.

The far-reaching consequences of accepting the conclusion of the "guessing" argument, as Putnam (1994, 295 n65) subsequently notes, mark a disanalogy with the transition to non-Euclidean geometry which motivated his advocacy of **QL**. However, this need not vitiate the overall program. An allied strategy would be to side-step the Kochen–Specker argument by giving **QL** a many-valued semantics.[109] To generalize this point, we may observe that there is a variety of possible candidates for a calculus on which a semantics for **QL** might be constructed, and that that which is most efficient at preserving our common sense physical intuitions need not be **K**. But if the semantics for **QL** are nonclassical, then Dummett's argument that **QL** meets his condition (2) does not go through. He would only be able to show that **QL** were (2*b*) if the paracomplete calculus which provided its semantics could be shown to be so.

However ingenious this may be, it proceeds on the assumption that Dummett's analysis of logical difference is unexceptionable. As we have seen, a relationship of intelligibility is central to this account. Systems which are mutually intelligible (2*b*) are seen as mere terminological relabellings, and not interestingly different. This sort of logical difference is recognizable as that of Quine's (1970, 81) heterodox logician who employs "and" for disjunction and "or" for conjunction. Quine's antirevisionist thesis is that all apparent logical revision can be so characterized; of course, Dummett wants to leave some scope for logical revision. Mutually unintelligible systems (1*a*) are incommensurable at the level of logic, and represent a dispute at the level of the theory of meaning (Dummett 1976a, 288f.). We have already observed that **J**, naturally Dummett's paradigm example of a dispute at the level of the theory of meaning, is (1*b*) rather than (1*a*), that is, it is intelligible to the classicist. By parity, we may assume that (2*a*) logics are treated similarly to (1*b*) logics, and thus that whenever **K** and the nonstandard system are not mutually intelligible, they receive the same analysis as mutually unintelligible systems. Thus Dummett's position is a simple dilemma: Either the difference between the nonstandard and classical systems is merely relabeling, or the two systems are utterly incommensurable. Like Cardinal Newman (1839), he holds that "when men understand what each other mean, they see ... that controversy is either superfluous or hopeless." We shall suggest that this is a false dilemma.

Dummett's position is reminiscent of the account of the divergence of scientific theories advanced by Feyerabend.[110] On this account, when two theories differ significantly there are changes of meaning in apparently common terms which are sufficiently substantial to make the two theories incommensurable. That is to say that neither theory is intelligible from the perspective of a practitioner of the other theory. Hence on Feyerabend's account we must forfeit two of the familiar strategies for theory comparison: consistency and derivability. If the theories are incommensurable they cannot be inconsistent, nor can one encompass the other. Some of Feyerabend's critics (for example, Laudan 1977, 143) have concluded that this amounts to an abandonment of any possibility of objective comparison. In fact, he advanced a variety of strategies for theory comparison, most of which appeal to some broader common factor between theories which are not semantically comparable.[111] This analogy may seem strained, since Dummett's basis of comparison is the theory of meaning and he explicitly rejects any role for empirical considerations, whereas at least one of Feyerabend's bases of comparison is empirical observation and he explicitly rejects semantic comparability.[112] However, the crucial difference is that Dummett is talking about logic, whereas Feyerabend is talking about empirical science. In both cases they argue that theories should be assessed by their fit to the appropriate normative constraints since the terms in which the theories are expressed are semantically incomparable. The theory of meaning is a normative constraint on logic, just as empirical observation is a normative constraint on science; logical theories are expressed in terms of logical constants which, for Dummett, are semantically incomparable, because not mutually intelligible in cases of genuine difference, just as for Feyerabend scientific terms are semantically incomparable, since in cases of genuine difference the theories in which they occur are incommensurable.

A corollary of this account of theory appraisal is that there are two possibilities for theory divergence. We may disagree either about which set of normative criteria is appropriate or we may disagree about which theory best captures an agreed set of criteria. But Dummett (1976a, 288) is exclusively concerned with the former, hence the only prospect he sees for **QL** is in the revision of the theory of meaning.[113] Should the other species of disagreement be so readily dismissed? It may seem eccentric to regard **K** and **QL** as competitors to be appraised by exactly the same class of criteria, although in other disputes, such as that with relevance logic, this seems more plausible. However, the **QL** case does not exhibit the radical discontinuity of normative criteria that characterizes the dispute with intuitionistic logic. The Dummettian classification excludes the possibility of the coevolution of logical theory and normative criteria. Where the dispute is not explicitly couched in terms of the revision of the purpose for which the logic is to be employed, it is not unreasonable to expect that, while key features of the criteria are preserved, others may be revised in the light of developments in the theory. In this evolutionary rather than revolutionary scenario we would expect that many—and hope that all—of the meanings of the logical constants may be preserved.

2.2.4. Quantum Logic and Meaning Variance

In pursuit of an account of evolutionary change, the analogy between Dummett's account of logical revision and Feyerabend's account of scientific theory revision is once more of use. In response to Feyerabend's thesis of the semantic incomparability of theoretical terms, his critics advanced formal accounts of how terms may be retained across the transition between theories. For example, Arthur Fine (1967, 237f.) argues that

> Term S in theory T is carried over into the theory T' [if] the following circumstances are present:

1. There is a characterization of S in T that is
   a. both meaningful and true in T', and
   b. such that, in appropriate and typical situations in which T is employed, this characterization could be offered as a definition of S or as an explanation of what S means in T.
2. There are conditions C that can be formulated in T', such that
   a. objects of T' that satisfy C are suitable objects for T;
      i. if S is a predicate term, then, whenever objects satisfying C satisfy S in T, they satisfy S in T';
      ii. if S is an operation term, then the result in T of applying S to objects satisfying C is the same as the result in T' of applying S to the same objects;
      iii. if S is a term for a magnitude, then...[114]

Can we apply this analysis to logical revision? Further assessment will require us to cash it out in logical terms. Thus theory T becomes **K**, theory T' becomes **QL**, and the terms, S, whose meanings we would wish to see carried over, are the standard metalogical vocabulary—whose definitions are common to both systems and thus readily satisfy both clauses—and all the logical constants. Introduction and elimination rules for the constants in a natural deduction or sequent calculus system would be the most plausible candidates for characterizations meeting Fine's clause (1)(b).

Since Gentzen, there has been an extensive program of looking for the meaning of the logical constants in these sequent calculus or natural deduction operational rules. In so far as this amounts to an attempt to reduce semantics to syntax, it has met with considerable problems.[115] However, as Fine is at pains to point out, our present requirements fall short of a demand for the meaning of S, and hence for the meanings of the constants; thus more modest proposals, such as that of Kosta Došen (1989, 1997), should be adequate. He defends syntactic "ultimate analyses" as sufficiency conditions for the identity of the constants. The operational rules for a constant in a sequent calculus presentation show how an ultimate analysis in terms of the structural rules may be conducted. Hence the question of the identity of the system is separated from that of the identity of the constants, showing how the same constants could figure in more than one system. Identical constants can occur in distinct systems if their operational rules are preserved. As Došen (1989, 367) shows, the relevance system **LR** can be derived from the same class of operational rules as **K**, with the difference between the logics occurring wholly at the level of the structural rules. Hence we have, in Došen's "ultimate analyses," characterizations of the constants of **K** that are meaningful and true in **LR**, and which could be offered as an explanation of what the constants mean in **K**. This satisfies Fine's clause (1); if we specify classes of propositions with sufficient care we should also be able to provide a condition C which recaptures **K** and thereby meets clause (2). Therefore, if we trust Fine's analysis to provide sufficient grounds for meaning retention, we have a demonstration that Dummett's classification is not exhaustive, since the transition from **K** to **LR** is not merely a matter of relabeling. Can **QL** be similarly analyzed, or must it fall into the Dummettian dilemma of mere heuristic extension versus fundamental revision of normative constraints?

Sequent calculus and natural deduction presentations have been developed for **QL** (notably in Nishimura 1980, Cutland and Gibbins 1982, and Delmas-Rigoutsos 1997). However, none of these systems shares the operational rules of **K**: Either an additional nonclassical operational rule is required for negation or additional clauses concerning the compatibility of the premisses must be introduced. Although it would be premature to rule out future developments in this field, we do not yet have a system in which the constants share an "ultimate analysis" with those of **K**. Hence we cannot show that **QL** meets Fine's conditions for meaning invariance as robustly as **LR**.

However, we may be able to meet these conditions with something less formal. In Putnam's (1969, 189f.) original defense of quantum-logical revisionism he enumerates nine "basic properties" of the constants which hold in **QL**:

1. p implies $p \vee q$;
2. q implies $p \vee q$;
3. if p implies r and q implies r, then $p \vee q$ implies r;
4. p, q together imply $p \wedge q$;
5. $p \wedge q$ implies p;
6. $p \wedge q$ implies q;
7. p and $\neg p$ never both hold ($p \wedge \neg p$ is a contradiction);
8. $(p \vee \neg p)$ holds;
9. $\neg\neg p$ is equivalent to p.

(1), (2), and (3) closely resemble disjunction introduction and elimination; (4), (5), and (6) closely resemble conjunction introduction and elimination; (7) closely resembles negation elimination; and (9) is double negation elimination. To this we may add something approximating to negation introduction, say "if p implies absurdity, then $\neg p$ holds," since by orthocomplementation $p \leq \bot \Rightarrow \top \leq p^{\perp}$. Hence we have characterizations of the salient constants that are meaningful and true in **QL** and could be offered as explanations of their meaning in **K**; the first clause of the Fine criteria is met.

As we have already observed, compatible **QL** propositions generate **K**, hence we can also meet his second clause by making the recapture condition C a compatibility relation on the propositions of **QL**, specifically that for any a, b meeting C, $a \wedge (\neg a \vee b) \leq b$. This can only be "center right" recapture, since we are seeking to articulate a program which rejects both of Dummett's alternatives: unintelligibility, which would mandate a "left-wing" response to recapture, and mere relabeling, which would allow "reactionary" recapture (at most). Thus the constants of **QL** satisfy at least one characterization of meaning invariance, and we have a motivation for regarding them as evolving out of the constants of **K** rather than as being added onto those constants as additional terminology. Of course, a program of this character may not meet with success, but our purpose has been merely to show that it is not conceptually precluded.

However, Fine's characterization of meaning invariance is not unique: John Bell and Michael Hallett (1982, 363ff.) employ a different characterization to argue that the meaning of negation cannot be preserved by **QL**. On their account, a term t which occurs in two structures L and L' with common primitives a, b, ..., and is definable in terms of those primitives in one structure but not in the other, or is so definable in both but in nonequivalent ways, does not have the same meaning in both structures. As they show (365), classical negation and **QL** negation do not meet their condition. Classical negation can be defined set-theoretically solely in terms of the partial ordering on its underlying lattice; **QL** negation cannot. (As we noted in 2.2.1, it employs an orthogonality relation, expressive of mutual inconsistency, which corresponds to the perpendicularity of subspaces of a Hilbert space.) Bell and Hallett's condition for meaning invariance is much stronger than Fine's: It requires not only the existence of a common characterization of the disputed term, but also the nonexistence of inequivalent characterizations. Is their condition too strong?

Margaret Morrison (1986, 406ff.) has an argument that suggests that it is. She shows that on Bell and Hallett's account, simultaneity relative to an observer must change its meaning between Newtonian space-time (NST) and Minkowski space-time (MST) because it is uniquely definable in terms of "neither causally precedes" in NST but not in MST. Moreover, it can be shown that simultaneity cannot be otherwise defined in MST (Malament 1977, 299). Hence on Bell and Hallett's account, special relativity does not reconceptualize the understanding of space-time; it changes the subject of physics. Since this conclusion is unacceptable, we have a counterexample to their treatment. The following consideration may reinforce this assessment: Meaning invariance is claimed in two different sorts of cases: where theories compete with one another and where one theory succeeds another (Leplin 1969, 73). In the latter case we would expect the new theory to emerge out of the assumptions of its predecessor, perhaps retaining enough of the successful parts of that theory for it to persist as a limit case. In the former case we are comparing autonomous theories, presumably related as siblings by descent from some common ancestor, but unlikely to have enough in common for either to be a limit case of the other.

Although both accounts aim for generality, Bell and Hallett's is motivated by competition and Fine's by transition. This is explicit in Fine (1967, 237), who presents his task as identifying the "generally discernible circumstances which hold when a term is retained in the transition from one theory to another ... [and which] themselves provide the rationale for retaining the term." Although Bell and Hallett (1982, 363) talk of "the passage from one [theory] to the other," their account characterizes the two theories as beginning from a common set of primitives, a presentation more suggestive of competition than transition. Furthermore, each account is at its most persuasive when addressing the scenario by which it was motivated, and, conversely, at its most vulnerable when addressing the other scenario. As we saw, Morrison's counterexample to the Bell and Hallett account is an instance of transition; conversely, criticism of the Fine account typically employs an example of competition.[116] There is, of course, a sense in which **K** and **QL** are competing systems, but the basis of that competition is precisely that **QL** purports to supersede **K**. If the revisionist program under consideration were to be vindicated, **QL** would succeed **K** just as QM has succeeded classical mechanics. Hence it is Fine's account which is better suited to the sort of revision at issue; and it is Fine's account that supports our conception of that revision.

The last section began with Dummett's analysis of the prospects for logical revisionism in terms of either relabeling or unintelligibility. In this section we have shown this to be a false dilemma, and argued that the program for the adoption of **QL** occupies a middle position. Such a program may not succeed; but it is at least not conceptually impossible.

2.3. Relevance Logic

Both of the reform programs we have discussed so far have origins more or less independent of the classical logic research program. In contrast, the programs of relevance and paraconsistency addressed in the next two sections evolved in direct response to the perceived shortcomings of classical logic. One result of this difference is that it is necessary to consider a variety of different formal systems within each program to gain a convincing sense of either proposal.

2.3.1. Why Make Logic Relevant?

The disagreement which relevance logic has with classical logic is over the concept of logical consequence itself, or as its original advocates called it, "entailment." The two names most commonly associated with the relevance logic program are those of Alan Ross Anderson and Nuel Belnap. However, the crucial idea for the program was first voiced by Wilhelm Ackermann in a paper published in German in 1956, "Begründung einer strengen Implikation," that is, a foundation for a rigorous implication. This rigorous implication, Ackermann wrote,

> should express the fact that a logical connection holds between A and B, that the content of B is part of that of A, or however one wishes to express it. That has nothing to do with the truth or falsity of A or B. Thus one should reject the validity of the formula $A \to (B \to A)$, since it permits the inference of $B \to A$ from A,

> and since the truth of A has nothing to do with whether a logical connection holds between B and A. (113)

Thus the connective "$\rightarrow$" is intended to express the existence of a logical connection—entailment—between its components. It is, therefore, similar to the calculus of strict implication in extending classical logic with a new connective, "$\rightarrow$," sharing Lewis's dissatisfaction with the power of "$\supset$" to express such a connection. However, the theory of "$\rightarrow$" differs from that of "$\prec$" (strict implication) in two ways, ways that make the new calculus a rival and not just a supplement of classical logic.

First, the calculus of entailment, as Anderson and Belnap came to call it, rejects not only the paradoxes of material implication, such as $A \rightarrow (B \rightarrow A)$, noted in the quotation from Ackermann, but also the paradoxes of strict implication. Thus, although $A \vee \neg A$ and $\neg(A \wedge \neg A)$ are theses, $B \rightarrow (A \vee \neg A)$ and $(A \wedge \neg A) \rightarrow B$ are not. Actually, the calculus of entailment can be seen as completing the project that Lewis started but failed to complete. For Lewis had written in 1914:

> That the merely contrary to fact implies anything is repugnant to common sense. But does the impossible—the absurd supposition—imply anything and everything? And is the necessarily true, whose denial is absurd, implied by any proposition whatever? When we include S9 in our postulates, we assume that this is the case. ... If one object to the notion that absurdities imply anything, and that the necessarily true is implied by anything, then it is only necessary to substitute M6 ... for S9. ... This change will eliminate the above theorems and others which have a like significance. (Lewis 1914, 245–246)

Even in 1917, he could write: "A relation which does not indicate relevance of content is merely a connection of 'truthvalues', not what we mean by a 'logical' relation or 'inference'" (Lewis 1917, 356). Lewis, like MacColl before him,[117] set out a project of relevance, but eventually settled for a system containing the strict implicational paradoxes. MacColl said he was forced to this by "the exigencies of logic,"[118] Lewis (1914, 246) that such implications were exemplified in everyday reasoning. Ackermann showed that logic did not so force them; and Anderson and Belnap (1975, §5) diagnosed the fallacies of relevance and modality that misled Lewis.

But if, for example, $(A \wedge \neg A) \rightarrow B)$ is not a valid entailment, then we should hope that B should not be a logical consequence of $A \wedge \neg A$ either. This brings us to the second way in which the theory of "$\rightarrow$" departs from that of "$\prec$." Coherence requires us actually to revise the basis of classical consequence to which "$\rightarrow$" is added. All the tautologies in truth-functional connectives are valid in relevance logic; but the consequence relation is different, even over the truth-functional vocabulary. $A \wedge \neg A \vDash B$ is one consequence which fails; notoriously, so too is $(A \vee B) \wedge \neg A \vDash B$.

Lewis's most famous argument for the validity of $A \wedge \neg A \to B$ runs as follows:

	$A \wedge \neg A \to A$	by *Simplification*
so	$A \wedge \neg A \to A \vee B$	by *Addition*
but	$A \wedge \neg A \to \neg A$	by *Simplification* again
so	$A \wedge \neg A \to (A \vee B) \wedge \neg A$	by *Adjunction*
but	$(A \vee B) \wedge \neg A \to B$	often called *Disjunctive Syllogism* (DS) or *Modus Tollendo Ponens*
so	$A \wedge \neg A \to B$	by *Transitivity* (*Cut*).

We can repeat the argument with $\vDash$ in place of $\to$. So if we want to reject the paradoxes of strict implication by adding a connective "$\to$" for which they fail, and we want our theory of entailment to reflect, that is, to express in the language, our theory of consequence, we need to reject both $A \wedge \neg A \vDash B$ and, consequently, $(A \vee B) \wedge \neg A \vDash B$.

However, these two consequences appear to follow immediately both from Tarski's conditions on consequence and from his semantic analysis of the relation (Tarski 1930). First, consider the following moves:

	$A \vDash A$	by *Reflexivity*
so	$A \vDash A, B$	by *Monotonicity*
hence	$A, \neg A \vDash B$	by DS
so	$A \wedge \neg A \vDash B$	since the premise set is essentially conjunctive.

We have labeled the central move here "DS," for it consists in inferring from the fact that $A \vee B$ is a consequence of X (the conclusion set is essentially disjunctive) that B is a consequence of X given $\neg A$. If that move is legitimate, then from $A \vee B \vDash A, B$ we can infer $A \vee B, \neg A \vDash B$, that is, DS: $(A \vee B) \wedge \neg A \vDash B$.

Establishing the apparent validity of $A \wedge \neg A \vDash B$ is even quicker semantically. For the multiple-conclusion sense of consequence is that $X \vDash Y$ whenever every model of X satisfies some wff in Y, that is, there are no models of X that do not satisfy some wff in Y. But there are no models of $A \wedge \neg A$. So there are none which do not satisfy B. Hence, it seems, $A \wedge \neg A \vDash B$.

There needs to be a twofold revision of the account of $\vDash$ to accommodate the rejection of these inferences as irrelevant. Let us consider the formal account of $\vDash$ first. On the one hand, *monotonicity* seems warranted by the fact that the premise set is intended to be read conjunctively, the conclusion set disjunctively. For both of the following inferences (commonly referred to as *strengthening the antecedent* and *addition*, respectively) are relevantly valid:

if $A \vDash B$ then $A \wedge C \vDash B$,
so, arguably, $A, C \vDash B$
and if $A \vDash B$ then $A \vDash B \vee C$,
so, arguably, $A \vDash B, C$.

However, accepting *monotonicity* is problematic, on account of the following inferences (provable by the *deduction theorem* and DS):

if	$A, B \vDash A$ then $A \vDash B \to A$	(*Positive Paradox*)
and if	$A \vDash A, B$ then $A, \neg A \vDash B$	(*Ex Falso Quodlibet*)
whence	$\neg A \vDash A \to B$.	(*Negative Paradox*).

All three conclusions are relevantly unacceptable, so some move in their deduction must be rejected.

In both cases (*monotonicity* coupled with the *deduction theorem*, on the one hand, and coupled with DS on the other), the diagnosis offered by the program of relevance logic was that there is an equivocation, manifested in the use of sets as the components of the consequence relation. *Monotonicity* is valid, if one thinks of the terms of the consequence relation extensionally, as simply sets of premises and conclusions; but a tighter connection between their constituents is needed to warrant the *deduction theorem* (or *conditional proof*) and DS. Compare *strengthening the antecedent* with the following thought. Take the classically valid move of *importation*:

$$A \wedge B \to C \vDash A \to (B \to C).$$

Let $C = A$; then since by *strengthening the antecedent*, $A \wedge B \to A$ is valid, we can infer $A \to (B \to A)$, *positive paradox*. So *importation* is not relevantly valid. What is needed is some other (in fact, stronger) connection between A and B, one not subject to *strengthening the antecedent* (and *monotonicity*). Let us write $A \circ B$ for this connection (connective)—we call it "fusion." Then

$$A \circ B \to C \vDash A \to (B \to C)$$

is relevantly valid. Correspondingly, we need a similar connection between premise-formulae. Let us continue to use comma (,) for extensional combination (i.e., set union) and introduce semicolon (;) for this intensional connection. Then premise-collections are built up in two ways, forming what we can call "bunches," intensional or I-bunches and extensional or E-bunches (i.e., sets):

1. any wff is an I-bunch,
2. if X, Y are bunches, $X; Y$ is an I-bunch,
3. any nonempty set of bunches is an E-bunch,
4. nothing else is a bunch.

Thus bunches are the appropriate objects for premise-combination. Can they also act as conclusion-combination? Multiple conclusions are disjunctive. Extensional disjunction is the familiar truth-function, $\vee$. What is intensional disjunction? It is usually written "+," it resists *addition* (just as fusion resists *strengthening the antecedent*), and it satisfies DS (just as fusion satisfies *conditional proof*):

$$\text{from } A \to B + C \text{ we can relevantly infer } A \circ \neg B \to C,$$

and

$$\text{from } X \vDash B; Z \text{ we can relevantly infer } X; \neg B \vDash Z.$$

Ex falso quodlibet and *negative paradox* are blocked by the failure of *addition* for "+" (and *monotonicity* for ";").

Let $X[Y]$ denote a bunch X containing a subbunch Y at a distinguished place.[119] Then we can rewrite Tarski's conditions on consequence in a relevantly acceptable way as follows:

$X \vDash X$ (*Reflexivity*)
if $X[Y] \vDash Z[W]$ then $X[Y'] \vDash Z[W']$ (*Monotonicity*),

where $Y \subseteq Y'$ and $W \subseteq W'$, that is, Y', W' are E-bunches containing Y, W, respectively, and

if $X[Y] \vDash Z$ and $V \vDash Y$ then $X[V] \vDash Z$.[120] (*Cut*)

Compactness and *substitutivity* remain as before.[121]

So much for the formal theory. We still have a semantic puzzle, to solve which we need a revised version of the semantics. If $X \vDash Y$ holds whenever every model of X makes some member of Y true, how can EFQ fail?—indeed, if X and Y are now bunches, what is the semantic account of consequence?

Let us focus on EFQ first: $A \wedge \neg A \vDash B$. To invalidate this, it seems we must make $A \wedge \neg A$ true and B false. But surely we cannot make $A \wedge \neg A$ true?—that is why EFQ is thought to be valid. The solution comes from Kripke's semantics for modal logics, in particular, his semantics for nonnormal modal logics (Kripke 1965, §3), in which one considers nonnormal worlds, "worlds" which are not necessarily consistent or complete, and by interpreting "$\neg$" by a cross-world relation. We let $^*: W \to W$, the worlds or indices of the model structure, be such that $a^{**} = a$. Then we say that $\neg A$ is true at $a \in W$ provided A is false at a^*. Hence both A and $\neg A$ may turn out true at a, for some a; thus, taking "$\wedge$" truth-functionally, $A \wedge \neg A$ may be true at a while arbitrary B is false there: $A \wedge \neg A \nvDash B$.

Similarly, if A is true at a, B false there, and A false at a^*, then $A \vee B$ and $\neg A$ are true at a, consequently invalidating DS in the form $(A \vee B) \wedge \neg A \vDash B$.

Ackermann's initial system $\mathbf{\Pi'}$ was amended by Anderson and Belnap to form their system **E** of entailment, by dropping Ackermann's rule γ:

$$\text{from } A \vee B \text{ and } \neg A \text{ to infer } B.$$

This is a rule form of DS. It is an admissible rule of **E**, but not necessarily of **E**-theories, which was their reason for omitting it from Ackermann's formulation.

Subsequently, a range of neighboring systems was developed. Notable are **R**, the calculus of relevant implication, dropping the modal condition (*viz.*

restricted permutation) on the "$\rightarrow$" of **E** (much as "$\supset$" is related to "$\prec$"), and $\mathbf{R}^{\Box}$, the result of adding necessity explicitly to **R**, defining $A \Rightarrow B$ as $\Box(A \rightarrow B)$ for "$\rightarrow$" of **R**. Surprisingly, $\mathbf{R}^{\Box}$ turned out to differ slightly from **E** (matching "$\Rightarrow$" of $\mathbf{R}^{\Box}$ to "$\rightarrow$" of **E**), and interest has subsequently turned away from **E** to $\mathbf{R}^{\Box}$, but more particularly toward **R**.

E and **R** are now well-established examples of what in the 1990s came to be called "substructural logics," logics with restricted structural rules. This is a proof-theoretic characterization, embracing the relevance logics, linear logic, **BCK**-logic, and several others. *Monotonicity* (in Curry's notation, K) is a structural rule—we saw that we needed to restrict its application to the intensional combination in relevance logic. Another structural rule is *contraction* (W), contracting occurrences of X, X (that is, repetitions of premises or conclusions) to X. In **R** it holds for both the extensional and intensional combinations; in the system **RW** and linear logic it is restricted in various ways. Relevance logic is **BCW**-logic, in which fusion satisfies the structural rules B, C, and W; linear logic is **BC**-logic, and classical logic is **BCWK**-logic.[122] The system **E** is $\mathbf{BC^{**}W}$-logic, with restricted permutation, C^{**}.

The semantics of **E** and **R** was developed independently in four ways in the late 1960s and early 1970s. The key was to adapt Kripke's accessibility relation R of the modal logics to create a ternary relation of relative accessibility, where $Rabc$ might now be read "a and b are compossible relative to c" or "c makes true the fusions of what a and b make true," that is, if $a \vDash A$ and $b \vDash B$ (that is, A is true at a and B is true at b) and $Rabc$ then $c \vDash A \circ B$. The relational semantics for **R** is based on frames $\langle 0, W, R, ^* \rangle$ in this way (Anderson, Belnap, and Dunn 1992, §48). Kit Fine showed how one could even develop an operational semantics, replacing the relation R by a binary operation $\circ$ (fusion) on worlds, $\circ : W^2 \rightarrow W$.[123] Recall that "$\circ$" (the connective) is the "residual" of "$\rightarrow$":

$$X \circ A \vDash B \text{ iff } X \vDash A \rightarrow B.$$

In particular,

$$(A \rightarrow B) \circ A \vDash B \text{ iff } A \rightarrow B \vDash A \rightarrow B.$$

Hence

$$(A \rightarrow B) \circ A \vDash B.$$

Thus if $a \vDash A \rightarrow B$ and $b \vDash A$, $a \circ b \vDash (A \rightarrow B) \circ A$ and so $a \circ b \vDash B$ (exploiting a systematic ambiguity in the sense of "$\circ$").

We let an operational frame for **R** consist of a quintuple $\langle 0, W, \circ, ^*, \leq \rangle$ subject to certain constraints (loc. cit.). An interpretation I for **R** consists of an operational frame and an assignment $\mathrm{V}: SL \times W \rightarrow 2$ (where SL is the set of propositional letters) such that

3.1 if $\mathrm{V}(A, a) = 1$ and $a \leq b$ then $\mathrm{V}(A, b) = 1$ for $A \in SL$.

We extend V to a valuation ($a \vDash_{\mathrm{I}} A$, i.e., A is true at a under I):

4.1 $a \vDash_{\mathrm{I}} A$ if $\mathrm{V}(A, a) = 1$ for $A \in SL$,

4.2.1 $0 \vDash_{\mathrm{I}} \boldsymbol{t}$,

4.2.2 $a \vDash_{\mathrm{I}} \boldsymbol{T}$ for all $a \in W$,[124]

4.3 $a \vDash_{\mathrm{I}} \neg A$ if $a^* \nvDash_{\mathrm{I}} A$,

4.4.1 $a \vDash_{\mathrm{I}} A \wedge B$ if $a \vDash_{\mathrm{I}} A$ and $a \vDash_{\mathrm{I}} B$,

4.4.2 $a \vDash_{\mathrm{I}} A \vee B$ if $a \vDash_{\mathrm{I}} A$ or $a \vDash_{\mathrm{I}} B$,

4.4.3 $a \vDash_{\mathrm{I}} A \circ B$ if $b \vDash_{\mathrm{I}} A$, $c \vDash_{\mathrm{I}} B$ and $b \circ c \leq a$,

4.4.4 $a \vDash_{\mathrm{I}} A \rightarrow B$ if whenever $b \vDash_{\mathrm{I}} A$, $a \circ b \vDash_{\mathrm{I}} B$,

5.1 $a \vDash_{\mathrm{I}} X \circ Y$ if $b \vDash_{\mathrm{I}} X$, $c \vDash_{\mathrm{I}} Y$ and $b \circ c \leq a$,

5.2 $a \vDash_{\mathrm{I}} X$, where X is an E-bunch, if $a \vDash_{\mathrm{I}} Y$ for all $Y \in X$.

Then $X \vDash A$ if for every interpretation I, and all $a \in W$, $a \vDash_{\mathrm{I}} A$ whenever $a \vDash_{\mathrm{I}} X$.

To show that, for example, $A \vDash B \rightarrow A$ fails in R, take the frame $W = \{0, 1, 2\}$ with $\circ$, * and $\leq$ defined by the tables:

$\circ$	0	1	2	$*$
0	0	0	0	0
1	0	1	2	2
2	0	2	0	1

$\leq$	0	1	2
0	✓	✓	✓
1	×	✓	×
2	×	×	✓

and let $\mathrm{V}(A) = \{1\}$, $\mathrm{V}(B) = \{2\}$. Then $1 \vDash_{\mathrm{I}} A \wedge \neg A$ and $1 \vDash_{\mathrm{I}} (A \vee B) \wedge \neg A$, but $1 \nvDash_{\mathrm{I}} B$, so $A \wedge \neg A \nvDash B$ and $(A \vee B) \wedge \neg A \nvDash B$. Moreover, $1 \vDash_{\mathrm{I}} A$ but $1 \nvDash_{\mathrm{I}} B \rightarrow A$, since $2 \vDash_{\mathrm{I}} B$ but $2 = 1 \circ 2 \nvDash_{\mathrm{I}} A$. So $A \nvDash B \rightarrow A$.

2.3.2. Does Relevance Logic Recapture Classical Logic?

Recall the "political" spectrum of responses to the possibility of recapture, ordered by their degree of radicalism. Least radical is the reactionary position that the new system should be interpreted as an extension of the old. Next comes the center-right position, in which the old system is understood as a limit case of the new. The left-wing positions involve a rejection of recapture, either as formally valid but unilluminating, because of an incompatibility elsewhere within their research programs, or as formally untenable. Relevance logic provides an excellent illustration of this account, because all of these positions can be identified among the attitudes of its proponents.

In their classification of relevant attitudes, Belnap and Dunn distinguish *irrelevant logicians*, (including classical logicians) who see no connection between relevance and entailment; *relevant logicians in the wide sense*, who acknowledge the importance of a formal characterization of relevant entailment; *relevant logicians* proper, who accept systems such as those listed in the last section as offering such a characterization; and *relevantists*, who advocate these systems as attempts at an organon for natural argumentation.[125] Of these only the relevantists are genuinely revisionary of **K**. They can be further subdivided in terms of their response to DS, crucial to recapture because it is one way of representing the difference between relevant and irrelevant systems. Hence DS may be regarded as either always valid: *soft relevantism*, which collapses into an irrelevant system;[126] or sometimes valid: *hard relevantism*, or never valid: *true relevantism*.[127]

In terms of the political spectrum, relevant logicians, in so far as they retain **K**, are on the reactionary right; hard relevantists (if they can systematize the valid instances of DS sufficiently well) are either center right, if they acknowledge the cogency of the recaptured system, or center left, if they do not; and true relevantists are on the radical left (or the center left if they can offer some alternative method of recapture). So only hard (and perhaps true) relevantists will be seriously interested in developing a successful recapture criterion.

Hard relevantism can be subdivided in terms of the strategies employed to justify the valid instances of DS.[128] Belnap and Dunn suggest four possible strategies: (1) the "I'm all right, Jack" strategy: specifying a contradiction-free domain in which no counterexamples to DS could occur; (2) the deductivist's strategy: proceeding by analogy with the deductivist's response to inductive inference; (3) the "leap of faith" strategy: a specific version of (2), in which relevantly unacceptable inferences are defended "on faith as well as judgement"; (4) the "toe in the water": disjoining the (Ackermann)[129] falsity constant $\boldsymbol{f}$ to the conclusion of all relevantly unacceptable inferences. (2) and (3) are clearly insufficiently concrete proposals to be of present use.[130] John Burgess (1983) subdivides (1) into (1a) *systematic enthymematic relevantism*, in which the recapture domain is specified by conjoining additional premiss(es) to the inferences of that domain, and (1b) *hybrid relevantism*, in which the recapture domain is ensured by the presence of certain background assumption(s), or super-premiss(es).[131] He also identifies a further strategy: (5) *fission relevantism*, whereby (extensional) DS obtains whenever intensional DS, $\sim A, A+B \vdash B$, is valid for the same A and B.[132] Both (4) and (5) are formulated in ways that do not lead straightforwardly to a recapture constraint.

All five strategies are susceptible to criticism. Both (1a) and (1b) are open to the objection that they are either circular or regressive: (1a) essentially involves appending to disputed inferences an additional premiss asserting the legitimacy of that inference (Anderson, Belnap, and Dunn 1992, §80.4.1, 503), and (1b) can be shown to rely on an appeal to DS at a higher level (Burgess 1983, 52).[133] (For example, in Mortensen's presentation of [1b], the validity of DS is supposed to be assured in a domain of wffs which are negation-consistent and prime. That is to say that no more than one of A and $\sim A$, for all A, are contained in the domain and at least one of A and B is in the domain whenever $A \vee B$ is in the domain. Thus if $\sim A$ and $A \vee B$ are in the domain, then A or B must be in the domain, by primality, but it cannot be A, by consistency, so it must be B. This licenses DS within the domain, but employs DS in the metatheory—which must therefore be presumed to be prime and consistent for the strategy to work.) The common feature of these circularity or regress criticisms is that they turn on a scrupulosity about the justification of deduction which occurs elsewhere only in the motivation of generally skeptical theses. Hence (1a) is the first step of Carroll's tortoise (1895, 279) and (1b) exhibits the circularity conspicuous in most attempts to justify deduction (Haack 1976, 186, for example). Meyer (1978, 45f., with acknowledgment to Kripke) identifies the tortoise connection, drawing the moral that the difficulty of justification here is no greater than for relevantly unobjectionable inference.[134] The same point can be made for (1b), wherein we can appeal to familiar moves such as Goodman's (1954, 67) reflective equilibrium or Dummett's (1973d, 296) explanatory/suasive distinction. Even if we remain unconvinced, we are no worse off than usual.

A potentially more serious criticism of all five recapture strategies is that they miss the point of relevance logic (Read 1988, 145ff.). Relevance logic is motivated by dissatisfaction with the classical account of entailment, not by fear of inconsistency.[135] The specification of a locally consistent domain allays the latter concern, but not necessarily the former: Other counterexamples to classical inference may remain, preventing recapture. Such is the case in the true relevantist "Scottish plan" account of validity which one of us has espoused: It construes validity as "the impossibility of true premisses *fuse* false conclusion" rather than "the impossibility of true premisses *and* ($\wedge$) false conclusion" (Read 1988, 147, emphases added). This permits an inconsistency-free counterexample to DS: an assignment of A and B such that $(\sim A \wedge (A \vee B)) \circ \sim B$ is true.

An example lets A be "Socrates was a man" and B, "Socrates was a stone." Since A is true, $A \vee B$ is true; but it would not follow from the falsity of A that Socrates was a stone. (For that we would need the stronger—and false—intensional claim that "If Socrates was not a man, then he was a stone," that is, $A + B$.) However, this counterexample to DS does not occur in the domain specified by an intensional interpretation of the Mortensen recapture criteria of negation consistency and primality. On this understanding negation-consistency would be the noncotenability of A and $\sim A$ within the domain, that is, $\sim$ ((A is in the domain) $\circ$ ($\sim A$ is in the domain)). Primality would require that whenever $A \vee B$ was in the domain, that if A was not in the domain then B was, and vice versa, that is (A is in the domain) $+$ (B is in the domain). (As we would expect, the distinction between intensional and extensional constants would collapse in any domain satisfying these criteria.) If $\sim A \wedge (A \vee B)$ were in the domain, $\sim A$ would also be in the domain by extensional conjunction elimination, whence A would not be in the domain, by intensional consistency. But $A \vee B$ would be in the domain, by extensional conjunction elimination, hence B would be in the domain, by intensional primality. Therefore $\sim B$ would not be in the domain, by intensional consistency, so $\sim A \wedge (A \vee B)$ and $\sim B$ would not be cotenable. Thus these criteria specify a domain in which DS obtains—even for the true relevantist—and thereby recaptures **K**. (We have not established that this domain would be interestingly nonempty. Hence the true relevantist may still have compelling grounds for adopting one of the left-wing responses to this recapture result.)

Hence we have plausible grounds for regarding the theory of **R** as a successor to that of **K**: the product of a glorious revolution.[136] If the true relevantist characterizes **R** in such a way that recapture fails (although we have shown that he need not), then his logical theory should be regarded as a competitor to that of **K**: An inglorious revolution would have occurred. Since his claim is that **K** is so bad a choice of organon as to be just wrong, this should not disturb him. However, his quarrel with **K** is not as fundamental as that of the intuitionist.[137] The true relevantist's competition to the classicist's theory would acknowledge considerable common ground. Not only would the true relevantist argue that his constants were intended to analyze the same operations of natural argumentation that are addressed in **K**, he may also share the classicist's background theories.

2.3.3. Pure and Applied Semantics

It has been alleged that none of the semantics for relevance logic (specifically **R**) qualify it as a plausible reform proposal. Jack Copeland's (1979, 1983a, 1983b, 1986) version of this argument proceeds on two essentially distinct levels. The first level is a contention that the Routley–Meyer semantics for **R** are "pure" rather than "applied." This distinction between pure and applied semantics is intended to capture the difference between constructing a suitable algebraic structure at a wholly theoretical level, and providing a convincing philosophical explication of the components of such a structure. The second level of Copeland's critique is a claim that legitimately to persuade classical logicians of the advantages of **R** over **K**, the advocates of **R** should provide it not only with an applied semantics but with one that assigns classical meanings to all of the constants of **R**.

To assess Copeland's first level of criticism it is necessary to clarify the pure/applied distinction to which he alludes. At root, this may be seen as reflecting an important difference among the motivations for logical endeavor: the distinction we drew in 1.1 between "rough" and "smooth" logic. Examples of each may be identified among relevant and neighboring systems. Whereas abelian logic (Meyer and Slaney 1989) and linear logic (Troelstra 1992) are demonstrably smooth, a case can be made for the syntax and proof theory of **R** being, if anything, significantly rougher than those of **K**. The success of **R** as a progressive revision of **K**, with respect to its status as an organon, hangs on the success of this case—which is what Copeland wishes to dispute.

As Copeland himself (1983a, 197) observes, the distinction between pure and applied semantics has been made by many authors on different occasions.[138] Consequently, it may be drawn in several different ways:

1. One of the earliest statements of the distinction is due to Carnap, who distinguishes between pure semantics as the abstract semantics of for-

mal languages and applied, or descriptive, semantics as the empirically determined semantics of natural language (Blackburn 1995, 820, citing Carnap 1942).

2. In a related vein, the distinction may be thought of as parallel to that between pure and applied mathematics, as appears to be implicit in Plantinga's (1974, 127) statement that "applied semantics . . . places more conditions upon the notion of modelhood."[139] These pure and applied activities are of the same kind, but unlike the former, the latter is apt for, and informed by, application to empirical matters. The actual application is a different practice again—an applied mathematician working on partial differential equations may be conscious of the importance of his work to physics and engineering, but he will leave the actual application to physicists and engineers. Hence it might be said that this difference is really between pure and *applicable* semantics. The actual application would then be something else, such as parsing theory.

3. The distinction can be read as including the application of applicable semantics (applied semantics in sense (2)) within the definition of that discipline. Hence this is a distinction between theoretical and practical activities, wherein the latter is understood as including the former, as well as the means of applying it to the world. This is the sense which the distinction has for Kirwan (1978, 107; see Copeland 1983a, 197), and that which Dummett (1973d, 293) appeals to as "the distinction between a semantic notion of logical consequence, properly so called, and a merely algebraic one."

4. Applied semantics have also been introduced as exclusively the activity of applying pure semantics to the world. This is the fashion in which one of us at one time chose to address the matter, distinguishing formal semantics from the theory of meaning (Read 1988, 166).

Hence applied semantics may be understood as applicable semantics (2), the application of this sort of formalism (4), or both (3). It is a further question for interpretations (2) and (3) whether applied semantics and pure semantics are mutually exclusive or whether the former is a special case of the latter. All of these different senses are translatable into each other; the only danger is that it may be unclear which is intended. Henceforth we adopt interpretation (2), regarding applied semantics as a special sort of pure semantics, meeting additional conditions.

Semantics are crucial to rough logics, since they help link the smoother aspects of a system to the argumentation it aims to codify. Pure semantics are pursued entirely at the smooth level, by merely presenting another formal system onto which the syntax can be mapped. Only if pure semantics can be grounded convincingly in the natural language meanings of its components can an applied semantics, and thereby a rough logic, be established. To see how this works, the connection between logic and natural argumentation must be explored in a little more detail. The logics considered consist of syntactic systems of deductions, $\Gamma \vdash \Delta$, among essentially abstract well-formed formulae. A semantic interpretation for such a system maps these formulae onto the propositions of a system of inferences, $\Gamma' \vDash \Delta'$. The validity of a deduction is characterized in terms of its derivability from the syntactic rules of the system, whereas that of an inference is characterized in terms of the preservation of the inferential goal of the system (for present purposes, truth). So far this is just to connect one formal system to another; semantics can go no further than this. However, applied semantics generates a formal system which can be related to natural argumentation, a linkage which is accomplished by the parsing theory, which governs how the language of natural argumentation is formalized, and informed by the inferential goal and the background theories. In particular the background theories impose constraints on what sort of theory of meaning may be employed. It is within the theory of meaning that philosophical questions about how language is related to the world are addressed. Some logical research programs impose tight constraints on how these questions should be answered, as we saw with the Dummettian program in 2.1.1; others are more liberal.

Hence the semantics of a rough logic must be applicable, which is to say that they must be parsing theory-apt (PTA). This additional condition consists in the formal notions invoked by the semantics being such that they can be related to the world in some intuitively convincing fashion. The philosophical defense of this intuitive conviction is the responsibility of the theory of meaning; there may be competition among different theories of meaning as to which accomplishes this task most effectively for a given semantics. Parsing theory has the humbler task of formalizing natural argumentation. As far as possible, this should be neutral as regards theory of meaning, although it cannot be accomplished at all unless a theory of meaning can be attempted. A logic occurring within the sort of research program with which we are concerned must not only be rough, it must also be feasible as an organon. This might be understood to impose a further condition on the semantics (generality): that they should be interpretable in a way that permits the formalization of natural argumentation in general, rather than merely the argumentation of some specific discourse.[140] This might be paraphrased as the requirement that an organon be global rather than strictly local in application.[141] In a strictly local system recapture of any global system is effectively blocked, in that even if such a system could be recaptured it would be interpreted in a fashion that defeated its intent. A purist might insist that such systems fail to be logic, as they are not subject-independent.

2.3.4. Criticism of the Semantics for Relevance Logic

Copeland claims that there is not enough explanation of R or * to justify their status as applied semantics, and suspects that they may be ad hoc as well as inexplanatory. For these criticisms to be effective, Copeland must make his dissatisfaction explicit. Unfortunately for the clear exposition of his dialectic, although this explication is attempted in his earliest paper, the two levels of his argument which we identified in 2.3.3 are clearly distinguished only in later work.[142] Both I (at least for the positive reduct $\mathbf{R}_+$) and R are similar to their analogs in the possible worlds semantics for modal logics. Presumably Copeland has no criticism of such systems; indeed, much of the precedence he cites for the applied/pure semantics distinction is concerned with their success. The salient difference is the analog of possible worlds in the Routley–Meyer semantics: Unlike possible worlds, "set-ups" can be either incomplete or inconsistent. Initially he regards this widening as illegitimate, because it is harmful to the classical account of negation (Copeland 1979, 402), but that is to confuse the first and second levels of his argument. To be nonclassical, even if unexpectedly, is not to be inexplanatory. On a more sympathetic reading set-ups may seem a useful and progressive generalization of conventional modal semantics, proceeding along a path already taken by much less controversial systems such as Stalnaker's **C2** (1968, 34), the semantics for which includes the inconsistent world λ.[143] Of course, such generalizations will not appeal to modal realists, whose understanding of possible worlds is as (descriptions of) worlds as actual as the real world—unless they are prepared to countenance actual inconsistencies. (And if they can do that, they can employ the more economical dialetheist semantics, in which some propositions are evaluated as both true and false simultaneously. See 2.4.2.) However that leaves Copeland's criticism conditional on establishing the unintelligibility of any account of possible worlds other than modal realism.

In later writing Copeland (1986, 487) accepts the understanding of set-ups by analogy with possible worlds as a plausible basis for an applied semantics. However he then argues that if these worlds are possible, they cannot be inconsistent; if they appear to be inconsistent that must be a result of employing nonstandard negation (ibid., 488f.).[144] This is in part a reprise of his earlier argument, and in part an underestimate of the generalization of possible worlds in use, thereby punning on "possible." Of course, it would still be open for Copeland to assert the incoherence of any generalization sufficient to articulate the Routley–Meyer semantics, but as shown in the last paragraph, this presumes an argument that modal realism is the only intelligible reading of possible worlds. Moreover, his argument depends not just on his second level, but on a strengthening of it: that negation in relevance logic is not only not classical but not negation.

Of all the novelties of the Routley–Meyer system, Copeland concentrates most on the *-operation, as it is here that he believes his criticisms have the greatest chance of success. He contends that the *-operation is ad hoc because its properties were devised solely to preserve the negation axioms of **R** (Copeland 1979, 410). So they may have been; but ad hocness is a methodological complaint, not a historical one: What is at issue is whether there is a convincing rationale for *, not what prompted its discovery. If we can exhibit a plausible role for * in natural language, then we can answer this criticism, and also establish its intelligibility. Generalizing its application from worlds to propositions, Routley and Meyer (1973, cited in Copeland 1979) read A^* as a "weak affirmation" of A, where to affirm a proposition weakly is to refrain from affirming its negation. This seems plausible, but Copeland (1979, 409) objects that it is absurd to attribute any sort of affirmation to such inanimate or insensible individuals as may brutely refrain from the utterance of negations. However, as Meyer and Martin (1986, 312) remind us, the essential content of * is "failure to deny," which is well within the capabilities of rocks, infants, and suchlike.[145] The mistake is not in attributing weak affirmation to rocks, but in engaging them in conversation in the first place: Only if you initiate such conversations will you have cause to attribute propositions to rocks.

Taken together, these observations suggest that a sufficiently intuitive rendering may be given to all of the technicalities of the Routley–Meyer semantics to regard it as applied.

The second level of Copeland's argument is that any semantics employed in the advocacy of **R** over **K** must assign classical meanings to all the constants. First, this is in need of unpacking: The "classical meanings" of the constants is equivocal. On the strongest understanding we could regard the meaning as given by all the instances of use, but then only **K** could have classical meanings. This may be what Quine (1970, 81) intends in his discussion of the "deviant logician's predicament," but it cannot be what Copeland has in mind, unless he is setting the relevantist the impossible task of establishing that **K** is relevant. Conversely, we might seek to derive the meanings of the constants from their introduction and elimination rules; but in this case Copeland's criterion is just that which we introduced in 2.2.4 to defend the preservation of classical meanings by the constants of **QL**. We argued there that a similar case could be made for the constants of **R** (or at least the closely related **LR**) provided that it recaptured **K**, which we demonstrated in 2.3.2. Presumably Copeland is concerned with the definition of the interpretation function, I, which diverges noticeably from the classical in its treatment of implication and negation. Even if we concede that this difference is especially important, we must then ask to what use Copeland hopes to put it; does he have a plausible account of the advocacy of logical revision?

Although Copeland (1983a, 199) talks of preserving the classical meaning of *all* the propositional constants, his primary concern is with negation. (It is perhaps more easily excusable to say that the meaning of implication is not preserved (making the theory of **R** an ingloriously revolutionary competitor to that of **K**), since that is precisely the focus of the reform. However, an argument that the relevant constant represents a progressive precisification of the concept imperfectly articulated by the classical constant, and thus that the theory of **R** is a glorious successor to that of **K**, would appear equally applicable to both implication and negation.) Copeland maintains that the classical meaning of negation must be preserved if the classicist is to be convinced that DS is invalid. But if we are to convince the classicist *qua* classicist that DS is invalid then we must mislead him, for DS is valid in **K**. If the question is rather whether DS is an acceptable move in natural argumentation, then the classicist has no monopoly over the understanding of the terms involved. The relevantist must show that his system is a preferable formalization to the classicist's **K**; if this involves the reform of negation as well as implication, then so much the better—if we improve the understanding of both constants—provided each reform is well motivated. Relevant logicians have claimed the disambiguation of the intuitions underlying negation as an incidental achievement of their program (Meyer and Martin 1986, 310). Hence we can draw a distinction between Boolean negation, which captures the intuition that not-A is true precisely when A is false, and De Morgan negation, which captures the intuition that not-A is true when A implies something objectionable.[146] In **K** these negations collapse into each other, but in relevance logic they can be distinguished. Since **R** can be conservatively extended to $\mathbf{R}^{\neg}$, by introducing Boolean negation which captures the missing classical features, in particular validating DS, Copeland's concerns can be allayed—as he acknowledges (1986, 486).

This raises the question of how best to understand such "classical relevance logics" incorporating both flavors of negation. Copeland is inclined to dismiss De Morgan negation as not really negation at all. Hence he denies that any system incorporating it can be interestingly paraconsistent, and presumably shares with Meyer the belief that relevance logics will come to be regarded as conservative extensions of **K** by the addition of non–truth-functional implication operators rather than as rivals, a transition similar to that made by modal logics. Conversely, it is possible to argue that De Morgan negation is the true heir to the imperfectly articulated negation of **K**, and that it is Boolean negation that is either unintelligible or not recognizable as negation (Priest 1990, 209). This would make good on our earlier claim that the theory of **R** could be regarded as a glorious successor to that of **K**. We shall return to this debate in the next section. Finally it might be argued that the correct inference to draw from the relevantist account of negation is that it is not a univocal notion, and that a good logic should be sensitive to the variety of its possible uses. Such a logic would have a classical reduct, but one with a much narrower range of employment than **K** *simpliciter*, and in this respect it would represent a departure from the classical program.

2.4. Paraconsistent Logic

In recent years some of the most sustained and trenchant criticism of the classical program has come from the advocates of paraconsistency. The focus of their proposed reforms is the classical treatment of inconsistency. In **K** arbitrary propositions may be derived from inconsistent premisses, since A, $\sim A \vdash B$, *ex contradictione quodlibet* or ECQ, is a valid inference. Any logic with a consequence relation for which ECQ is a rule may be said to *detonate*, or to be *explosive* (Priest and Routley 1989a, 151). Any inconsistent theory which is closed under an explosive logic, such as **K**, will be *trivial*, that is, it will contain all the propositions of the underlying language (ibid.). (A theory is *inconsistent* iff there is some A such that the theory contains both A and

$\sim A$.) The aim of paraconsistency is to formalize systems that are not explosive, so that inconsistent but nontrivial theories may be closed under them. These systems are called *paraconsistent logics* and the inconsistent, nontrivial theories in which they are employed are called *paraconsistent theories*.[147]

2.4.1. What Is Paraconsistency?

Paraconsistency is the focus of two closely related yet fundamentally distinct research programs. Although all levels of these programs differ, at least to some degree, the key point of divergence lies in the background theory, the philosophical assumptions constraining the choice of formal system. The *weakly paraconsistent* research program shares the classical background assumption that the world is consistent.[148] Hence its aim is to provide an account of situations in which some of the information under consideration is presumed to be in error: Corrupt computer databases, conflicts of laws, human belief systems, and confusions in the development of science have all been cited as exemplifying this phenomenon.[149] All of these cases are most readily modeled as inconsistent theories in which some form of logical inference applies. Yet in each case, some discrimination between good and bad information is still possible, so the theories cannot be trivial, hence their inference relation must be paraconsistent. The stronger paraconsistent research program, known as *dialetheism*, holds that theories of this kind may be accurate descriptions of the world, and thus forbears from assuming the consistency of the world in its background theory.[150] This program is often motivated by the exhibition of alleged inconsistencies in (the best accounts of) the world, such as paradoxes of self-reference and antinomies in the foundations of mathematics and in accounts of motion (Priest 1987, part three). However, subscription to the program does not strictly require belief in the inconsistency of the world, merely agnosticism about its consistency (Routley, Meyer, Plumwood, and Brady 1982, 60ff.).

The weak paraconsistentist may accept the familiar assumptions of classical background theory unamended, hence his dispute with the classicist is wholly at the level of logic. By contrast, the dialetheist, as we have seen, must diverge from classical background theory. Does that mean that the real focus of the dialetheist's dispute with the classicist is outside the logic? We drew this conclusion for the major intuitionist research programs, which we saw (in 2.1.4) to be most successfully defended through their independently compelling nonclassical background theories. However, the revision of the classical background proposed by the dialetheist (at least *qua* dialetheist) is not as comprehensive, nor is its advocacy as remote from the choice of logic. For the minimum revision of the classical background theory required for dialetheism is toleration of inconsistencies in the world. But strictly speaking, consistency and inconsistency are properties of theories, not of the world. So this revision of the classical background theory amounts to advocacy of a paraconsistent system as essential for the best description of the world. The content of any such advocacy must ultimately turn on the comparison of logics, not of background theories. For both the dialetheist and the weak paraconsistentist the foci of their disputes with the classicist are in the foreground of their logical theories.

The foreground of a logical theory contains the formal system and its attendant metatheory and semantics, and also a parsing theory and an inferential goal. Paraconsistency requires no more than modest revision of the last two of these components. Both paraconsistent programs promise a degree of conceptual simplification of the classical parsing theory, since they obviate the need for a procrustean reinterpretation of all apparent contradictions as hidden equivocations,[151] and because they offer the prospect of a simple articulation of the concepts of naïve semantics (Priest 1987, 157ff.), which require appeal to hierarchies of metalanguages in the standard classical presentation. Both programs can also seek to avail themselves of the attitude to truth preservation—the classical inferential goal—exhibited by the proponents of relevance logic: that they aim to revise only the concept of preservation, and not that of truth, and are therefore engaged in a wholly logical task. We shall consider how successfully this attitude may be maintained.

2.4.2. Can Paraconsistency Be Formalized?

What is the formal content of the paraconsistent programs? Both programs require a consequence relation in which ECQ is blocked. There are several ways of revising **K** that achieve this. Perhaps the most direct is to retain the inference $A \wedge {\sim}A \vdash B$, but block ECQ by excluding the rule of adjunction, $A, B \vdash A \wedge B$, and thereby blocking the inference $A, {\sim}A \vdash A \wedge {\sim}A$.[152] The resultant nonadjunctive systems thus tolerate inconsistency, but detonate in the presence of explicit contradictions. This makes them unsuitable for the dialetheist program, since if the world is inconsistent, we would expect some contradictions to be true. There are also a number of reasons to doubt the suitability of nonadjunctive systems for the weak paraconsistency program (see Priest and Routley 1989a, 157ff. and 171ff). In particular, adjunction is such a fundamental feature of our understanding of conjunction that it is hard to see how any nonadjunctive constant could adequately represent conjunction.[153] Furthermore, should this drawback be remedied through the extension of a nonadjunctive system by a different conjunction constant, and, say, then valid inferences close enough to ECQ to endanger (at least the motivating intuitions of) paraconsistency, such as $A\&A, {\sim}(A\&A) \vdash B$, may be obtained (Priest and Routley 1989a, 160).

Alternative routes to the formalization of paraconsistency run through a reconsideration of the rules for implication. Classical (material) implication has certain properties that are inimical to paraconsistency. Chief among these is the negative paradox of implication (NPI), ${\sim}A \vdash A \rightarrow B$, which leads to ECQ in the presence of *modus ponens*, an indispensable feature of any reasonably recognizable implication, as follows:

$$\cfrac{A \qquad \cfrac{\sim A}{A \to B}\text{ NPI}}{B}\text{ MP}\ .$$

Many paraconsistent systems, such as the nonadjunctive systems considered before and the broadly relevant systems addressed next, also drop the positive paradox of implication (PPI), $B \vdash A \to B$. However, it is possible to formulate paraconsistent systems which retain this rule, known as positive-plus systems. Chief among these are the sequence of systems $\mathbf{C}_n$, for $0 \leq n \leq \omega$, developed by da Costa (1974, 498f.).[154] $\mathbf{C}_0$ is just $\mathbf{K}$, presented axiomatically in the manner of Kleene (1952, §19, 82). Da Costa introduces a new "consistency" operator $^\circ$, such that A° iff $\sim(A \wedge \sim A)$, which is intended to be understood as "A is not a source of inconsistency." The $^\circ$ operator may be iterated, with $A^{(n)}$ standing for $A^\circ \wedge A^{\circ\circ} \wedge \cdots \wedge A^{\circ\cdots\circ}$. Each $\mathbf{C}_n$ nonconservatively extends $\mathbf{C}_0$ by additional axioms for $^{(n)}$, which state that the consistency of wffs ensures the consistency of their combinations: $A^{(n)} \wedge B^{(n)} \to (A \wedge B)^{(n)}$, $A^{(n)} \wedge B^{(n)} \to (A \vee B)^{(n)}$, and $A^{(n)} \wedge B^{(n)} \to (A \to B)^{(n)}$, and by substituting the axiom $B^{(n)} \to ((A \to B) \to ((A \to \sim B) \to \sim A))$ for the classical axiom $((A \to B) \to ((A \to \sim B) \to \sim A))$ (da Costa 1974, 500). In the limit case, $\mathbf{C}_\omega$, this axiom is simply omitted.

There are a number of problems with the positive-plus account of paraconsistency. In the first place, retaining PPI in paraconsistent systems involves the sacrifice of inferences such as *modus tollens* and transposition (TRANS.), $A \to B \vdash \sim B \to \sim A$ (Priest and Routley 1989a, 177). Otherwise the system can be shown to detonate:

$$\cfrac{\sim A \qquad \cfrac{\cfrac{A}{B \to A}\text{ PPI}}{\sim A \to \sim B}\text{ TRANS.}}{\sim B}\to\text{I}\ .$$

Moreover, it is possible to derive explosive inferences of the form $A \wedge \sim A \wedge A^{(n)} \vdash B$ in most of the interesting inconsistent theories for which $\mathbf{C}_n$ might be hoped to offer a paraconsistent formulation (ibid., 167).

Some of the more general problems for positive-plus systems foreshadow difficulties common to all paraconsistent logics to which we shall return in later sections. First, it is disputable whether da Costa's constant "$\sim$" offers an adequate account of negation. As the sketch of the $\mathbf{C}_n$ systems makes clear, the law of noncontradiction, $\vdash \sim(A \wedge \sim A)$, does not apply in da Costa's paraconsistent systems—indeed, its addition to any of them causes a collapse into the explosive $\mathbf{C}_0$. The law of noncontradiction might be regarded as a necessary part of any adequate analysis of negation (see 2.2.4). In particular, its failure suggests that, in traditional terms, A and $\sim A$ are subcontraries rather than contradictories (Priest and Routley 1989a, 165). Two statements are contraries when it is logically impossible for them both to be true, and subcontraries when it is logically impossible for them both to be false (Strawson 1952, 25).[155] Statements that are both contraries and subcontraries are contradictories. While a statement may have many contraries and subcontraries, its contradictory is unique, and should be picked out by the negation of the statement. The informal definitions may be formalized as: A and B are contraries iff $\vdash \sim(A \wedge B)$; A and B are subcontraries iff $\vdash A \vee B$ (Priest and Routley 1989a, 165).[156] $\vdash A \vee \sim A$ is a theorem of all of da Costa's systems, but $\vdash \sim(A \wedge \sim A)$ is not a theorem of any of them, so the $\sim$ constant generates subcontraries, not contradictories, and fails to analyze negation. Furthermore, although it would be possible to augment $\mathbf{C}_n$ by a contradictory-forming negation constant, any such constant would satisfy ECQ, and so fail to be paraconsistent (ibid., 166).

The second problem arises from the so-called Curry paradoxes of naive semantics and naive set theory.[157] The first of these may be regarded as a generalization of the (strengthened) liar paradox, a statement that says of itself that it is not true. If such a statement is not true then it is true, but if it is true then it is not true, so it is both true and not true. Many paraconsistent systems can accept this conclusion, and dialetheists regard it as evidence for the inconsistency of the world. The Curry paradoxes are more recalcitrant. The semantic form is a statement, A, which says of itself that if it is true then so is B, where B may be any arbitrary statement, that is, A says (or is equivalent to) $\mathrm{T}A \rightarrow B$. Application of the truth scheme of naive semantics, $P \leftrightarrow \mathrm{T}P$, yields $(\mathrm{T}A \rightarrow B) \leftrightarrow \mathrm{T}A$. From this we may reason to an arbitrary conclusion as follows:

$$\dfrac{\dfrac{\dfrac{\mathrm{T}A \rightarrow (\mathrm{T}A \rightarrow B)}{\mathrm{T}A \rightarrow B}\,\textsc{abs} \qquad (\mathrm{T}A \rightarrow B) \rightarrow \mathrm{T}A}{\mathrm{T}A}\,\textsc{mp} \qquad \dfrac{\mathrm{T}A \rightarrow (\mathrm{T}A \rightarrow B)}{\mathrm{T}A \rightarrow B}\,\textsc{abs}}{B}\,\textsc{mp},$$

providing that the implication constant obeys the absorption principle (ABS), $A \rightarrow (A \rightarrow B) \vdash A \rightarrow B$ (Priest and Routley 1989a, 172f.). An analogous result arises from the application of the abstraction principle of naive semantics, $A(x) \leftrightarrow x \in \{\, x : A(x) \,\}$, to a "Curried" version of the Russell paradox (let $C = \{\, x : x \in x \rightarrow B \,\}$).

ABS is dependent on the structural rule of contraction (W), since:

$$\dfrac{\dfrac{\dfrac{\dfrac{A \rightarrow (A \rightarrow B) \ \vdash\ A \rightarrow (A \rightarrow B)}{A, A \rightarrow (A \rightarrow B) \ \vdash\ A \rightarrow B}\,\textsc{mp}}{A, A, A \rightarrow (A \rightarrow B) \ \vdash\ B}\,\textsc{mp}}{A, A \rightarrow (A \rightarrow B) \ \vdash\ B}\,\textsc{w}}{A \rightarrow (A \rightarrow B) \ \vdash\ A \rightarrow B}\,\rightarrow\textsc{i} .$$

Because W is admissible in most familiar systems, in particular $\mathbf{J}$ and $\mathbf{K}$, ABS is a property of implication in these systems (ibid., 176f).[158] $\mathbf{C}_\omega$ and $\mathbf{C}_1$ can be shown to extend the positive reducts of $\mathbf{J}$ and $\mathbf{K}$ respectively, by introduction of the da Costa $\sim$, hence their implications exhibit ABS, as do those of the other $\mathbf{C}_n$ systems (ibid., 177). Thus these systems are trivialized by Curry paradoxes, and are unsuitable for some of the most important paraconsistent applications.

The remaining route to the formalization of paraconsistency is more promising—and more familiar. All relevance systems block ECQ, since it is a blatantly irrelevant inference. And, as we saw in 2.3.4, some relevance systems may be given a paraconsistent semantics.[159] However, most systems of semantics for relevance logics preclude the simultaneous ascription of truth and falsity to the same proposition, contrary to our expectations of a dialetheic semantics. Furthermore, a system can be paraconsistent despite validating some irrelevant inferences: The relevance and paraconsistent programs overlap but do not coincide (Read 1988, 138f.). We will follow Priest in calling all systems that achieve paraconsistency by rejecting DS *broadly relevant*, even if they are not strictly relevant.

A variety of different systems of logic and of semantics have been proposed within the broadly relevant approach, however, several of the more interesting systems coincide in their zero-degree reducts (that is, for $\sim$, $\wedge$ and $\vee$).[160] This reduct can be characterized in semantic terms by matrices resembling Kleene's (1952, §64, 332ff.) strong matrices, but with both truth and the middle value designated. (Conventionally, these matrices are understood as paracomplete, with the middle value therefore undesignated.) Formally, this gives us a set of valuations $\mathrm{V} = \{-1, 0, 1\}$ and a valuation function v such that $v(\sim A) = -v(A)$, $v(A \wedge B) = \min[v(A), v(B)]$ and $v(A \vee B) = \max[v(A), v(B)]$ (Avron 1994, 219). The consequence relation may then be characterized as $\Gamma \vdash \Delta$ iff for all valuations v, if $v(A) \geq 0$ for all $A \in \Gamma$ then $v(B) \geq 0$ for some $B \in \Delta$ (where Γ and Δ are sets of propositions) (ibid., 225). An equivalent, but somewhat more perspicuous, presentation of this reduct may be given by thinking of the truth values as sets of the standard values, identifying 1 with $\{\boldsymbol{t}\}$, -1 with $\{\boldsymbol{f}\}$ and 0 with $\{\boldsymbol{t}, \boldsymbol{f}\}$ (Priest 1987, 94f.).

The treatment of implication is somewhat more problematic. Material implication could be introduced by definition, but would not satisfy *modus ponens*, since there are countermodels to $A, \sim A \vee B \vDash B$, as we would expect in a "broadly relevant" system. We suggested that MP was an indispensable feature of implication. However, its failure does prevent the derivation of the Curry paradoxes, and on these grounds material implication has been recommended as a suitable implication for the system adumbrated in the last paragraph (Goodship 1996, 158). We return to this suggestion shortly. To block the derivation of these paradoxes while retaining MP requires an intensional, non–truth-functional implication for which ABS fails. This can be achieved in a similar fashion both in relevance systems, such as Routley's **DK**, and in irrelevant systems, such as Priest's **LP**.[161] However, both approaches require substantial semantic innovation, and are therefore exposed to criticism similar to that of the semantics for relevance logic discussed in 2.3.4. The semantics for **LP** are simpler than those for **R**, since they can dispense with the *-operation and employ a binary rather than a ternary accessibility relation (devices primarily introduced to avoid dialetheism). However, to block ABS, the accessibility relation is obliged to be nonreflexive, an equally startling development (Priest 1987, 107).[162] Semantics for **DK** may be developed in a similar fashion, although with less simplicity (ibid., 114), or algebraically (Priest and Routley 1989a, 179f.); both approaches require counter-intuitive assumptions.

Despite these drawbacks, the broadly relevant approach to paraconsistency is still the most promising, and we concentrate on it for the remainder of this chapter. It is genuinely paraconsistent—there is no source of trivializing inferences, not even the Curry paradoxes—and it is suitable for dialetheism. However, it is still susceptible to the charge that its constants do not have the senses that they purport to have, and therefore that it only succeeds by changing the subject. To respond to this accusation the paraconsistentist must be able to offer an argument that his constants formalize the same intuitions as the classical constants. We turn to this issue in 2.4.4, but first we address the recapture of classical logic in paraconsistent logic.

2.4.3. Classical Recapture in Paraconsistent Logic

Classical recapture is an important result for both paraconsistent programs. All paraconsistentists advocate the employment of systems that can tolerate inconsistency, but they acknowledge that inconsistency will be rare in most discourses, and unknown in some (ibid., 144). Dialetheists propose their program as a successor to the classical program. Hence they see the retention of **K** as a limit case, usable in consistent situations, as evidence of the methodological superiority of dialetheism (ibid., 148; Priest 1989a, 143). Thus the pursuit of a satisfactory account of classical recapture has been the focus of much important work within the dialetheist program, and this issue will repay a little careful attention.

Since we are now chiefly concerned with broadly relevant paraconsistent systems, it is natural to expect there to be close analogs between the paraconsistent and relevance accounts of recapture. In 2.3.3 we analyzed the role of recapture for relevance logics by rehearsing a fourfold itemization of recapture strategies (derived from Anderson, Belnap, and Dunn 1992, §§80.4.1ff., 503ff.). Belnap and Dunn distinguish (1) the "I'm all right, Jack" strategy: specifying a contradiction-free domain; (2) the deductivist's strategy: proceeding by analogy with the deductivist's response to inductive inference; (3) the "leap of faith" strategy: defending the disputable inferences "on faith as well as judgement"; (4) the "toe in the water": disjoining a notion of falsity to the conclusion of all disputable inferences. The focus of all accounts of recapture are the "quasi-valid" inferences: inferences that are classically valid but fail in **LP**, the analog of weak counterexamples in intuitionism (Priest 1989b, 625). The simplest of these approaches is (1): If the domain of discourse is free from contradictions, then the paraconsistent countermodels to the quasi-valid inferences will be absent, and classical inference will be employable without reservation. The difficulty is to specify a condition by which the consistency of a domain could be guaranteed. The consistency operator of da Costa's systems $\mathbf{C}_n$, $A^\circ \equiv_{\text{def}} \sim(A \wedge \sim A)$, would be futile in **LP**, since $\vdash A^{(n)}$ is a theorem of **LP**. Indeed, it may be shown that no such propositional operator on single wffs of **LP** will do the trick (unless **LP** is augmented by nullary constants: Priest 1987, 139). Nor can we specify the consistent domain as the class of wffs for which ECQ holds: If this criterion is expressed in a paraconsistent metalanguage it is compatible with the presence of some A and B such that B is not a consequence of A and $\sim A$ (Batens 1990, 219). Something more sophisticated is required.

At different points Priest appeals to all four strategies.[163] However, his chief account (Priest 1987, 141f.) seeks to formalize the intuitions behind (3).[164] To this end he appeals to the fact that all theorems of **K** are also theorems of **LP**, to establish that whenever $A_1, \ldots, A_n \vdash B$ is a quasi-valid inference, $\vdash \sim(A_1 \wedge \cdots \wedge A_n) \vee B$ is a theorem of **LP** (Priest 1989b, 625). Writing G for $(A_1 \wedge \cdots \wedge A_n)$, we may then reason as follows:

$$\dfrac{\dfrac{\dfrac{G \qquad \sim G \vee B}{G \wedge (\sim G \vee B)}\,\wedge\text{I}}{(G \wedge \sim G) \vee (G \wedge B)}\,\text{DIST.} \qquad \dfrac{G \wedge \sim G}{(G \wedge \sim G) \vee B}\,\vee\text{I} \qquad \dfrac{\dfrac{G \wedge B}{B}\,\wedge\text{E}}{(G \wedge \sim G) \vee B}\,\vee\text{I}}{(G \wedge \sim G) \vee B}\,\vee\text{E}\,.$$

Hence, if we can accept the premisses of $A_1, \ldots, A_n \vdash B$, and accept that it is quasi-valid, we can accept the disjunction of the conclusion with the "crucial contradiction" of the inference, $(A_1 \wedge \cdots \wedge A_n) \wedge \sim(A_1 \wedge \cdots \wedge A_n)$ (Priest 1987, 143). Since all (crucial) contradictions are at least false, the account so far is a version of strategy (4). Priest proceeds from here by appealing to

> Principle R: If a disjunction is rationally acceptable and one of the disjuncts is rationally rejectable, then the other is rationally acceptable. (ibid., 141)

Representing rational acceptability and rejectability as modal propositional operators $\blacklozenge\mathfrak{A}A \equiv_{\text{def}}$ "it is rationally permissible to accept A" and $\blacklozenge\mathfrak{R}A \equiv_{\text{def}}$ "it is rationally permissible to reject A" (Doherty 1998, 488), Priest's principle R may be interpreted as $\blacklozenge\mathfrak{A}(A \vee B), \blacklozenge\mathfrak{R}A \vdash \blacklozenge\mathfrak{A}B$. (We may write $\sim\blacklozenge\sim$ as $\blacksquare$, rational obligation, and retain $\Diamond$ and $\Box$ for the alethic modalities of possibility and necessity.) Because $\blacklozenge\mathfrak{A}$ is closed under logical entailment of its arguments, we may now reason from the acceptability of the premisses of a quasi-valid inference to the acceptability of its conclusion, providing that the crucial contradiction is rationally rejectable:

$$\dfrac{\dfrac{\blacklozenge\mathfrak{A}G \qquad \blacklozenge\mathfrak{A}(\sim G \vee B)}{\blacklozenge\mathfrak{A}((G \wedge \sim G) \vee B)}\,\text{Closure of }\blacklozenge\mathfrak{A} \qquad \blacklozenge\mathfrak{R}(G \wedge \sim G)}{\blacklozenge\mathfrak{A}B}\,\text{R}\,.$$

Hence if $A \wedge \sim A$ is rationally rejectable for all the wffs, A, of some domain, then that domain will recapture **K**.

There are a number of problems with this account. Chief among these is that we might obtain some proposition P which it was rational both to accept and to reject. We might then argue:

$$\dfrac{\dfrac{\blacklozenge\mathfrak{A}P}{\blacklozenge\mathfrak{A}(P \vee Q)}\text{ Closure of }\blacklozenge\mathfrak{A} \qquad \blacklozenge\mathfrak{R}P}{\blacklozenge\mathfrak{A}Q}\text{ R },$$

which would establish the rational acceptability of an arbitrary proposition, or even of $0 = 1$ (Goodship 1996, 155f.). In anticipation, Priest (1987, 123) asserts that joint rational acceptance and rejection is not possible, on the grounds that acceptance and rejection must be manifest in behavior, and the two behaviors could not be manifested simultaneously.

Nevertheless, certain paradoxical propositions have been advanced as counterexamples to this claim (initially in Littman 1991, cited in Priest 1995, 61). Such propositions are of the form A: "It is irrational to believe A." (That is, $A = \blacksquare\mathfrak{R}A$.) If one believed A, one would believe that it is irrational to believe it, which would be irrational. So it *is* irrational to believe A, so A is true, and so one ought to believe it. Hence we can conclude that we ought to accept A, because it is true, and that we ought to reject it, because believing it would be irrational. A similar conclusion is derived in Smiley's (1993, 22) presentation of a (strengthened) liar paradox: "This statement is untrue." For Priest (1987, 90) this proposition is both true and untrue.[165] Hence Smiley argues that we are rationally obliged to accept it, because it is true, and to reject it, because it is untrue. Thus he claims that Priest must abandon his claim that rejection and acceptance are incompatible, or acknowledge that truth and untruth are also incompatible.

Goodship (1996, 153) regards these paradoxes (so-called Littman sentences, after Littman 1991) as telling against Priest's claim for the incompatibility of rational acceptance and rejection, which she construes as "one cannot be rationally obliged to both accept and reject something," and thereby against R. However, as Doherty (1998, 483) points out, Priest's incompatibility claim is not concerned with $\blacksquare\mathfrak{A}A$ and $\blacksquare\mathfrak{R}A$, but with $\blacklozenge\mathfrak{A}A$ and $\blacklozenge\mathfrak{R}A$; it alleges an incompatibility of rational permissibility, not of rational obligations. Priest (1993, 40; 1987, 243) is quite prepared to countenance incompatible rational obligations, "rational binds," since he denies that ought implies can, and in particular that $\blacksquare P \rightarrow \blacklozenge P$. For any Littman sentence A, it follows that $\blacksquare\mathfrak{A}A \wedge \blacksquare\mathfrak{R}A$, a counterexample to Goodship's incompatibility thesis, but not to Priest's. If we could arrive at $\blacklozenge\mathfrak{A}A \wedge \blacklozenge\mathfrak{R}A$, we would have a counterexample to Priest's incompatibility thesis (although not to Goodship's), but no such proposition can be derived from a Littman sentence, unless $\blacksquare P$ implies $\blacklozenge P$, which Priest (1987, 243) denies. In the absence of such strengthened counterexamples, R would be undamaged by this attack.

However, Doherty (1998, 488) suggests two further problems for R. First, he thinks that a proposition such as "It's raining" is sufficient to generate a strengthened counterexample, since it could be rationally accepted at one time and place (or under one precisification) but rationally rejected at another, making it both rationally acceptable and rationally rejectable. But this point trades on equivocation of precisely the kind that the employment of propositions or statements rather than sentences is designed to avoid (Strawson 1952, 4). "It's raining," shorn of any context, as it must be for Doherty's purposes, is not a proposition. Of course, this exception-barring stratagem—responding to issues such as vagueness or time through the parsing theory—is one that nonclassical logicians such as Priest typically wish to minimize. But it can still be the most appropriate response when these issues are not of immediate concern: A formalism should not contain any more logical machinery than necessary. And when these issues are addressed within the logic,[166] the intolerability of the required equivocation is laid bare. Doherty (in correspondence) is also concerned that if rational acceptability and rejectability are spelt out in terms of epistemic probability, as Priest (1987, 143) suggests they might be, then their assessment, and therefore the extent of the recapture domain, will be subjective, varying from individual to individual, and from time to time. But providing that the assignment of probabilities is at least internally coherent, this might be thought to be a harmless, even welcome feature for a paraconsistent system: that it should be able to accommodate contrasting intuitions about the extent of consistency. All that is required is that, wherever and however the boundary is drawn, the recapture domain should validate all and only classically valid inferences, and this result would be secured by R.

A more serious problem for Priest's account of recapture in terms of rational acceptance and rejection concerns its apparent indebtedness to specifically classical concepts. He wishes to argue that **LP** is sufficient for all our needs; unlike some (weak) paraconsistentists he is not content to retain a classical metalanguage. Indeed Priest (1987, 88f.) rejects the distinction between object and metalanguage. Hence he is vulnerable to arguments suggesting that some disputed classical principles are ineradicable from his program.[167] If such arguments carry any weight they present Priest with a dilemma: either he must smuggle in concepts from the system he claims to have superseded or concede the unintelligibility within his (supposedly self-sufficient) system of material essential to the formulation of that system. Moreover, he cannot without circularity respond to this dilemma by an appeal to recapture, if recapture is itself dependent on classical principles.[168] Even inessential inexpressibility is problematic, at least polemically, since Priest (1987, 24ff.) has promoted his system over more familiar classical and paracomplete responses to paradox by stressing the natural-language expressive completeness apparently exhibited by his system, but not by its competitors.[169] If his system is also expressively weaker than natural language he loses this competitive advantage.

The situation can be most easily understood in terms of the spectrum of responses to recapture described in 1.4. We have seen that all paraconsistentists require a connection to classical logic: to describe consistent domains and to defend themselves against a charge of having changed the subject by introducing novel constants. So both "left-wing" responses must be rejected:

The radical left-wing response claims that recapture must fail because of an incompatibility between the formal systems; the center-left response claims that recapture is insignificant because of incompatibilities elsewhere in the logical theories. Priest must also reject the reactionary response, which would reduce paraconsistent logic to an extension of **K**, because he holds that paraconsistent logic is the "One True Logic."[170] This leaves only the center-right position: an understanding of **K** as a limit case of a more general theory, a position which Priest has good independent reasons for wishing to endorse.

However, Priest's account of acceptance and rejection endangers the tenability of this position. We have seen that if these propositional operators are adequate for the justification of R, they must be incompatible. But *in*compatibility is a negative property. It cannot be expressed suitably in terms of paraconsistent (that is De Morgan) negation, as $\sim\exists A(\blacklozenge\mathfrak{A}A \wedge \blacklozenge\mathfrak{R}A)$, since, as Priest (1987, 90f.) himself concedes, this would not rule out $\exists A(\blacklozenge\mathfrak{A}A\wedge\blacklozenge\mathfrak{R}A)$,[171] which would permit the detonation of R. Perhaps Priest could devise a more sophisticated—genuinely exclusive—account of incompatibility. But then he would be open to the introduction of Boolean negation ($\neg$) into his system by the definition (Batens 1990, 216):

i. $\neg A \vdash \sim A$;

ii. $\neg A$ and A are incompatible.

This extension of **LP** would also be an extension of **K**, and thus mandate a reactionary response to recapture.

We shall discuss the significance of Boolean negation for paraconsistency in greater detail in the next section. But even if its introduction here could somehow be blocked, the reactionary response would still seem to be the most credible. If rejection and acceptance are taken seriously, then **LP** and similar systems may be understood as generated within a calculus of acceptance and rejection based on **K**. In effect, we have been utilizing just such a calculus informally in our articulation of Priest's account of recapture. The basis of the formal presentation of such a system would be the equivalencies: $v(A) = \{\boldsymbol{t}\}$ iff $\blacklozenge\mathfrak{A}A\wedge\blacklozenge\mathfrak{R}{\sim}A$; $v(A) = \{\boldsymbol{t}, \boldsymbol{f}\}$ iff $\blacklozenge\mathfrak{A}A\wedge\blacklozenge\mathfrak{A}{\sim}A$; $v(A) = \{\boldsymbol{f}\}$ iff $\blacklozenge\mathfrak{R}A\wedge\blacklozenge A{\sim}A$.[172] Negation could then be De Morgan (paraconsistent) within the scope of the $\blacklozenge\mathfrak{A}$ and $\blacklozenge\mathfrak{R}$ operators, but Boolean (classical) elsewhere. Specifically, we can see that the $\blacklozenge\mathfrak{R}$ operator cannot be formulated without a characterization of exclusion, that is, of Boolean negation. For, from the foregoing equivalencies, we can see that $\blacklozenge\mathfrak{A}A$ iff $\boldsymbol{t} \in v(A)$ and $\blacklozenge\mathfrak{A}{\sim}A$ iff $\boldsymbol{f} \in v(A)$, but $\blacklozenge\mathfrak{R}A$ iff $\boldsymbol{t} \notin v(A)$ and $\blacklozenge\mathfrak{R}{\sim}A$ iff $\boldsymbol{f} \notin v(A)$. The negations within $\boldsymbol{t} \notin v(A)$ and $\boldsymbol{f} \notin v(A)$ must be Boolean lest these statements be compatible with $\boldsymbol{t} \in v(A)$ and $\boldsymbol{f} \in v(A)$, respectively. (If $\boldsymbol{t} \in v(A)$ is compatible with $\boldsymbol{t} \notin v(A)$, then $\blacklozenge\mathfrak{A}A$ is compatible with $\blacklozenge\mathfrak{R}A$, which is ruled out in Priest's informal characterization of these terms.)

It might be thought that Priest's (1991) modified account of recapture—wherein the quasi-valid inferences are default assumptions within a nonmonotonic system—might fare better (Goodship 1996, 157). However, as Priest (1991, 322) makes clear, this system (**LPm**) improves on his earlier account only by offering a less contrived formal theory of reasoning, not by offering a clearer explication of how, if "we remain within the domain of the consistent, classical logic is perfectly acceptable." In particular, **LPm** is no more able than **LP** to specify the classical consistency of a domain. Moreover, **LPm** does not preserve the classical account of inconsistency, since ECQ can never be validated (Priest 1991, 326). So a consistent domain closed under **LPm** is not equivalent to the same domain closed under **K**. The move to **LPm** would not seem to remedy the difficulties with recapture. There are still some possibilities remaining: Priest could argue that the circularity in his definition of rejectability is not vicious, or he could embrace the reactionary approach to recapture, by abandoning global paraconsistency, while retaining dialetheism. Both avenues require further development, although only the first would still be revisionary of logic.

2.4.4. Boolean Negation and Curry Implication

We saw in 2.4.2 that the main problem for the formalization of paraconsistent systems is the provision of accounts of negation and implication that reflect our intuitions but resist trivialization. Specifically, we are concerned with *Boolean negation* ($\neg$) and *Curry implication* ($\hookrightarrow$), which we shall take to be any negation constant satisfying ECQ and any implication constant satisfying MP and ABS. We have seen that the unrestricted presence of these rules is explosive. There are a number of points that the classicist may make to exploit this apparent predicament for the paraconsistentist.

First, he may argue as follows:[173] Boolean negation and Curry implication are intelligible notions. They are absent from any genuinely paraconsistent logic, on pain of trivialization. So paraconsistent logics are expressively incomplete. The issues which dialetheism claims to resolve, such as the paradoxes of self-reference, may be addressed in consistent, expressively incomplete systems: there is no need to endorse a paraconsistent system. There are several lines of reply. Firstly, the weak paraconsistentist is untouched by this criticism: the superiority of his program for the analysis of inconsistent theories does not rest on expressive completeness. Second, it is disputable whether systems intolerant of inconsistency are, *ceteris paribus*, preferable to systems which can tolerate inconsistency. Traditionally, the inconsistency of a theory has been regarded as catastrophic, but precisely because of traditional features of logic, such as ECQ. If these are absent, inconsistency is a less compelling criticism. In effect, the assumption that paraconsistency is a desperate measure, only to be countenanced when every other option has been exhausted, begs the question against paraconsistent logic. Moreover, whereas the expressive incompleteness of the consistent treatments of the paradoxes of self-reference typically affects notions employed in the treatment itself, $\neg$ and $\hookrightarrow$ are not employed at any stage of Priest's account (Priest 1990, 202).[174]

However, the main paraconsistent response to this argument is more bold: a denial that $\neg$ and $\hookrightarrow$ represent intelligible notions. At first sight, this seems extraordinary, since both constants can be introduced into a system such as **LP**, either proof-theoretically or semantically. However, it is a familiar point that constants may not be introduced by arbitrary stipulation of proof-theoretic rules: Some additional constraints must be satisfied (Prior 1960). Various candidate constraints have been mooted, either semantic, which leads to the second means of introducing these constants, or proof-theoretic. Constraints of the latter kind are generally based on the requirement that the new constant should extend the underlying system conservatively (Belnap 1962).[175] Both $\neg$ and $\hookrightarrow$ trivialize paraconsistent systems containing truth-predicates satisfying the truth schema of naive semantics, and are therefore not conservative of such systems. However, as Priest (1990, 205) acknowledges, they are conservative of (some) paraconsistent systems without truth-predicates. Since these truth-predicates conservatively extend the systems to which they are appended, it is only the combined presence of the rules for $\neg$ or $\hookrightarrow$ with such a truth-predicate that is nonconservative. Hence Priest (ibid.) concludes that the conservative extension test cannot tell us which of these is to blame, and is therefore ineffectual in defense of the intelligibility of $\neg$ or $\hookrightarrow$.

The constants must be included within a semantics before they can be accepted as intelligible. The **LP** semantics for $\neg$ may be expressed truth-conditionally, as:

> $\neg A$ is true iff A is not true;
>
> $\neg A$ is false iff A is not false,

in contrast to the semantics for De Morgan negation:

> $\sim A$ is true iff A is false;
>
> $\sim A$ is false iff A is true. (Priest 1990, 207)

Hence the truth conditions for $\neg$ incorporate negation. If this negation is De Morgan negation, then A and $\neg A$ may be true together, and therefore there are counterexamples to ECQ for $\neg$. If that negation is Boolean negation then these truth conditions cannot settle the issue of whether $\neg$ is intelligible, since they must presume its intelligibility, question-beggingly. Similarly, Priest (ibid., 212) claims that the semantics for $\hookrightarrow$ must be given either in terms of the existing constants, in which case the derivation of the Curry paradoxes is blocked, or in terms of itself, which would be question-begging as a defense of its intelligibility.

Hence the paraconsistentist can respond to the classical argument with the claim that he is not compelled to concede the intelligibility of $\neg$ and $\hookrightarrow$. The classicist may respond to this that these constants are perfectly intelligible to *him*, and that they capture indispensable features of natural argumentation: If the paraconsistentist persists in using his own constants instead, then he has changed the subject. This is a dispute about the location of the hard core of the characterization of implication and negation, within (a) the classical program, and (b) natural argumentation. If the features possessed by $\neg$ and $\hookrightarrow$, but not by $\sim$ and $\rightarrow$, that is ECQ and ABS, were part of the hard core of (a), then it would be impossible fully to characterize $\neg$ and $\hookrightarrow$ without them. Hence any adequate characterization of these constants in **K** would be false in **LP**, precluding the identification of $\neg$ with $\sim$ and $\hookrightarrow$ with $\rightarrow$. Even so, it would still be possible for the paraconsistentist to argue that ECQ and ABS were absent from the hard core of (b), and pursue an ingloriously nonclassical program.

We have already seen that many of the same difficulties afflict the provision of an applied semantics for contraction-free logics, such as **LP**, as for relevance logics. Yet the search for applied dialetheic semantics is an ongoing program, and there are some grounds for optimism. The requirement for nonreflexive worlds, interpreted as situations in which the laws of logic differ, may be unexpected, but at least the most counterintuitive features of the Routley–Meyer semantics for relevance logics, such as the *-operator, are not needed (Priest 1992, 299). Goodship (1996, 158f.) suggests retaining contraction and resisting Curry paradoxes by employing only the material conditional, $\supset$, for which MP fails. This would make an applied semantics much easier to achieve, but would necessitate a justification of the surprising claim that MP is not one of the core intuitions of implication. Some solace might be found in the result that MP for $\supset$ is valid in **LPm**.

If center-right classical recapture can be motivated, then much the same diagnosis can be made for **LP** as for the relevance systems assessed in 2.3. **LP** would be a glorious revision of **K** with at least the prospect of a sufficiently applied semantics for the system to be viable as an organon. However, we saw in the last section that although center-right recapture is claimed by Priest, that claim cannot yet be regarded as substantiated. His account of recapture is in danger both of question-begging in the definition of rational rejectability and of accidentally conceding the intelligibility of Boolean negation. If these defects cannot be remedied, then Priest will be forced into a version of Dummett's dilemma:[176] Either **K** and **LP** will be mutually unintelligible, or **LP** will be expressible within an extension of **K**. Both strategies may be coherent ways of articulating dialetheism, but neither is Priest's program.

3. Conclusion

Our first priority in these concluding remarks is to underscore some of the key aspects of the argument of the preceding sections. The first section addressed the dynamics of logic, explaining how and why changes of logic occur and how they can be justified by an exploration of a methodology of theory change for logic. Our account began (1.1) with an account of the broader context of logical systems: logical theories, which include not only syntax, semantics, and metatheory, but also a parsing theory, a set of inferential goals, and the background philosophical theories by which these goals are informed. In 1.2

we offered a clarification of the idea of "revolutions" in the formal sciences. We distinguished four salient possibilities: three sorts of revolution, which we called "inglorious," "glorious," and "paraglorious," and no revolution at all (stasis). A glorious revolution is a transition between theories in which the key components of the original theory are preserved, despite changes in their character and relative significance. In a paraglorious revolution, new key components are added, but in an inglorious revolution key components are lost. If Dummett's dilemma adequately describes the prospects for logical revision, then only stasis and inglorious revolution can occur. Hence, to show this to be a false dilemma, it suffices to show that there can be glorious (and/or paraglorious) revolutions in logic. Not only would this produce a richer characterization of logical revision, but also, a fortiori, an answer to Quine's (1970, 81) alleged predicament.[177]

But the most important aspect of any dynamics of logic must be an account of the diachronic character of logical theories. Here (1.3) we appealed to the treatment of research programs and research traditions by Imre Lakatos (1970) and Larry Laudan (1977, 78ff.), offering a synthesis of the two approaches and exploring how they may be adapted to the case of logic. One result of this treatment was an explanation of how the temptation to see logic as irrevisable arises from confusion between research programs of different depths within the same tradition. In 1.4 we introduced a characterization of "recapture," the means by which the inference relation of one system may be preserved as a special case within another system. As we were to show in section 2, classical recapture provides an explanation of the special status that **K** retains in most nonclassical systems which does not diminish their originality. Returning to Lakatos (1976) in 1.5, we derived a characterization of the possible responses to anomalous data within a research program, which we termed "heuristic context": the practices characteristic of a specific stage in the development of a research program.

This measure of how open a program is to reform and revision generated a range of historically familiar positions, and assembled them into an implicit hierarchy (1.6). The hierarchy begins with ways of dealing with recalcitrant features of natural argumentation that do not involve revision of logic, and continues with the adoption of nonrival logics. However, the most interesting levels are those in which logic must be revised. We distinguished between "restriction of logic," "wholesale reform of logic," and "change of the subject matter of logic." Restriction of logic avoids an anomaly by moving to a new logic which lacks previously valid inferences and theorems. Wholesale reform of logic builds on the former move by exposing to criticism and reformulation the elements of a logical theory beyond the logical system, including metalogical concepts, such as that of consequence, background theories, and the inferential goal. Change of the subject matter of logic is a change of inferential goal precipitated by nonconservative revision of background theories. This shifts the focus of the dispute from the discipline of logic to whatever discipline threw up the conflicting background theories.

In section 2 we applied the picture developed in section 1 to four specific reform proposals. The purpose of these case studies is twofold: collectively, they serve to demonstrate the applicability of our general picture of logical reform to some of the most extensively discussed proposals; and individually, they offer an opportunity to explore the finer detail of a variety of different debates within especially illustrative contexts. The first case study (2.1) was a discussion of intuitionistic logic (**J**). After exploring the detail of the principal research programs by which **J** is advocated, we demonstrated that translations between **J** and **K** and between **J** and **S4** do not establish equivalence between the related systems. We showed that **J** recaptures **K**. The significance of metalogic and proof theory for logical revision has been the focus of an ongoing debate in the philosophical advocacy of **J**. We exhibited this as a false lead, at least as far as both of the chief intuitionistic research programs are concerned. In contrast with most other reform proposals, the point of conflict always retreats to the background theories. This suggests that the heuristic context of these programs is at the final level of the hierarchy we developed in 1.6: change of subject matter.

The second case study (2.2) was of Birkhoff and von Neumann's **QL**, and the program proposing that quantum mechanics would be better understood if this system were adopted as an organon. This program differs from the other case studies in its overtly empirical motivation, although, as we demonstrated in 2.1, this does not stand up to close scrutiny. However, the program still raises some crucial philosophical issues. Hence we used it to explore how a nonclassical system can be cotenable with important classical background theories, and to provide a worked example of our response to Dummett's dilemma.

The third case study (2.3) was concerned with systems of relevance logic. In the programs considered in the first two case studies, there is little room for dispute over which nonclassical system is best adapted to the program's positive heuristic. However, the diversity of possible syntactic and semantic systems is an important feature of the relevance and paraconsistent programs. The relevance program also provides an excellent illustration of the possible responses to recapture, since all four of them are instantiated within various implementations of the program. We concentrated in this case study on the importance of semantics for logical revision, asking what sort of semantics must be provided for a system before it can be advanced as a viable reform proposal. The focus of the treatment was a critique of Copeland's claim that the linkage of the structures of relevance logic to natural argumentation is too weak to justify the application.

Finally, the fourth case study (2.4) was of paraconsistent logic, perhaps the most controversial of serious reform proposals. In this chapter we took particular care to explore some of the intricacy of evolving a system to fit the demands of philosophical background theories. We also examined the mechanism of recapture, which has recently been at the center of some of the most interesting and generally applicable discussion of paraconsistent systems. The irony of this is that, as we demonstrated in 2.4.3, classical recapture is much harder to achieve in paraconsistent systems than in the other nonclassical systems in this chapter.

At the end of 1.6 we promised an important positive application for the change of subject matter level of the hierarchy of logical heuristic contexts. One role that a transition at this level can play is the facilitation of a glorious revolution brought about by shifting the program onto new foundations offering higher standards of rigor and improved generality. Klein's *Erlanger Programm* may be understood as a move of this sort within geometry. Klein's achievement was to found geometries not in more or less arbitrary lists of axioms, but in the invariants under groups of transformations, each group corresponding to a different geometry (Klein 1893, cited in Boyer and Merzbach 1989, 548f.). Thus "geometry" was reified from a subdiscipline of mathematics to an object of mathematical study, reconstructing an ancient subject on the modern foundations of group theory and linear algebra (Marquis 1998, 186f.).

We may now discern two contrasting prognoses for the near future of research into the logic of natural argumentation. This is often portrayed (Haack 1974; Sarkar 1990, for example) as a continuing dispute among a proliferation of largely unrelated, competing nonclassical programs, each of which seeks the status of sole successor to classical logic. However, within the heuristic context appropriate to the highest level of the hierarchy, change of subject matter, this proliferation of logics may be understood to represent a refinement of logical method. The original quarry, the best logic for natural argumentation, has given way to something of higher generality: a structure that integrates the best features of a plurality of logics—an *Erlanger Programm* for logic. The articulation of such a structure as applied to natural argumentation is still in its earliest stages, but much recent work toward the provision of a general account of logical systems may lend itself to the advancement of this program.[178] Because **K** would be subsumed within this structure as a key component, the program might best be regarded as a treatment not of *non*classical logic but of *post*classical logic.

Further Reading

For many years, students of alternative logics were ill-served by the publishing trade, with few monographs, no textbooks, and key results often found only in inaccessible journals or circulating in samizdat form. In recent years, the situation has improved beyond recognition.

No fewer than three introductory textbooks (Bell, DeVidi, and Solomon 2001; Priest 2001; Beall and Van Fraassen 2003) are now available, as well as an accessible collection of survey articles (Goble 2001). Each of these works considers all of the programs we discussed in section 2, except quantum logic. Both of the major encyclopedias of philosophy to have appeared in the last decade (Zalta 1995–; Craig 1998) offer extensive coverage of nonclassical logics, in contrast with previous such endeavors. Even greater detail may be found in the vastly expanded second edition of the *Handbook of Philosophical Logic* (Gabbay and Guenthner 2001–).

More specific projects include important historical work on the early years of the intutionistic program (including Hesseling 2003) and several collections of new work on quantum logic (including Coecke, Moore, and Wilce 2000; Dalla Chiara, Giuntini, and Greechie 2004; Weingartner 2004). The relevance and paraconsistent programs have benefited from the consolidation of recent results in textbooks (Restall 2000; Mares 2004) and continue to produce new work (Priest, Beall, and Armour-Garb 2004). Perhaps most interestingly of all there is evidence of a renewed fascination with the phenomenon of logical pluralism itself (Brown and Woods 2001; Beall and Restall 2006).

Notes

1. We refer to formal systems by boldface acronyms, to avert confusion with the broader programs by which they are advocated. Hence, by **K** we mean classical logic, propositional unless clearly first-order by context.

2. The difference between consequence and deductive systems corresponds to Tennant's (1996, 351f.) distinction between, respectively, gross and delicate proof theory.

3. For example, **K** has the same theorems in $\neg$, $\wedge$, $\vee$, and $\supset$ (if $A \supset B$ is defined as $\neg A \vee B$) as the relevance system **R**.

4. Note that Dummett uses this terminology for a different distinction, specifically, he defines smooth logics as systems in which the rules of inference and proof coincide, and rough logics as systems in which they do not (Dummett 1973a, 436).

5. A browse through any issue of the *Journal of Symbolic Logic* will furnish numerous examples.

6. The theoreticity of observation originates with Pierre Duhem (see his 1904, 145ff., and the discussion in Gillies 1993, 132ff.) and is widely discussed in modern philosophy of science. A contrast may be drawn between two versions of this position: a "harmless" position which simply exhibits the dependency of observations on theory, and a stronger, more philosophically contentious position which denies that observation and theory can be clearly distinguished (Wright 1992, 159ff.; Lakatos 1970, 96ff.). Only the "harmless" position is assumed here.

7. For an impression of the difficulties of this activity, see Walton (1996).

8. See section 2 below and chapter 17 of this volume.

9. Gillies (1992, 5) distinguishes "Franco-British" from "Russian" revolutions in similar terms to our contrast of "glorious" and "inglorious" revolutions. Our terminology may exhibit unabashed persuasive definition, but it sidesteps the historical exegesis prompted by Gillies: Why is the French revolution more like the British than the Russian? What are the start and end points of each revolution?

10. As pursued in many of the contributions to Lakatos and Musgrave (1970), notably Toulmin (1970). For Kuhn's own account of normal science see Kuhn (1962), 23ff.

11. Which is why Gillies (1992, 5) thinks the French Revolution was glorious, since he includes the 1815 restoration of Louis XVIII within its scope. However, this indicates the instability of assessments made on such a large scale, since there seems

no good reason why he should not have gone further still and included the overthrow of Louis-Philippe in 1848, which makes the whole affair inglorious.

12. This problem is exacerbated by the epistemological confusion discussed shortly. Also see 2.2.4.

13. In Crowe's terminology, glorious and (tacitly) paraglorious revolutions are "formational events" and inglorious revolutions are "transformational events."

14. $\mathbf{R}^{\neg}$, the conservative extension of **R** with Boolean negation, is a common extension of **K** and **R**; see Meyer (1986) for details.

15. For instance, Dauben (1984, 62), Gillies (1992, 6), and Dunmore (1992). However, Crowe has moved from denying that there any revolutions in mathematics (1975, 19) to suggesting that even inglorious revolutions may be possible (1988, 264f.; 1992, 313).

16. "Research programs" are introduced properly in 1.3; for the time being we will use this phrase informally, with its literal sense.

17. Larvor (1997, 52) has an alternative argument to this conclusion: That although mathematicians seldom misreport the *phenomena* of their discipline, they still err in the *explanations* they offer for these phenomena. This exhibits the importance of maintaining the distinction between science and subject matter in logic and mathematics.

18. This is a simplification of Lakatos's epistemology: See Larvor (1998), 64 and Motterlini (2002).

19. For further discussion of the merits of this strategy see Preston (1997), 169ff. It has been suggested that Lakatos (1976) exhibits a methodological anarchism absent from MSRP, since in this work he counsels against the unconditional acceptance of any methodological rule (Larvor 1998, 87). However, *that* sort of anarchism is endorsed in Lakatos 1970 (51), wherein the positive heuristic is placed outwith the hard core; the contrast with the later Feyerabend is in the very existence of enduring methodological constraints.

20. Ziman (1985, 2 fig. 1) gives a helpful diagram of the nesting of research programs (although his methodology is not explicitly Lakatosian). He also notes (p. 6) that any detailed picture of the interdependency of different areas of research will be immensely topologically complicated. However, this need not diminish the illustrative force of a suitable simplification.

21. This picture is inevitably an idealization, in so far as it presumes a system-independent characterization of the hard core of different programs. This concern is mitigated by the degree of rational reconstruction involved in any articulation of research programs.

22. Some optimists would disagree—for example, Horgan (1996)—although not Lakatos, who is consistently antihistoricist (see Larvor 1998, 29). In the case of logic, such optimism would correspond to Kant's view that logic is one of the "few sciences that can attain a permanent state, where they are not altered any more" (Kant 1992, 534, see also 438). The dangers of this position are manifest in Kant's further opinion that this permanent state had been attained by Aristotle.

23. After the usage of Laudan (1977, 78ff.). Our characterization meets two of the defining conditions of his account (common metaphysical and methodological assumptions, and tolerance of a variety of different, even mutually exclusive, constituent programs), but perhaps not the third (specification of certain exemplary theories within the tradition).

24. Aristotelian syllogism and an early implementation of classical logic respectively. See Haack (1974), 26ff.

25. Quine's position is discussed at length by Plantinga (1974), 222ff.

26. This section is based on Aberdein (2001a).

27. The earliest usage we have been able to find of the word "recapture" to describe a relationship of this kind is Priest (1987), 146, although such relationships have been discussed in other terms for much longer. Sometimes this has been in a weaker sense, as the reproduction of the theorems of the prior system, or in a stronger sense, as the reproduction of the proofs of that system. See Corcoran (1969, 154ff.) distinction between logistic, consequence and deductive systems discussed in 1.1.

28. A sequent is, generally, a pair of sets of wffs; a single-conclusion sequent is a pair of a set of wffs and a wff.

29. See the distinction between expressive power and deductive power drawn by Rautenberg (1987, xvi).

30. "Perhaps ... any genuine 'logical system' should contain classical logic as a special case": van Benthem (1994), 135. Kneale and Kneale (1962, 575) also seem to be committed to this view.

31. In this case the situation is complicated by Belnap and Dunn's (1981) claim not to embrace the radical left stance themselves; rather they attribute it to "the true relevantist," whose position they wish to criticize.

32. "Intuitionists ... deny that the [classical] use [of the logical constants] is coherent at all": Dummett (1973b), 398. But see Dummett (1973a), 238 for a more conciliatory intuitionist response to recapture.

33. See, for example, Lewis (1932), 70 and chapter 12 of this volume.

34. Susan Haack acknowledges the possibility of such a limitation to her attempt to define rivalry on purely syntactic grounds, although her choice of examples downplays its likelihood (Haack 1974, 6).

35. Whose *How to Solve It* (Pólya 1945) he translated into Hungarian. Compare, for instance, Pólya (1945), xxxvif. and Lakatos (1976), 127f.

36. We propose "monster-exploiting" as a shorthand for what Lakatos calls "the method of proofs and refutations," into which he subsumes subsidiary methods of "lemma-incorporation" and "content-increasing" (64). Bloor (1983, 145 n12) suggests the more colorful "monster-embracing," citing Caneva (1981). However, Caneva (1981, 108f.) actually uses the misleading "monster-assimilating" here, reserving "monster-embracing" as an (equally misleading) synonym for "primitive exception-barring."

37. The conjecture was first published in Euler 1758, although it had been anticipated in a manuscript of Descartes's (Lakatos *loc. cit.*). As Worrall and Zahar note in their preface to Lakatos 1976 (ix), a recurring criticism of Lakatos (1976) is that it is derived from a narrow diet of examples, beyond which it is not applicable. Their hope that this complaint could be answered with additional case studies from Lakatos's thesis, omitted in its earlier journal publication, seems precipitate: see Anapolitanos (1989), 337. However, in recent years much more of the history of mathematics has been fruitfully explored on Lakatosian terms (discussed at length in Larvor 1997, 43ff.).

38. Lakatos (1976, 14ff.) extracts a wide variety of such monster-barring responses from the literature.

39. See fig. 7, ibid., 17 for a helpful illustration.

40. In reinforcement of this assimilation, Bloor (1983, 139 n2) notes the presumably serendipitous employment of similar analogies by logicians, for example: "For some [the Lewis principles, $(A \wedge \neg A) \rightarrow B$ and $A \rightarrow (B \vee \neg B)$] are welcome guests, whilst for others they are strange or suspect" (Makinson 1973, 26).

41. The resultant hierarchy partially overlaps a similar account, from which the quoted headings are taken, developed by Haack (1978), 153ff.

42. For example, an uncharitable reading of Strawson (1952), 88.

43. The more conservative of Strawson's two approaches to nondenoting terms, as reconstructed by Nerlich (1965), 34.

44. For a discussion of Frege's treatment of vagueness, see Williamson (1994), 37ff.

45. Russell (1905, 484) criticizes Frege's proposal as "plainly artificial"—a fairly swift response given the comparatively limited reception of Frege's work at that time. See Haack (1974), 127f.

46. A more charitable reading of Strawson (1952), 88 would be to the same effect (see note 42).

47. But see McCawley (1981), 222ff. and Davis (1998) for criticism of the Gricean program and its claims of progress.

48. See Read (1988), 179ff. for an application of Wittgenstein's proposal. Similar methods have been used against other anomalies: for example, the treatment of the paradoxes of material implication in Balzer (1993), 76.

49. See Sorensen (1999), 159f. for criticism of this suggestion and broader discussion of the employment of orientation in logical notation.

50. Van Fraassen's paracomplete supervaluations have been dualized to paraconsistent "subvaluations" by Hyde (1997) (see 2.4.2). A theory (and any logic admitting it) is said to be "paracomplete" if for some, but not all, of its propositions, neither the proposition nor its negation is true; that is, $B \vdash A, \neg A$ fails. Dually, a theory (and any logic admitting it) is said to be "paraconsistent" if for some, but not all, of its propositions, both the proposition and its negation is true; that is, $A, \neg A \vdash B$ fails.

51. Kripke (1975, 64 n18) is particularly insistent on this.

52. The identity of the more radical proposal should be obvious for most of the examples, except perhaps Russell's misleading form analysis. Here the competitor theory is Meinong's account of nonexistent objects, which Russell (1905, 482ff.) criticizes extensively (and perhaps unjustly). Meinong's program did not have a formal logical presentation at this time, although proposals for remedying this have been published subsequently, for example, Parsons (1980) and Jacquette (1996).

53. Propositions employing higher-order quantifiers appear anomalous because they seem to be inexpressible in first-order logic. A rearguard claim that they can be expressed elsewhere in the language is monster-adjustment, since by redefining the anomalous vocabulary as nonlogical, it prevents conflict with the prevailing logical theory. In contrast, a claim that a certain discourse is unintelligible represents a principled delimitation of the subject matter of logic: an exception-barring move, as discussed.

54. See 1.4 for details of recapture and the possible responses to it.

55. Examples of this process are the progress of the Lewis modal systems, such as **S4**, from apparent rivals to **K** to extensions of **K**; and the understanding of **J**, interpreted as merely a calculus of (classical) provability, in the light of the Gödel–McKinsey–Tarski translation (see 2.1.2).

56. This system was first suggested as a progressive revision of classical logic some 30 years later, notably in Putnam (1969); see 2.2 for further discussion.

57. Koetsier (1991) (cited in Larvor 1997, 53) complains that Seidel seems unaware of the importance of his methodological innovation; but, as Larvor (1997) responds, proper assessments of significance require historical perspective.

58. Typically through the application of situation theory, as in Devlin (1991): particularly programmatic passages may be found at 10f. and 295ff. But see Mares (1996) and Restall (1996), wherein situation theory is assimilated to the less comprehensively revisionist relevance logic program.

59. Such as that of Johnson and Blair (1997, 161), who "distinguish informal logic from formal logic, not only by methodology but also by its focal point . . . the cogency of the support that reasons provide for the conclusions they are supposed to back up." More extensive treatments may be found in Johnson (1996) and Walton (1998).

60. The two most frequently cited sources are Nye (1990) and Plumwood (1993). Although Nye (1990, 175) concludes her indictment of "masculine" logic with the claim that "there can be no feminist logic," her alternative, a dialectic of care (ironically derived from the work of male critics, such as Paul de Man), could be seen as a revision at the final level of our hierarchy—in which the word "logic" itself would be jettisoned, despite the retention of some of its methods. Plumwood's defense of relevance logic on feminist grounds is more conservative, and might be thought to belong in the previous level of the hierarchy. However, programs are not characterized by their formal calculi alone: Plumwood's revision of the classical background theories is clearly substantial and her program not necessarily continuous with that of the more orthodox advocates of relevance logic. Both positions have been heavily criticized, notably by Haack (1996, xvf.; 1998, 125 n9) and Curthoys (1997, 68ff.).

61. Of the examples given, only the informal systems and Nye's proposal require the loss of some key components of the formal system (indeed all of them, if she is taken at her word, as advocating the abolition of logic). Devlin (1991, 10) is clear that he regards **K** as a special case, and Plumwood's preferred formal system, **R**, also recaptures **K**, as we shall show in 2.3.

62. See note 50 for a definition.

63. There are exceptions to this attitude, as we shall see.

64. For instance, as attempted in Weyl's (1918) constructive set theory (cited in Quine 1970, 88) or Lorenzen's (1955) "operative mathematics" (cited in Körner 1960, 153f.).

65. This presentation is essentially due to Heyting 1956, 98f., 102f., but as presented in van Dalen (1986), 231. Further refinements of (iii) and (vi) due to Kreisel (1965, 129) may be introduced to ensure the decidability of the proof relation.

66. Earlier, partial, axiomatizations were produced by Glivenko and Kolmogorov.

67. This is actually a stronger requirement than strictly needed to obtain **J**, hence **J** may be given a multiple-conclusion presentation (Takeuti 1975; Read 1995, 229).

68. This originated in Dummett (1959a), and has been developed extensively in subsequent work, notably Dummett (1991) and Tennant (1997).

69. Such application is not without further difficulties: see Tennant (1997), 48, 403ff.

70. In the exposition of this argument we follow Tennant (1997), 176ff. Dummett has presented the argument in many different locations, notably his (1973a), 466ff.

71. In so far as Dummett discusses this possibility at all (for instance, in his 1982, 258f.), he substantially underestimates its feasibility (Tennant 1997, 169f.).

72. Ultimately, Tennant is no friend to the Gödelian Optimist, and wishes to argue that this position is either ad hoc or incoherent (ibid., 239). However, we shall not consider this argument here.

73. Kremer (1989, 58) suggests that the meaning theory of Brandom (1983) should do the trick.

74. Tennant's account of empirical discourse (1997, 403ff.) proceeds along similar lines (although his meaning theory is intended to motivate adoption of his version of intuitionistic logic).

75. For the distinction between rough and smooth logic, see 1.1. There are also many smooth applications of **J**, for example, in computer science.

76. There has been commentary both for (e.g., Schwartz 1987) and against (e.g., Read and Wright 1985) the proposal. Williamson (1994, 300 n13) briefly surveys the debate.

77. We follow Gallier (1991, 73) in presentation.

78. A more sophisticated double-negation translation, with several practical advantages for proof theory, has been published by J.-Y. Girard (1991).

79. McKinsey and Tarski (1948, 13f.) established the preservation of antitheorems, that is $\vdash_{\mathbf{J}} A$ only if $\vdash_{\mathbf{S4}} A^*$.

80. The GMT translation embeds **J** within a different reduct of **S4** from the Gödel translations.

81. Rasiowa and Sikorski (1953, 93) prove this for the GMT translation; Troelstra (1990, 297) shows how their proof may be generalized to Gödel's earlier translations.

82. Alternatively, we could think of this as using a double-negation translation to introduce the "missing" constants, $\vee$ and $\exists$, into $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$ by definition, which is how the scenario is envisaged by Gödel (Kneale and Kneale 1962, 679). Since the resultant system would be clearly equivalent to $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall}$, the underlying question is the same: Is $\mathbf{J}_{\neg,\rightarrow,\wedge,\forall} \cong \mathbf{K}$?

83. Remember that in 1.4 we distinguished between proper subsystems, which may have the same constants as the parent system, but only a subclass of the wffs, and proper reducts, which have only some of the constants of the parent system, but partition the class of inferences containing only those constants into the same valid and invalid subclasses, and are thus extended by the parent system.

84. Further examples in this vein can be found in (Dummett 1976a, 285) and at the end of (Brouwer 1912, 89).

85. For an argument that these results only establish completeness for at most the positive reduct of **J**, see Dummett (1977), 265ff.

86. These two claims may also be distinguished in Copeland's (1983a, 200, for example) criticism of the semantics for **R**.

87. Kolmogorov's position is discussed by von Plato (1994, 200ff.).

88. Brouwer (1952, 142) assumes an intuition of the continuum lacking in Kolmogorov's stricter constructivism.

89. Some constructivists have pursued this corollary, and advocated a negation-free logic (e.g., Griss 1946). However, this (rather extreme) move cannot help here, since we are seeking an explanation of how classical negation, but not classical disjunction, could be seen as intuitionistically acceptable.

90. The latter being what Dummett (1991, 332f.) calls the "ancillary use of non-classical constants."

91. Interpolation is one notable omission.

92. Actually, Tennant (1996, 382) seems unsure whether **K** preserves its preferred species of truth. *Qua* relevantist, this is perhaps understandable (see 2.2), but *qua* intuitionist his qualms seem to turn on a criticism of classical truth (1996, 361f.), which moves the focus of the debate away from the formal system and toward the goal of the system. We will see more of this move shortly.

93. The earliest account of harmony (Dummett 1973a, 396f.) expressly employs Belnap's conservative extension requirement, which was articulated in response to Prior (Belnap 1962).

94. An observation of Takeuti's (1975) cited by Gallier (1991, 40).

95. In both sequent calculus and natural deduction form. For example, $\mathbf{GKT}_i$ (Gallier 1991, 41) and $\mathbf{NJ}'$ (Ungar 1992, 56ff.), respectively.

96. Normalization theorems have been produced for various presentations of **K**: for example, Shoesmith and Smiley (1978), 366ff.; Weir (1986), 477f.; Ungar (1992), 150ff. Weir (1986, 466ff.) offers an inversion principle satisfied by **K** but not **J**, which he argues is more natural than Prawitz's version and offers an account of harmony for the classical constants.

97. For example, Weir (1986, 479) anticipates that the intuitionist might respond that his inversion principle favors stronger logics. Of course, in this case Prawitz's principle could be said to favor weaker logics.

98. Described as "revision of the scope of logic" in Haack's analogous hierarchy (Haack 1978, 155). There is some ambiguity in this use of "scope" (see Resnik 1996, 497).

99. Notably by Crispin Wright. For example, see his dissent from Rasmussen and Ravnkilde's claim that there are "no anti-realistically acceptable semantics which will validate classical logic for all statements not known to be effectively decidable" (Wright 1982, 468ff., citing Rasmussen and Ravnkilde 1982).

100. The following exposition is derived chiefly from Birkhoff and von Neumann 1936; van Fraassen 1974; Redhead 1987; and Foulis 1997.

101. Other systems of logic, such as the **R3** of Reichenbach (1944), have also been inspired by QM, but none of them have generated as much philosophical interest as **QL**.

102. A Hilbert space is a complete, normed inner product space. That is, there is a mapping assigning a real number to every element, and every pair of vectors has an inner product. The inner product function associates a scalar $\mathbf{u} \cdot \mathbf{v}$ with a pair of vectors $\mathbf{u}$ and $\mathbf{v}$ such that $\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u}$, $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$ and $n\mathbf{u} \cdot \mathbf{v} = n(\mathbf{u} \cdot \mathbf{v}) = \mathbf{u} \cdot n\mathbf{v}$, for any scalar n.

103. A lattice is a partially ordered set such that every pair of elements has a least upper bound and a greatest lower bound.

104. There are several possible ways of introducing a conditional into **QL**, none of them wholly satisfactory. This one, the "Sasaki hook," satisfies *modus ponens* and several other desirable constraints (Dickson 2001).

105. Putnam's program is first suggested in his (1962, 248), and articulated fully in his (1969).

106. Dummett's suggestion postdates his (1976a), although the spirit of this argument is present in that paper. The argument does not turn on the infinitude of the system considered: a suitable finite system is given by Kochen and Specker (1967). (See Putnam 1994, 294 n65).

107. "The realistic terms in which [Putnam] construes statements about quantum mechanical systems cannot but allow as legitimate a purely classical interpretation of the logical constants as applied to such statements." Dummett (1976a), 285.

108. For the opposite argument, in defense of a classical metalanguage, see 2.1.2.

109. Such an approach is developed in Kamlah (1981), 320ff. He employs Reichenbach's three-valued system **R3**, which was independently developed as a logic for QM, although a less esoteric system would do as well for our purposes.

110. Feyerabend's position fluctuates, and is plagued by difficulties of exposition not presently relevant. A version close to that advanced here is stated in Feyerabend (1962), 75.

111. Preston (1997, 117) lists eight strategies suggested by Feyerabend.

112. Feyerabend circa 1962, that is; he subsequently denied any normative role to empirical observation, notably in his (1975).

113. The approach resulting from acceptance of the result of the Kochen–Specker argument, considered already, would be in sympathy with this analysis.

114. We abbreviate clause (2)(b)(iii) because it is clearly irrelevant to the logical case.

115. For example, there is an extensive discussion of the problems which led Popper to abandon his allegiance to this program (as developed in Popper 1947a and 1947b) by Schroeder-Heister (1984).

116. For example, Hesse (1968), 48, a convincing response to which is given in Leplin (1969), 71ff.

117. See MacColl (1906) and Read (1998).

118. "It is true that in ordinary speech the conjunction *if* usually suggests some necessary relation between the two sentences it connects; but the exigencies of logic force us to adhere to our definition, $A : B = (AB')^{\eta}$ [i.e., $A \to B = \neg\Diamond(A \wedge \neg B)$] and disregard this suggested relation." MacColl (1901).

119. X is a subbunch of itself; Y and Z are subbunches of any bunch of which $X;Y$ is a subbunch; Y is a subbunch of any bunch of which any set containing Y is a subbunch.

120. This formulation of Cut is valid; however, it does not really recognize the multiple-conclusion nature of consequence. The full multiple-conclusion version of relevance logic has not yet been worked out.

121. The *compactness* condition is that any consequence of a set is a consequence of some finite subset of it. *Substitutivity* was articulated by Tarski in a later article (1936, 415). It expresses the fact that logical consequence is formal, that what follows from a set of premises does so in virtue of its form, not its particular content. To articulate this thought, we need a distinction between the logical and the descriptive vocabulary in the grammar.

122. B is associativity for fusion, or prefixing for "$\to$," and C is permutation for "$\to$."

123. See Anderson, Belnap, and Dunn (1992), §51, where it is called the "operational-relational semantics."

124. On $\boldsymbol{t}$ and $\boldsymbol{T}$, see note 129.

125. Anderson, Belnap, and Dunn 1992 §80, 489f. (The chronology of the ensuing dialectic may appear mysterious unless it is noted that this section predates Meyer [1978]. It was first published as Belnap and Dunn [1981], but circulated in typescript from 1976.)

126. *Pace* Tennant, whose systems are omitted from this catalog.

127. This subdivision is explicated in Routley (1984).

128. Such strategies are itemized by Anderson, Belnap, and Dunn (1992), §§80.4.1–4, 503ff.; Meyer (1978), 85; Burgess (1983), 47ff. and Bhave (1997), 403.

129. The Ackermann constants $\boldsymbol{t}$ and $\boldsymbol{f}$ (the true and the false) represent the conjunction of all logical truths and the disjunction of all logical falsehoods, respectively, whereas the Church constants $\boldsymbol{T}$ and $\boldsymbol{F}$ (the trivial and the absurd) represent the disjunction of all propositions and the conjunction of all propositions respectively (Anderson and Belnap 1975, §27.1.2, 342). The conjunction and disjunction used here are the extensional lattice constants, hence the conjunction of a set of propositions is the weakest proposition which implies every element of the set, and the disjunction the strongest proposition implied by every member of the set.

130. Meyer dismisses (2) as "recommended to the relevantist, not so much as a concrete option but as a brand of lunacy to which he, too, can aspire" (1978, 85). He is more favorable toward (3), viewing it as inevitable in the face of general skepticism about deduction (ibid., 94f.), to the distaste of Routley (1984). However, this cannot rest on a literal reading of (3), which advocates leaps of faith only for *some* inferences.

131. Burgess is critical of all the strategies he identifies: His purpose is to show that disjunctive syllogism represents a class of arguments for whose validity relevance logic is unable to account. He attributes (1a) to Routley (1981), although the position in Routley 1984 is closer to (1b), which Burgess attributes to Mortensen (specifically, his 1983). Belnap and Dunn only address (1a) in their criticism of (1) (Anderson, Belnap, and Dunn 1992, §80.4.1, 503).

132. This strategy may be found in Anderson and Belnap (1975), §16.1, 165f.

133. Priest (1989b, 624) makes the same criticism of the presentation of (1b) he finds in Routley and Routley (1972). For his preferred approach see 2.4.3.

134. Also see note 130. However, Routley argues that the tortoise's argument turns on a relevantly unacceptable conflation of exportation and importation, that is of the tactics of (1a) and (1b), and thus that (1b) is the only feasible version of (1) (Routley, Meyer, Plumwood, and Brady 1982, 30; Routley 1984).

135. One option would be to dismiss this original motivation as historical, and focus instead on the utility of relevance logic for reasoning in potentially inconsistent circumstances, a move encouraged by the adoption of either dialetheic or American plan semantics (see below). Ultimately, however, this is to give up on the positive heuristic of relevance logic and adopt that of paraconsistent logic instead (see 2.4).

136. Because **R** is not an extension of **K**—recall that extension is defined in terms of valid inference, not just theorems (see 1.4). The class of extensional theorems of **R** is equivalent to the class of theorems of **K** (Anderson and Belnap 1975, §24.1.2, 283f.), making **K** a proper reduct of **R** in this weaker, "logistic" sense.

137. Contra Belnap and Dunn, who suggest a parallel between intuitionism and true relevantism (Anderson, Belnap, and Dunn 1992, §80, 489). This suggestion is criticized at length by Meyer (1978, 18ff.) who notes a variety of disanalogies—such as **J**'s origins in an already articulated philosophical system and its intrinsic non–truth-functionality—which suggest that the intuitionist's dissent is more fundamental.

138. He attributes the terminology to Plantinga (1974, 126ff.) and also quotes Dummett (1973d, 293f.), and Kirwan (1978), 107 with approval.

139. In the same passage Plantinga offers "depraved semantics" as a synonym for applied semantics, although it seems inappropriate to regard as depraved something that must satisfy extra conditions. Some of his other remarks suggest that his position should be seen as closer to interpretation (3) or (4).

140. An example of a system with a semantics which clearly fails the generality condition is Michalski, Chilansky, and Jacobsen's 12-valued system (where each value corresponds to a month of the year) for employment in the diagnosis of plant disease (Haack 1978, 214). Less esoterically, we shall suggest that the American plan may fail this condition.

141. This is approximately the distinction that Haack makes between local and global pluralism (Haack 1974, 42ff.; 1978, 223ff.). We differ from her in excluding only *strictly* local systems, which resist even the paraphrase of general argumentation—that is, they do not recapture any system that could represent general argumentation, under their semantics.

142. Although the distinction is first suggested by Copeland (1979, 406), its importance to the dialectic only becomes clear in his 1983a (200). This confusion serves to illuminate what Copeland (1983a, 199ff.) takes to be a deplorable misreading of his 1979 in Routley, Routley, Meyer, and Martin (1982).

143. A generalization of standard possible world semantics to paraconsistent logic with recapture of the standard system has been worked out in detail in Mares (1997).

144. It is apposite to recall the difference between a "world" in which both a proposition and its negation are true, and a "world" in which a proposition is both true and false. Both situations involve (at least) the generalization of some classical notion: in the former, either the characterization of negation or that of deducibility; in the latter, the understanding of truth and falsity. The latter situation, a much graver revision requiring reappraisal of inferential goal and background theory, is never required by the Routley–Meyer semantics.

145. Clarifying the relations between denial, rejection and negation is crucial to the understanding of nonclassical accounts of negation (see Priest 1993, 36ff.). We shall return to this in 2.4.3.

146. Of course, both flavors of negation may be represented implicationally: Boolean in terms of Church falsehood, as $A \rightarrow \boldsymbol{F}$, De Morgan in terms of Ackermann falsehood, as $A \rightarrow \boldsymbol{f}$ (Meyer 1986, 302ff.). The novelty of De Morgan negation is that it captures this intuition *alone*.

147. The term "paraconsistent" was introduced by Miró Quesada in 1976, although systems of this character have a much longer history (Arruda 1989, 127).

148. "Weakly paraconsistent" is Routley's terminology (Routley, Meyer, Plumwood, and Brady 1982, 59).

149. Priest (1987) and Priest and Routley (1989b) both contain discussion of these and other examples, some of which have been addressed at greater length elsewhere, for example, Meheus (1993); French and da Costa (2002); Abe and Pujatti (2001).

150. Priest (1987, 4) offers an etymology for this neologism. Dialetheic systems have also been called "dialectical logics," a name which (perhaps unduly) emphasizes their connection to the Hegelian and Soviet traditions of "dialectical philosophy" (Routley, Meyer, Plumwood, and Brady 1982, 60 n2).

151. A clear example of the monster-adjustment strategy may be seen in Empson's assurance that "grammatical machinery may be assumed which would make the contradiction into two statements" (Empson 1930, 196) and Rescher's "difference-of-respect" procedure (Rescher 1973, 96).

152. This strategy originates in Jaskowski (1948). More recent systems in this tradition include those of Rescher and Brandom (1980) and Schotch and Jennings (1989).

153. Hyde (1997, 652ff.) defends his nonadjunctive system on the grounds that the problem is dual to a similar drawback in supervaluational systems (they are non-subjunctive, since $A \vee B \vDash A, B$ fails). This gives him a nice *ad hominem* argument (ibid., 654 n13) against David Lewis, who elsewhere advocates supervaluationism (Lewis 1970, 228f.; 1976, 70), but is not otherwise a defense of nonadjunctive "conjunction."

154. In other papers da Costa formalized quantified extensions of these systems, $\mathbf{C}_n^*$, and quantified systems with identity, $\mathbf{C}_n^=$.

155. Strawson harmlessly simplifies these definitions; most traditional logicians additionally specified that contraries may both be false and that subcontraries may both be true, making contrary, subcontrary and contradictory mutually exclusive (e.g., Watts 1724, 198; Whately 1826, 34).

156. This formalization might be disputed, but not obviously to da Costa's advantage. In particular, Slater (1995, 452) claims that something stronger is required, and therefore that *all* paraconsistent "negations" are really just subcontrary forming.

157. Originating as a generalization of the Russell paradox in Curry (1942).

158. It also holds for strict implication and the intensional implications of some of the most popular Anderson-Belnap relevant systems, such as **E** and **R** (although not their contraction-free relatives **EW** and **RW**).

159. The contrast between weak paraconsistency and dialetheism corresponds to that between epistemic and ontic readings of the auxiliary truth values of the American plan semantics (see Anderson, Belnap, and Dunn 1992, §81, 506ff.).

160. Notable exceptions include the four-valued systems, such as Belnap and Dunn's American plan semantics, which are paracomplete as well as paraconsistent.

161. **DK** is one of Routley's "depth relevance" systems; an axiomatization may be found in Routley, Meyer, Plumwood, and Brady (1982), 289. The adequacy of this system for resisting triviality from the Curry paradoxes is established in Brady (1989). A semantic characterization of implication for **LP** is given in Priest (1987), 106. Priest explains his preference for an irrelevant system, 110ff.

162. Reflexivity is retained for the actual world, and Priest defends the worlds where it fails as "logically impossible situations," where different laws of logic apply (Priest 1992, 292).

163. (1): Priest (1987), 146, where he suggests the use of the Church falsity constant $\boldsymbol{F}$; (2): ibid., 145; (3): ibid., 141f.; (4): ibid., 146. Although Priest does not refer to Belnap and Dunn's classification, he does cite the paper in which it originated (ibid., 140).

164. Priest has modified the account of recapture given here, in (1991), 322ff. The latter account is technically superior, as he observes in (1996b), 655 n9, but is still motivated by the same considerations (Priest 1991, 322; *pace* Goodship 1996, 156, who sees the accounts as diverging).

165. Priest argues that the conditional employed in the truth schema is not contraposible, and thus distinguishes falsity from untruth. Since he regards simultaneous truth and untruth as no more problematic than simultaneous truth and falsity, this does not seem to be an indispensable feature of his project. Without it, the difference between the strengthened liar and its simpler variant ("This statement is false") would disappear (Doherty 1998, 489 n23).

166. As both have been in paraconsistent logic: time in Priest (1987), 204ff.; vagueness in many different systems, summarized by Hyde (1997, 645f.).

167. We discussed similar arguments against quantum logic in 2.2.2–3.

168. In contrast, the intuitionist, who as we suggested in 2.1.2, might appeal to recapture in response to criticism of the use of **K** in completeness proofs for **J**. The intuitionist can establish his recapture criterion, decidability, entirely on his own resources.

169. Many critics of Priest have attacked this claim. The most important attacks, to which we return in the next section, allege the inexpressibility of one or more of the logical constants. Other attacks of this kind include those of Denyer (1989) and Everett (1994), answered by Priest (1989c and 1996a, respectively).

170. Note that this is a feature of his global paraconsistency (monism about paraconsistent logic) rather than his dialetheism (agnosticism about the consistency of the world).

171. Priest (1993, 39 n8) later remarks that " 'exclusive' ... must mean more than that the conjunction cannot be true"—but he does not say what else is needed.

172. This account may be understood as an interpretation of the "couple semantics" of Batens (1982), cited in his (1990, 212 n10).

173. An argument of this kind is attributed to Thomason (1986) by Priest (1990, 203).

174. In contrast, the treatment of the liar paradox by Kripke (1975) takes the paradox to be nontrue and nonfalse, notions that are ineffable within the formalism.

175. More sophisticated constraints, such as harmony (Dummett 1973a, 397), typically incorporate this requirement. See 2.1.3.

176. See 2.2.3 for a derivation from Dummett (1976a), 285.

177. The complexity of Quine's views on logical revision makes faithful exegesis difficult, but an influential reading is that apparent changes of logic can always be explained as resulting from superficial relabeling, like the consequences of mistranslation (Haack 1974, 14f.; Morton 1973, 503ff.). This would make Quine's view approximate to one fork of Dummett's dilemma. A more sophisticated view of Quine's position would allow for the possibility of either fork (see Quine 1970, 96; Levin 1979, 57ff.; Priest 2003). Since on this reading Quine's position is equivalent to Dummett's, we stick to the naive interpretation of the deviant logician's predicament, which possesses an interest independent of its provenance.

178. Promising leads include Belnap's display logic (Anderson, Belnap, and Dunn 1992, §62, 294ff.), Feferman's theory of finitary inductively presented logics FS_0 (Feferman 1989), Gabbay's labeled deductive systems (Gabbay 1996), Beall and Restall's logical pluralism (Beall and Restall 2000, 2006), and Sambin's basic logic (Sambin, Battilotti, and Faggian 2000; Sambin 2002). One of us has treated this program at greater length elsewhere (Aberdein 2001b, from which the last two paragraphs are adapted).

References

Abe, J., and L. Pujatti. 2001. A meta-interpreter based on paraconsistent legal knowledge engineering. *Logic and Logical Philosophy* 9: 129–136.

Aberdein, A. 2000. *Revising Logic*. Ph.d. diss., University of St Andrews.

Aberdein, A. 2001a. Classical recapture. In *Prospettive della Logica e della Filosofia della Scienza*, eds. V. Fano, M. Stanzione, and G. Tarozzi, 11–18. Catanzaro, Italy: Rubettino.

Aberdein, A. 2001b. Post-classical logic? *Beiträge der Österreichischen Ludwig Wittgenstein Gesellschaft* 9(1): 22–28.

Ackermann, W. 1956. Begründung einer strengen Implikation. *Journal of Symbolic Logic* 21: 113–128.

Anapolitanos, D. A. 1989. Proofs and refutations: A reassessment. In *Imre Lakatos and Theories of Scientific Change*, eds. K. Gavroglu, Y. Goudaroulis, and P. Nicolacopoulos, 337–345. Dordrecht: Kluwer.

Anderson, A. R., and N. Belnap. 1975. *Entailment: The Logic of Relevance and Necessity* 1. Princeton, N.J.: Princeton University Press.

Anderson, A. R., N. Belnap, and M. Dunn. 1992. *Entailment: The Logic of Relevance and Necessity* 2. Princeton, N.J.: Princeton University Press.

Arruda, A. I. 1989. Aspects of the historical development of paraconsistent logic. In *Paraconsistent Logic: Essays on the Inconsistent*, eds. G. Priest, R. Routley, and J. Norman, 99–130. Munich: Philosophia Verlag.

Avron, A. 1994. What is a logical system? In *What Is a Logical System?*, ed. D. Gabbay, 217–238. Oxford: Clarendon.

Balzer, N. 1993. *The Human Being as a Logical Thinker*. Amsterdam: Rodopi.

Batens, D. 1982. A bridge between two-valued and many-valued systems: n-tuple semantics. In *Proceedings of the Twelfth International Symposium on Multiple-valued Logic*. Los Angeles, Calif.: IEEE.

Batens, D. 1990. Against global paraconsistency. *Studies in Soviet Thought* 39: 209–229.

Beall, J. C., and G. Restall. 2000. Logical pluralism. *Australasian Journal of Philosophy* 78: 475–493.

Beall, J. C., and G. Restall. 2006. *Logical Pluralism*. Oxford: Oxford University Press.

Beall, J. C., and B. C. van Fraassen. 2003. *Possibilities and Paradox: An Introduction to Modal and Many-Valued Logic*. Oxford: Oxford University Press.

Bell, J., D. DeVidi, and G. Solomon. 2001. *Logical Options: An Introduction to Classical and Alternative Logics*. Peterborough: Broadview.

Bell, J., and M. Hallett. 1982. Logic, quantum logic and empiricism. *Philosophy of Science* 49: 355–379.

Belnap, N. 1962. Tonk, plonk and plink. *Analysis* 22: 130–134.

Belnap, N., and M. Dunn. 1981. Entailment and disjunctive syllogism. In *Philosophy of Language/Philosophical Logic*, eds. G. Fløistad and G. H. von Wright. The Hague: Nijhoff.

Bhave, S. V. 1997. Situations in which disjunctive syllogism can lead from true premisses to a false conclusion. *Notre Dame Journal of Formal Logic* 38: 398–405.

Birkhoff, G., and J. von Neumann. 1936. The logic of quantum mechanics. *Annals of Mathematics* 37: 823–843.

Blackburn, S. 1995. Semantics. In *The Oxford Companion to Philosophy*, ed. T. Honderich, 820. Oxford: Oxford University Press.

Bloor, D. 1978. Polyhedra and the abominations of Leviticus. *British Journal for the History of Science* 11: 245–272.

Bloor, D. 1983. *Wittgenstein: A Social Theory of Knowledge*. London: Macmillan.

Boyer, C., and U. Merzbach. 1989. [1991]. *A History of Mathematics*. 2nd ed. New York: Wiley.

Brady, R. 1989. The non-triviality of dialectical set theory. In *Paraconsistent Logic: Essays on the Inconsistent*, eds. G. Priest, R. Routley, and J. Norman, 437–471. Munich: Philosophia Verlag.

Brandom, R. 1983. Asserting. *Noûs* 17: 637–650.

Brouwer, L. E. J. 1912. [1983]. Intuitionisme en formalisme. Translated by A. Dresden as "Intuitionism and formalism." In *Philosophy of Mathematics: Selected Readings*, reprint, eds. P. Benacerraf and H. Putnam, 77–89. Cambridge: Cambridge University Press.

Brouwer, L. E. J. 1952. Historical background, principles and methods of intuitionism. *South African Journal of Science* 49: 139–146.

Brown, B., and J. Woods, eds. 2001. *Logical Consequence: Rival Approaches.* London: Hermes Science Publications.

Bub, J. 1989. The philosophy of quantum mechanics. *British Journal for the Philosophy of Science* 40: 191–211.

Bub, J. 1991. The problem of properties in quantum mechanics. *Topoi* 10: 27–34.

Burgess, J. P. 1983. Common sense and relevance. *Notre Dame Journal of Formal Logic* 24: 41–53.

Caneva, K. 1981. What should we do with the monster? Electromagnetism and the psychosociology of knowledge. *Sociology of the Sciences Yearbook* 5: 101–131.

Carnap, R. 1942. *Introduction to Semantics.* Cambridge, Mass.: Harvard University Press.

Carnap, R. 1950. *Logical Foundations of Probability.* Chicago: Chicago University Press.

Carroll, L. 1895. What the tortoise said to Achilles. *Mind* n.s. 4: 278–280.

Coecke, B., D. Moore, and A. Wilce. 2000. Operational quantum logic: An overview. In their *Current Research in Operational Quantum Logic: Algebras, Categories, Languages*, 1–36. Dordrecht: Kluwer.

Copeland, B. J. 1979. On when a semantics is not a semantics: Some reasons for disliking the Routley–Meyer semantics for relevance logic. *Journal of Philosophical Logic* 8: 399–413.

Copeland, B. J. 1983a. Pure semantics and applied semantics: A response to Routley, Routley, Meyer, and Martin. *Topoi* 2: 197–204.

Copeland, B. J. 1983b. A rejoinder to Routley, Routley, Meyer, and Martin. *Journal of Non-Classical Logic* 2: 61–66.

Copeland, B. J. 1986. What is a semantics for classical negation? *Mind* 95: 478–490.

Corcoran, J. 1969. Three logical theories. *Philosophy of Science* 36: 153–177.

Craig, E., ed. 1998. *Routledge Encyclopedia of Philosophy*, 10 vols. London: Routledge.

Crowe, M. J. 1967. Science a century ago. In *Science and Contemporary Society*, ed. F. J. Crosson, 105–126. Notre Dame, Ind.: University of Notre Dame Press.

Crowe, M. J. 1975. [1992]. Ten "laws" concerning patterns of change in the history of mathematics. In *Revolutions in Mathematics*, reprint, ed. D. Gillies, 15–20. Oxford: Clarendon.

Crowe, M. J. 1988. Ten misconceptions concerning mathematics and its history. *Minnesota Studies in the Philosophy of Science* 11: 260–277.

Crowe, M. J. 1992. A revolution in the historiography of mathematics? In *Revolutions in Mathematics*, ed. D. Gillies, 306–316. Oxford: Clarendon.

Curry, H. 1942. The inconsistency of certain formal logics. *Journal of Symbolic Logic* 7: 115–117.

Curthoys, J. 1997. *Feminist Amnesia: The Wake of Women's Liberation.* London: Routledge.

Cutland, N. J., and P. F. Gibbins. 1982. A regular sequent calculus for quantum logic in which $\wedge$ and $\vee$ are dual. *Logique et Analyse* 25: 221–248.

da Costa, N. C. A. 1974. On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic* 15: 497–510.

Dalla Chiara, M. L. 1986. Quantum logic. In *Handbook of Philosophical Logic* 3, eds. D. Gabbay and F. Guenthner, 427–469. Dordrecht: Reidel.

Dalla Chiara, M. L., R. Giuntini, and R. Greechie, eds. 2004. *Reasoning in Quantum Theory: Sharp and Unsharp Quantum Logics*. Berlin: Springer-Verlag.

Dauben, J. 1984. Conceptual revolutions and the history of mathematics: Two studies in the growth of knowledge. In *Revolutions in Mathematics*, ed. D. Gillies, 49–71. Oxford: Clarendon.

Davis, W. A. 1998. *Implicature: Intention, Convention and Principle in the Failure of Gricean Theory*. Cambridge: Cambridge University Press.

de Swart, H. 1976. Another intuitionistic completeness proof. *Journal of Symbolic Logic* 41: 644–664.

Delmas-Rigoutsos, Y. 1997. A double deduction system for quantum logic based on natural deduction. *Journal of Philosophical Logic* 26: 57–67.

Denyer, N. 1989. Dialetheism and trivialisation. *Mind* 98: 259–263.

Devlin, K. 1991. [1995]. *Logic and Information*, 2nd ed. Cambridge: Cambridge University Press.

Dickson, M. 2001. Quantum logic is alive $\wedge$ (it is true $\vee$ it is false). *Philosophy of Science* 68: S274–S287.

Doherty, A. 1998. Laura Goodship on Priest's principle R. *Australasian Journal of Philosophy* 76: 480–490.

Douglas, M. 1970. *Natural Symbols: Explorations in Cosmology*. London: Barrie & Jenkins.

Douglas, M. 1975. Self-evidence. In her *Implicit Meanings: Essays in Anthropology*, 276–318. London: Routledge.

Došen, K. 1989. Logical constants as punctuation marks. *Notre Dame Journal of Formal Logic* 30: 362–381.

Došen, K. 1997. Logical consequence: A turn in style. In *Logic and Scientific Methods*, eds. M. L. dalla Chiara, K. Doets, D. Mundici, and J. van Benthem, 289–311. Dordrecht: Kluwer.

Duhem, P. 1904. [1954]. *The Aim and Structure of Phycial Theory*. Trans. P. P. Weiner. Reprint, Princeton, N.J.: Princeton University Press.

Dummett, M. 1959a. [1978]. Truth. In his *Truth and Other Enigmas*, reprint, 1–24. London: Duckworth.

Dummett, M. 1959b. [1978]. Wittgenstein's philosophy of mathematics. In his *Truth and Other Enigmas*, reprint, 166–185. London: Duckworth.

Dummett, M. 1973a. [1981]. *Frege: Philosophy of Language*, 2nd ed. London: Duckworth.

Dummett, M. 1973b. [1978]. The philosophical basis of intuitionistic logic. In his *Truth and Other Enigmas*, reprint, 215–247. London: Duckworth.

Dummett, M. 1973c. [1978]. The significance of Quine's indeterminacy thesis. In his *Truth and Other Enigmas*, reprint, 375–419. London: Duckworth.

Dummett, M. 1973d. [1978]. The justification of deduction. In his *Truth and Other Enigmas*, reprint, 290–318. London: Duckworth.

Dummett, M. 1976a. [1978]. Is logic empirical? In his *Truth and Other Enigmas*, reprint, 269–289. London: Duckworth.

Dummett, M. 1976b. [1993]. What is a theory of meaning? (II). In his *The Seas of Language*, reprint, 34–93. Oxford: Clarendon.

Dummett, M. 1977. *Elements of Intuitionism*. Oxford: Clarendon.

Dummett, M. 1982. [1993]. Realism. In his *The Seas of Language*, reprint, 230–276. Oxford: Clarendon.

Dummett, M. 1991. *The Logical Basis of Metaphysics*. Cambridge, Mass.: Harvard University Press.

Dunmore, C. 1992. Meta-level revolutions in mathematics. In *Revolutions in Mathematics*, ed. D. Gillies, 209–225. Oxford: Clarendon.

Einstein, A. 1916. [1920]. *Relativity: The Special and General Theory. A Popular Exposition*. Trans. R. Lawson. Reprint, London: Methuen.

Empson, W. 1930. [1961]. *Seven Types of Ambiguity*. Reprint, Harmondsworth: Penguin.

Epstein, R. 1995. *The Semantic Foundations of Logics: Propositional Logics*, 2nd ed. New York: Oxford University Press.

Euler, L. 1758. Elementa doctrinæ solidorum. *Novi Commentarii Academiæ Scientiarum Petropolitanæ* 4: 109–140. Reprinted in his *Opera Omnia* (1) 26: 71–93.

Everett, A. 1994. Absorbing dialetheia? *Mind* 103: 413–419.

Feferman, S. 1989. [1994]. Finitary inductively presented logics. In *What Is a Logical System?*, reprint, ed. D. Gabbay, 297–328. Oxford: Clarendon.

Feyerabend, P. 1962. Explanation, reduction & empiricism. *Minnesota Studies in the Philosophy of Science* 3: 28–97.

Feyerabend, P. 1975. *Against Method*. London: Verso.

Fine, A. 1967. Consistency, derivability and scientific change. *Journal of Philosophy* 64: 231–240.

Fine, A. 1974. On the completeness of quantum theory. *Synthese* 29: 257–289.

Finkelstein, D. 1969. Matter, space and logic. *Boston Studies in the Philosophy of Science* 5: 199–215.

Foucault, M. 1966. [1970]. *Les Mots et les Choses*. Translated by A. Sheridan as *The Order of Things: An Archaelogy of the Human Sciences*. London: Routledge.

Foulis, D. 1997. A half century of quantum logic—what have we learned? In *Quantum Structures and the Nature of Reality*, eds. D. Aerts and J. Pykacz, 1–36. Dordrecht: Kluwer.

Frege, G. 1879. [1967]. *Begriffsschrift*. In *From Frege to Gödel: A Sourcebook in Mathematical Logic 1879–1931*, ed. J. van Heijenoort, 5–82. Translated by S. Bauer-Mengelberg. Cambridge, Mass.: Harvard University Press.

French, S., and N. C. A. da Costa. 2002. Inconsistency in science: A partial perspective. In *Inconsistency in Science*, ed. J. Meheus, 105–118. Dordrecht: Kluwer.

Gabbay, D. 1996. *Labelled Deductive Systems*, 1. Oxford: Clarendon Press.

Gabbay, D., and F. Guenthner, eds. 2001–. *Handbook of Philosophical Logic*, 2nd ed. Eighteen volumes. Berlin: Springer.

Gallier, J. 1991. Constructive logics. Part I: A tutorial on proof systems and typed λ-calculi. *Digital Paris Research Laboratory Reports* 8.

Gardner, M. 1971. Is quantum logic really logic? *Philosophy of Science* 38: 508–529.

Gentzen, G. 1933. [1969]. Über das Verhältnis zwischen intuitionistischer und klassischer Arithmetik. Translated by M. E. Szabo as "On the relation between classical and intuitionist arithmetic." In *The Collected Papers of Gerhard Gentzen*, ed. M. E. Szabo, 53–67. Amsterdam: North-Holland.

Gentzen, G. 1935. [1969]. Untersuchungen über das logische Schliessen. Translated by M. E. Szabo as "Investigations into logical deduction." In *The Collected Papers of Gerhard Gentzen*, ed. M. E. Szabo, 68–131. Amsterdam: North-Holland.

Gibbins, P. 1987. *Particles and Paradoxes: The Limits of Quantum Logic*. Cambridge: Cambridge University Press.

Gibbins, P., and D. Pearson. 1981. The distributive law in the two-slit experiment. *Foundations of Physics* 11.

Gillies, D. 1992. Introduction. In his *Revolutions in Mathematics*, 1–14. Oxford: Clarendon.

Gillies, D. 1993. *Philosophy of Science in the Twentieth Century: Four Central Themes*. Oxford: Blackwell.

Girard, J.-Y. 1991. A new constructive logic: Classical logic. *Mathematical Structures in Computer Science* 1: 255–296.

Goble, L., ed. 2001. *The Blackwell Guide to Philosophical Logic*. Oxford: Blackwell.

Gödel, K. 1933a. [1986]. Zur intuitionistischen Arithmetik und Zahlentheorie. Translated as "On intuitionistic arithmetic and number theory." In his *Collected Works* 1, ed. S. Feferman et al., 5–82. Oxford: Oxford University Press.

Gödel, K. 1933b. [1990]. Eine Interpretation des intuitionistischen Aussagenkalküls. Translated as "An interpretation of the intuitionistic propositional calculus." In his *Collected Works* 2, ed. S. Feferman et al., 300–303. Oxford: Oxford University Press.

Goldstein, L. 1992. Smooth and rough logic. *Philosophical Investigations* 15: 93–110.

Goodman, N. 1954. *Fact, Fiction and Forecast*. London: Athlone.

Goodship, L. 1996. On dialethism. *Australasian Journal of Philosophy* 76: 153–161.

Gray, J. 1992. The nineteenth-century revolution in mathematical ontology. In *Revolutions in Mathematics*, ed. D. Gillies, 226–248. Oxford: Clarendon.

Grice, P. 1975. [1991]. Logic and conversation. In *Conditionals*, reprint, ed. F. Jackson, 155–175. Oxford: Oxford University Press.

Griss, G. F. C. 1946. Negationless intuitionist mathematics. *Koninklijke Nederlandsche Akadamie van Wetenschappen* 49: 1127–1133.

Gutting, G. 1989. *Michel Foucault's Archaeology of Scientific Reason*. Cambridge: Cambridge University Press.

Haack, S. 1974. [1996]. *Deviant Logic: Some Philosophical Issues*. In her *Deviant Logic, Fuzzy Logic: Beyond the Formalism*, reprint with corrections, 1–177. Chicago: Chicago University Press.

Haack, S. 1976. [1996]. The justification of deduction. In her *Deviant Logic, Fuzzy Logic: Beyond the Formalism*, reprint, 183–191. Chicago: Chicago University Press.

Haack, S. 1978. *Philosophy of Logics*. Cambridge: Cambridge University Press.

Haack, S. 1996. *Deviant Logic, Fuzzy Logic: Beyond the Formalism*. Chicago: Chicago University Press.

Haack, S. 1998. *Manifesto of a Passionate Moderate*. Chicago: Chicago University Press.

Hallett, M. 1979. Towards a theory of mathematical research programmes. *British Journal for the Philosophy of Science* 30: 1–25, 135–159.

Hesse, M. 1968. Fine's criteria of meaning change. *Journal of Philosophy* 65: 46–52.

Hesseling, D. E. 2003. *Gnomes in the Fog: The Reception of Brouwer's Intuitionism in the 1920s*. Basel: Birkhäuser.

Heyting, A. 1930. Die formalen Regeln der intuitionistischen Logik. Die formalen Regeln der intuitionistischen Mathematik. *Sitzungsberichte der preußischen Akademie von Wißenschaften, physikalicsh-mathematische Klasse* 42–71; 311–327.

Heyting, A. 1956. [1983]. Disputation. In *Philosophy of Mathematics: Selected Readings*, reprint, eds. P. Benacerraf and H. Putnam, 66–76. Cambridge: Cambridge University Press.

Heywood, P., and M. Redhead. 1983. Nonlocality and the Kochen-Specker paradox. *Foundations of Physics* 13: 481–499.

Hilbert, D. 1926. [1983]. Über das Unendliche. Translated by E. Putnam and G. J. Massey as "On the infinite." In *Philosophy of Mathematics: Selected Readings*, eds. P. Benacerraf and H. Putnam, 183–201. Cambridge: Cambridge University Press.

Horgan, J. 1996. *The End of Science: Facing the Limits of Knowledge in the Twilight of the Scientific Age*. London: Little, Brown.

Hyde, D. 1997. From heaps and gaps to heaps of gluts. *Mind* 106: 641–660.

Jacquette, D. 1996. *Meinongian Logic: The Semantics of Existence and Nonexistence*. Berlin: de Gruyter.

Jaskowski, S. 1948. [1969] Rachunek zdań dla systemów dedukcyjnych sprzecznych. *Studia Societatis Scientarium Torunensis A* 1: 57–77. Translated as "Un calcul des propositions pour les systèmes déductifs contradictoires." *Studia Logica* 24: 143–157.

Jevons, W. S. 1876. *Logic*. London: Macmillan.

Johansson, I. 1936. Der Minimalkalkül, ein reduzierter intuitionistischer Formalismus. *Compositio Mathematicæ* 4: 119–136.

Johnson, R. H. 1996. *The Rise of Informal Logic*. Newport News, Va.: Vale.

Johnson, R. H., and J. A. Blair. 1997. Informal logic in the twentieth century. In *Historical Foundations of Informal Logic*, eds. D. Walton and A. Brinton, 158–177. Aldershot: Ashgate.

Kamlah, A. 1981. The connexion between Reichenbach's three-valued and von Neumann's lattice-theoretical quantum logic. *Erkenntnis* 16: 315–325.

Kant, I. 1992. *Lectures on Logic*. Trans. J. M. Young. Cambridge: Cambridge University Press.

Kirwan, C. 1978. *Logic and Argument*. London: Duckworth.

Kleene, S. C. 1952. *Introduction to Metamathematics*. Amsterdam: North-Holland.

Klein, F. 1893. A comparative review of recent researches in geometry. *Bulletin of the New York Mathematical Society* 2: 215–249.

Kneale, W., and M. Kneale. 1962. *The Development of Logic*. Oxford: Clarendon.

Kochen, S., and E. Specker. 1967. The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics* 17: 549–587.

Koetsier, T. 1991. *Lakatos's Philosophy of Mathematics*. Amsterdam: North-Holland.

Kolmogorov, A. N. 1925. [1967]. O printsipe tertium non datur. Translated as "On the principle of excluded middle." In *From Frege to Gödel: A Sourcebook in Mathematical Logic 1879–1931*, reprint, ed. J. van Heijenoort, 416–437. Cambridge, Mass.: Harvard University Press.

Körner, S. 1960. *The Philosophy of Mathematics*. London: Hutchinson.

Kreisel, H. 1965. Mathematical logic. In *Lectures on Modern Mathematics* 3, ed. T. L. Saaty, 95–195. New York: Wiley.

Kremer, M. 1989. Logic and meaning: The philosophical significance of the sequent calculus. *Mind* 98: 50–72.

Kripke, S. 1965. Semantical analysis of modal logic II. Non-normal modal propositional calculi. In *The Theory of Models*, ed. J.W. Addison et al., 206–220. Amsterdam: North-Holland.

Kripke, S. 1975. [1984]. Outline of a theory of truth. In *Recent Essays on Truth and the Liar Paradox*, reprint, ed. R. Martin, 53–81. Oxford: Clarendon.

Kuhn, T.S. 1962. [1970]. *The Structure of Scientific Revolutions*, 2nd ed. Chicago: Chicago University Press.

Kuhn, T.S. 1977. Second thoughts on paradigms. In *The Structure of Scientific Theories*, ed. F. Suppe, 459–482. Urbana: University of Illinois Press.

Lakatos, I. 1970. [1978]. Falsification and the methodology of scientific research programmes. In his *Philosophical Papers* 1, reprint, eds. J. Worrall and G. Currie, 8–101. Cambridge: Cambridge University Press.

Lakatos, I. 1971. [1978]. History of science and its rational reconstructions. In his *Philosophical Papers* 1, reprint, eds. J. Worrall and G. Currie, 102–138. Cambridge: Cambridge University Press.

Lakatos, I. 1976. *Proofs and Refutations: The Logic of Mathematical Discovery*, eds. J. Worrall and E. Zahar. Cambridge: Cambridge University Press.

Larvor, B. 1997. Lakatos as historian of mathematics. *Philosophia Mathematica* 5: 42–64.

Larvor, B. 1998. *Lakatos: An Introduction*. London: Routledge.

Laudan, L. 1977. *Progress and its Problems: Towards a Theory of Scientific Growth*. Berkeley: University of California Press.

Leplin, J. 1969. Meaning variance and the comparability of theories. *British Journal for the Philosophy of Science* 20: 69–75.

Levin, M.E. 1979. Quine's view(s) of logical truth. In *Essays on the Philosophy of W. V. Quine*, eds. R.W. Shahan and C. Swoyer, 45–67. Sussex: Harvester.

Lewis, C.I. 1914. The calculus of strict implication. *Mind* 23: 240–247.

Lewis, C.I. 1917. The issues concerning material implication. *Journal of Philosophy* 14: 350–356.

Lewis, C.I. 1932. [1968]. Strict implication. In *Logic and Philosophy: Selected Readings*, reprint, ed. G. Iseminger, 67–76. New York: Appleton-Century-Crofts. [reprinted from C.I. Lewis and C.H. Langford. 1932. *Symbolic Logic*].

Lewis, D. 1970. [1983]. General semantics. In his *Philosophical Papers* 1, reprint, 189–232. New York: Oxford University Press.

Lewis, D. 1976. [1983]. Survival and identity. In his *Philosophical Papers* 1, reprint, 55–77. New York: Oxford University Press.

Littman, G. 1991. *What problems does dialetheism pose for rationality?* Hon. diss., University of Queensland.

Lorenzen, P. 1955. *Einführung in die operative Logik und Mathematik*. Berlin: Springer.

MacColl, H. 1901. Letter to Bertrand Russell dated July 19th. Hamilton, Ontario: Russell Archives.

MacColl, H. 1906. *Symbolic Logic and its Applications*. London: Longmans, Green.

Maczynski, M.J. 1971. Boolean properties of observables in axiomatic quantum mechanics. *Reports on Mathematical Physics* 2: 135–150.

Makinson, D.C. 1973. *Topics in Modern Logic*. London: Methuen.

Malament, D. 1977. Causal theories of time and the conventionality of simultaneity. *Noûs* 11: 293–300.

Mares, E. 1996. Relevant logic and the theory of information. *Synthese* 109: 345–360.

Mares, E. 1997. Who's afraid of impossible worlds? *Notre Dame Journal of Formal Logic* 38: 516–526.

Mares, E. 2004. *Relevant Logic: A Philosophical Interpretation*. Cambridge: Cambridge University Press.

Marquis, J.-P. 1998. Epistemological aspects of the application of mathematics to itself. In *Philosophy and the Many Faces of Science*, eds. D. A. Anapolitanos, A. Baltas, and S. Tsinorema, 183–194. Lanham, MD: Rowman & Littlefield.

McCawley, J. D. 1981. *Everything that Linguists Have Always Wanted to Know about Logic** (**But Were Ashamed to Ask*). Oxford: Blackwell.

McKinsey, J. 1939. Proof of the independence of the primitive symbols of Heyting's calculus of propositions. *Journal of Symbolic Logic* 4: 155–158.

McKinsey, J., and A. Tarski. 1948. Some theorems about the sentential calculi of Lewis and Heyting. *Journal of Symbolic Logic* 13: 1–15.

Meheus, J. 1993. Adaptive logic in scientific discovery: The case of Clausius. *Logique et Analyse* 36: 359–391.

Meyer, R. 1978. Why I am not a relevantist. *A.N.U. Logic Group Research Paper* 1.

Meyer, R. 1985. Proving semantical completeness "relevantly" for R. *A.N.U. Logic Group Research Paper* 23.

Meyer, R. 1986. Sentential constants in R and $R^{\neg}$. *Studia Logica* 45: 301–327.

Meyer, R., and E. P. Martin. 1986. Logic on the Australian plan. *Journal of Philosophical Logic* 15: 305–332.

Meyer, R., and J. Slaney. 1989. Abelian logic (from A to Z). In *Paraconsistent Logic: Essays on the Inconsistent*, eds. G. Priest, R. Routley, and J. Norman, 245–289. Munich: Philosophia Verlag.

Moore, G. E. 1919. External and internal relations. *Proceedings of the Aristotelian Society* 20: 40–62.

Morrison, M. 1986. Quantum logic and the invariance argument—a reply to Bell and Hallett. *Philosophy of Science* 53: 403–411.

Mortensen, C. 1983. The validity of disjunctive syllogism is not so easily proved. *Notre Dame Journal of Formal Logic* 24: 35–40.

Morton, A. 1973. Denying the doctrine and changing the subject. *Journal of Philosophy* 70: 503–510.

Motterlini, M. 2002. Reconstructing Lakatos. A reassessment of Lakatos' epistemological project in the light of the Lakatos Archive. *Studies in the History and Philosophy of Science* 33: 487–509.

Nerlich, G. 1965. Presupposition and entailment. *American Philosophical Quarterly* 2: 33–42.

Newman, J. H. 1839. [1997]. Faith and reason, contrasted as habits of mind. In his *Fifteen Sermons Preached before the University of Oxford between A.D. 1826 and 1843 in the Definitive Third Edition of 1872*, 176–201. Reprint, Notre Dame, Ind.: University of Notre Dame Press.

Nishimura, I. 1980. Sequential method in quantum logic. *Journal of Symbolic Logic* 45: 339–350.

Nye, A. 1990. *Words of Power: A Feminist Reading of the History of Logic*. London: Routledge.

Parsons, T. 1980. *Nonexistent Objects*. New Haven, Conn.: Yale University Press.

Passmore, J. 1957. [1966]. *A Hundred Years of Philosophy*, 2nd ed. London: Penguin.

Peirce, C. S. 1896. The regenerated logic. *Monist* 7: 19–40.

Plantinga, A. 1974. *The Nature of Necessity*. Oxford: Clarendon.

Plumwood, V. 1993. The politics of reason: Towards a feminist logic. *Australasian Journal of Philosophy* 71: 436–464.
Pólya, G. 1945. [1957]. *How to Solve It: A New Aspect of Mathematical Method*, 2nd ed. London: Penguin.
Popper, K. 1947a. Logic without assumptions. *Proceedings of the Aristotelian Society* 47: 251–292.
Popper, K. 1947b. New foundations for logic. *Mind* 56: 193–235.
Prawitz, D. 1977. Meaning and proofs: On the conflict between classical and intuitionistic logic. *Theoria* 43: 2–40.
Prawitz, D. 1981. Philosophical aspects of proof theory. In *Philosophy of Language/Philosophical Logic*, eds. G. Fløistad and G. H. von Wright, 235–277. The Hague: Nijhoff.
Preston, J. 1997. *Feyerabend: Philosophy, Science and Society*. Cambridge: Polity.
Priest, G. 1987. *In Contradiction*. Dordrecht: Nijhoff.
Priest, G. 1989a. Classical logic aufgehoben. In *Paraconsistent Logic: Essays on the Inconsistent*, eds. G. Priest, R. Routley, and J. Norman, 131–148. Munich: Philosophia Verlag.
Priest, G. 1989b. Reductio ad absurdum et modus tollendo ponens. In *Paraconsistent Logic: Essays on the Inconsistent*, eds. G. Priest, R. Routley, and J. Norman, 613–626. Munich: Philosophia Verlag.
Priest, G. 1989c. Denyer's $ not backed by sterling arguments. *Mind* 98: 265–268.
Priest, G. 1990. Boolean negation and all that. *Journal of Philosophical Logic* 19: 201–215.
Priest, G. 1991. Minimally inconsistent LP. *Studia Logica* 50: 321–331.
Priest, G. 1992. What is a non-normal world? *Logique et Analyse* 35: 291–302.
Priest, G. 1993. Can contradictions be true? *Proceedings of the Aristotelian Society* Supp. 67: 35–54.
Priest, G. 1995. Gaps and gluts: Reply to Parsons. *Canadian Journal of Philosophy* 25: 57–66.
Priest, G. 1996a. Everett's trilogy. *Mind* 105: 631–647.
Priest, G. 1996b. On inconsistent arithmetics: A reply to Denyer. *Mind* 105: 648–659.
Priest, G. 2001. *An Introduction to Non-Classical Logic*. Cambridge: Cambridge University Press.
Priest, G. 2003. On alternative geometries, arithmetics and logics: A tribute to Łukasiewicz. *Studia Logica* 74: 441–468.
Priest, G., J. C. Beall, and B. Armour-Garb, eds. 2004. *The Law of Non-Contradiction: New Philosophical Essays*. Oxford: Oxford University Press.
Priest, G., and R. Routley. 1989a. Systems of paraconsistent logic. In *Paraconsistent Logic: Essays on the Inconsistent*, eds. G. Priest, R. Routley, and J. Norman, 151–186. Munich: Philosophia Verlag.
Priest, G., and R. Routley. 1989b. Applications of paraconsistent logic. In *Paraconsistent Logic: Essays on the Inconsistent*, eds. G. Priest, R. Routley, and J. Norman, 367–393. Munich: Philosophia Verlag.
Prior, A. 1960. The runabout inference ticket. *Analysis* 21: 38–39.
Putnam, H. 1962. [1975]. It ain't necessarily so. In his *Mathematics, Matter and Method: Philosophical Papers* 1, 237–249. Cambridge: Cambridge University Press.
Putnam, H. 1965. [1975]. A philosopher looks at quantum mechanics. In his *Mathematics, Matter and Method: Philosophical Papers* 1, 130–158. Cambridge: Cambridge University Press.

Putnam, H. 1969. [1975]. Is logic empirical? Reprinted as "The logic of quantum mechanics." In his *Mathematics, Matter and Method: Philosophical Papers* 1, 174–197. Cambridge: Cambridge University Press.

Putnam, H. 1974. How to think quantum logically. *Synthese* 29: 55–61.

Putnam, H. 1975. Introduction: Science as approximation to truth. In his *Mathematics, Matter and Method: Philosophical Papers* 1, vii–xiv. Cambridge: Cambridge University Press.

Putnam, H. 1983. Vagueness and alternative logic. In his *Realism and Reason: Philosophical Papers* 3, 271–286. Cambridge: Cambridge University Press.

Putnam, H. 1985. A quick Read is a wrong Wright. *Analysis* 45: 203.

Putnam, H. 1994. Comments and replies. In *Reading Putnam*, eds. P. J. Clark and R. L. V. Hale, 242–295. Oxford: Blackwell.

Quine, W. V. 1953. Reference and modality. In *From a Logical Point of View*, 139–159. Cambridge, Mass.: Harvard University Press.

Quine, W. V. 1970. *Philosophy of Logic*. Englewood Cliffs, N.J.: Prentice Hall.

Ramsey, F. 1927. Facts and propositions. *Proceedings of the Aristotelian Society* Supp. 7: 153–170.

Rasiowa, H. and R. Sikorski. 1953. Algebraic treatment of the notion of satisfiability. *Fundamenta Mathematicæ* 40: 62–95.

Rasmussen, S., and J. Ravnkilde. 1982. Realism and logic. *Synthese* 52: 378–437.

Rautenberg, W. 1987. Introduction. In his *Ω-Bibliography of Mathematical Logic* 1. New York: Springer Verlag.

Read, S. 1988. *Relevant Logic: A Philosophical Examination of Inference*. Oxford: Blackwell.

Read, S. 1995. *Thinking about Logic: An Introduction to the Philosophy of Logic*. Oxford: Oxford University Press.

Read, S. 1998. MacColl and strict implication. *Nordic Journal of Philosophical Logic* 3: 59–83.

Read, S. 2003. Logical consequence as truth-preservation. *Logique et Analyse* 183–184: 479–493.

Read, S. and C. Wright. 1985. Hairier than Putnam thought. *Analysis* 45: 56–58.

Redhead, M. 1987. *Incompleteness, Nonlocality and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics*. Oxford: Clarendon.

Redhead, M. 1994. Logic, quanta and the two-slit experiment. In *Reading Putnam*, eds. P. J. Clark and R. L. V. Hale, 161–175. Oxford: Blackwell.

Reichenbach, H. 1944. *Philosophic Foundations of Quantum Mechanics*. Berkeley: University of California Press.

Rescher, N. 1973. *The Primacy of Practice*. Oxford: Blackwell.

Rescher, N. and R. Brandom. 1980. *The Logic of Inconsistency*. Oxford: Blackwell.

Resnik, M. 1985. Logic: Normative or descriptive? The ethics of belief or a branch of psychology? *Philosophy of Science* 52: 221–238.

Resnik, M. 1996. Ought there to be but one logic? In *Logic and Reality: Essays on the Legacy of Arthur Prior*, ed. B. J. Copeland, 489–517. Oxford: Clarendon.

Restall, G. 1996. Information flow and relevant logics. In *Logic, Language and Computation* 1, eds. J. Seligman and D. Westerståhl, 463–477. Stanford, Calif.: C.S.L.I.

Routley, R. 1981. *Exploring Meinong's Jungle and Beyond*. Canberra: ANU Press.

Routley, R. 1984. Relevantism, material detachment and the disjunctive syllogism argument. *Canadian Journal of Philosophy* 14: 167–188.

Routley, R. and R. Meyer. 1973. The semantics of entailment I. In *Truth, Syntax and Modality*, ed. H. Leblanc, 199–243. Amsterdam: North-Holland.

Routley, R., R. Meyer, V. Plumwood, and R. Brady. 1982. *Relevant Logics and Their Rivals* 1*: The Basic Philosophical and Semantical Theory*. Atascadero, Calif.: Ridgeview.

Routley, R., and V. Routley. 1972. The semantics of first degree entailment. *Noûs* 6: 335–359.

Routley, R., V. Routley, R. Meyer, and E. Martin. 1982. On the philosophical bases of relevant logic semantics. *Journal of Non-Classical Logic* 1: 71–105.

Russell, B. 1903. *Principles of Mathematics*. Cambridge: Cambridge University Press.

Russell, B. 1905. On denoting. *Mind* 14: 479–493.

Sambin, G. 2002. Steps towards a dynamical constructivism. In *Proceedings of the XI International Congress of Logic, Methodology and Philosophy of Science, Krakow, August 1999*, eds. P. Gärdenfors, K. Kijania-Placek, and J. Wolenski, 261–284. Dordrecht: Kluwer.

Sambin, G., G. Battilotti, and C. Faggian. 2000. Basic logic: Reflection, symmetry, visibility. *Journal of Symbolic Logic* 65: 979–1013.

Sarkar, T. K. 1990. Some systems of deviant logic: A unifying approach. In *Foundations of Logic and Language: Studies in Philosophical and Non-Standard Logic*, ed. P. K. Sen, 122–181. Calcutta: Allied.

Schotch, P. and R. Jennings. 1989. On detonating. In *Paraconsistent Logic: Essays on the Inconsistent*, eds. G. Priest, R. Routley, and J. Norman, 306–327. Munich: Philosophia Verlag.

Schroeder-Heister, P. 1984. Popper's theory of deductive inference and the concept of a logical constant. *History and Philosophy of Logic* 5: 79–110.

Schwartz, S. 1987. Intuitionism and sorites. *Analysis* 47: 179–183.

Seidel, P. L. 1848. Note über eine Eigenschaft der Reihen, welche discontuirliche Functionen darstellen. *Abhandlungen der mathematisch-physikalischen Classe der Königlich-Bayerischen Akademie der Wissenschaften* 5: 379–393.

Shoesmith, D. J., and T. Smiley. 1978. *Multiple-Conclusion Logic*. Cambridge: Cambridge University Press.

Sklar, L. 1992. *Philosophy of Physics*. Boulder, Colo.: Westview.

Slater, B. H. 1995. Paraconsistent logics? *Journal of Philosophical Logic* 24: 451–454.

Smiley, T. 1960. Sense without denotation. *Analysis* 20: 125–135.

Smiley, T. 1993. Can contradictions be true? *Proceedings of the Aristotelian Society* Supp. 67: 17–33.

Sorensen, R. 1999. Mirror notation: Symbol manipulation without inscription manipulation. *Journal of Philosophical Logic* 28: 141–164.

Stalnaker, R. 1968. [1991]. A theory of conditionals. In *Conditionals*, reprint, ed. F. Jackson, 28–45. Oxford: Oxford University Press.

Strawson, P. F. 1952. *Introduction to Logical Theory*. London: Methuen.

Takeuti, G. 1975. *Proof Theory*. Dordrecht: North-Holland.

Tarski, A. 1930. Fundamentale Begriffe der Methodologie der deduktiven Wissenschaften, I. *Monatshefte für Mathematik und Physik* 37: 361–404. English translation in his 1956 *Logic, Semantics, Metamathematics*, trans. J. H. Woodger, 60–109. Oxford: Clarendon.

Tarski, A. 1936. Uber den Begriff der logischen Forschung. *Actes du Congrès International de Philosophie Scientifique* 7: 1–11. English translation in his 1956 *Logic, Semantics, Metamathematics*, trans. J. H. Woodger, 409–420. Oxford: Clarendon.

Tennant, N. 1996. Delicate proof theory. In *Logic and Reality: Essays on the Legacy of Arthur Prior*, ed. B. J. Copeland, 351–385. Oxford: Clarendon.

Tennant, N. 1997. *The Taming of the True*. Oxford: Clarendon.

Thagard, P. 1982. From the descriptive to the normative in psychology and logic. *Philosophy of Science* 49: 24–42.

Thomason, R. 1986. Paradoxes and semantic representation. In *Theoretical Aspects of Reasoning about Knowledge*, ed. J. Halpern. San Francisco: Morgan Kaufman.

Toulmin, S. 1970. Does the distinction between normal and revolutionary science hold water? In *Criticism and the Growth of Knowledge*, eds. I. Lakatos and A. Musgrave, 39–47. Cambridge: Cambridge University Press.

Troelstra, A. S. 1990. Introductory note [to Gödel 1933b]. In *Collected works* 2, ed. S. Feferman et al., 296–299. Oxford: Oxford University Press.

Troelstra, A. S. 1992. *Lectures on Linear Logic*. Stanford, Calif.: C.S.L.I.

Ungar, A. M. 1992. *Normalization, Cut-elimination and the Theory of Proofs*. Stanford, Calif.: C.S.L.I.

van Benthem, J. 1994. General dynamic logic. In *What Is a Logical System?*, ed. D. Gabbay, 107–139. Oxford: Clarendon.

van Dalen, D. 1986. Intuitionistic logic. In *Handbook of Philosophical Logic* 3, eds. D. Gabbay and F. Guenthner, 225–339. Dordrecht: Reidel.

van Fraassen, B. 1966. Singular terms, truth-value gaps and free logic. *Journal of Philosophy* 63: 481–495.

van Fraassen, B. 1973. Semantic analysis of quantum logic. In *Contemporary Research in the Foundations and Philosophy of Quantum Theory*, ed. C. A. Hooker. Dordrecht: Reidel.

van Fraassen, B. 1974. [1975]. The labyrinth of quantum logics. In *The Logico-Algebraic Approach in Quantum Mechanics* 1, reprint, ed. C. A. Hooker, 577–607. Dordrecht: Reidel.

Veldman, W. 1976. An intuitionistic completeness theorem for intuitionistic predicate logic. *Journal of Symbolic Logic* 41: 159–166.

von Neumann, J. 1932. [1996]. *Mathematical Foundations of Quantum Mechanics*. Princeton, N.J.: Princeton University Press. Translation of *Grundlagen der Quantenmechanik*. Berlin: Springer Verlag.

von Plato, J. 1994. *Creating Modern Probability: Its Mathematics, Physics and Philosophy in Historical Perspective*. Cambridge: Cambridge University Press.

Walton, D. 1996. *Argument Structure: A Pragmatic Theory*. Toronto: University of Toronto Press.

Walton, D. 1998. *The New Dialectic: Conversational Contexts of Argument*. Toronto: University of Toronto Press.

Watts, I. 1724. [1762]. *Logick, or the Right Use of Reason in the Inquiry after Truth*, 11th ed. Edinburgh: Donaldson.

Weingartner, P., ed. 2004. *Alternative Logics: Do Sciences Need Them?* Berlin: Springer.

Weir, A. 1986. Classical harmony. *Notre Dame Journal of Formal Logic* 27: 459–482.

Weyl, H. 1918. [1960]. *Das Kontinuum*. Translated as *The Continuum*. New York: Chelsea.

Whately, R. 1826. [1850]. *Logic*, 2nd ed. London: Griffin.

Williamson, T. 1994. *Vagueness*. London: Routledge.

Wittgenstein, L. 1921. [1961]. *Tractatus Logico-Philosophicus*. Trans. D. Pears and B. McGuinness. London: Routledge.

Wright, C. 1982. [1992]. Realism, bivalence and classical logic. In his *Realism, Meaning and Truth*, reprint, 458–478. Oxford: Blackwell.

Wright, C. 1992. *Truth and Objectivity*. Cambridge, Mass.: Harvard University Press.

Zalta, E. N., ed. 1995–. *Stanford Encyclopedia of Philosophy*. Stanford, Calif.: C.S.L.I. http://plato.stanford.edu.

Ziman, J. 1985. Pushing back frontiers—or redrawing maps! In *The Identification of Progress in Learning*, ed. T. Hägerstrand, 1–12. Cambridge: Cambridge University Press.

15

Philosophy of Inductive Logic: The Bayesian Perspective

Sandy Zabell

1. Introduction

Inductive inference dates from Aristotle, but resisted precise quantitative formulation until the rise of mathematical probability. In the three-and-a-half centuries since the birth of mathematical probability in 1654, mathematicians, statisticians, and philosophers have been fascinated by its deep implications for inductive inference. By the time of Laplace, a rich model for inductive inference emerged that sometimes elicited, in the hands of the less capable, excesses of approbation and aversion for more than a century. In the twentieth century the theory underwent a radical transformation at the hands of Frank Ramsey, Bruno de Finetti, and their followers, and the modern theory of subjective probability emerged, complete with a detailed model of the inductive process.

This chapter describes the logic of inductive inference as seen through the eyes of the modern theory of personal probability, including a number of its recent refinements and extensions. The structure of the chapter is as follows. After a brief discussion of mathematical probability, to establish notation and terminology, the gradual evolution of the probabilistic explication of induction from Bayes to the present is recounted. Our interest is not in this history per se (fascinating as it is), but in its use to highlight the key assumptions, criticisms, refinements, and achievements of that theory. Along the way, the structure of the modern theory is presented, and its relation to the problem of induction discussed.

1.1. Probability

Mathematical probability is a model of uncertainty, uncertainty about the past or the future. Like all mathematical models, it consists of a structure constrained by a set of assumptions termed axioms. The intended heuristic interpretation of the mathematical model motivates further definitions and permits us to formulate conjectures regarding consequences of the axioms. There is a generally agreed mathematical formulation of this, dating from Kolmogorov (1933). A *probability space* is a triple $(\Omega, \mathcal{A}, P)$, consisting of a set Ω (the *sample space* of atomic possibilities), a collection $\mathcal{A}$ of subsets of Ω (the *events* of interest), and a set function P assigning to each event a number between 0 and 1 (the *probability* of the event).

One interprets this model as follows. The sample space Ω represents the set of possibilities for an uncertain outcome; for example, if we are rolling a six-sided die, one may take Ω to be the set $\{1, 2, 3, 4, 5, 6\}$. The subsets of the sample space represent particular events of interest; for example, the event "rolling the die resulted in an even number" corresponds to the set $\{2, 4, 6\}$. (It is a useful, and for our purposes harmless abuse of terminology to blur the distinction between a linguistic proposition, an empirical event such a proposition might describe, and the subset of the sample space corresponding to such an event. For our purposes events, or the propositions describing them, *are* subsets of the sample space.) The event A is said to *occur* if the uncertain outcome ω lies in A: thus $\omega \in A \subseteq \Omega$. The probability of an event is most commonly interpreted as representing either an epistemic quantity (such as a degree of belief regarding the occurrence of the event) or an aleatory one (such as its frequency of occurrence in a population or sequence of trials).

The sample space Ω may be any set whatever, countable or uncountable. The *Kolmogorov axioms* make certain assumptions regarding the collection of events $\mathcal{A}$ and the probability function P. The collection of events is assumed to be an *algebra*: Thus, if $\emptyset$ represents the empty set (the *impossible event*), A^c the *complement* of an event ($\Omega - A$, the elements of Ω not in A), and $A \cup B$, $A \cap B$ respectively the union and intersection of A and B (the elements of Ω that lie in either or both A and B, respectively), then it is assumed that (a) both $\emptyset$ and Ω are elements of $\mathcal{A}$; (b) if $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$; and (c) if $A, B \in \mathcal{A}$, then $A \cup B$, $A \cap B \in \mathcal{A}$. (These assumptions as stated are deliberately redundant: if $\emptyset \in \mathcal{A}$, then $\Omega = \emptyset^c \in \mathcal{A}$; and if $A \cup B \in \mathcal{A}$, then $A \cap B = (A^c \cup B^c)^c \in \mathcal{A}$.)

Finally, there are the axioms that the probability function P satisfies. These are:

1. If $A \in \mathcal{A}$, then $P(A) \geq 0$.
2. If $A, B \in \mathcal{A}$ and $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.
3. $P(\Omega) = 1$.

That is, the function P is assumed to be a nonnegative, additive set function assigning a unit measure to the set Ω.

If the algebra $\mathcal{A}$ satisfies the further property that it is closed under *countable* unions (and therefore, by De Morgan's laws, under countable intersections as well), that is, if $A_1, A_2, \ldots \in \mathcal{A}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$, then the algebra $\mathcal{A}$ is said to be a *σ-algebra*. If the probability function is also assumed to satisfy the property that

$$P\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i) \quad \text{provided } A_i \cap A_j = \emptyset,\ 1 \leq i < j < \infty,$$

(that is, if the sets $A_1, A_2, \ldots$ are all *pairwise disjoint*), then the probability function P is said to be *countably additive*, and otherwise *finitely additive*. (Strictly speaking, a finitely additive probability measure is one that is only assumed to be finitely additive, but could be countably additive; just as in logic a finitely satisfiable set of wffs might in fact be satisfiable.) Most mathematicians assume that the algebra $\mathcal{A}$ is a σ-algebra, and the probability function P countably additive; but some subjectivists prefer to assume only that $\mathcal{A}$ is an algebra, and P finitely additive. We assume the former.

1.2. The Nature of Probability

> "When I use a word," Humpty Dumpty said in rather a scornful tone, "it means just what I choose it to mean—neither more or less." "The question is," said Alice, "whether you can make words mean different things." "The question is," said Humpty Dumpty, "which is to be master—that's all." (Lewis Carroll, *Alice in Wonderland*)

The choices that are made in the interpretation of a probability model reflect the differing views of what probability is or might be. To begin, probabilities can be either *aleatory* or *epistemic*; that is, they are either properties of things (objects or events, chance phenomena), or relate to our knowledge about things. Both aleatory and epistemic probabilities come in different flavors. Aleatory probabilities can be either propensities or frequencies; and if frequencies, either finite or infinite. Such probabilities have only an indirect relation to inductive inferences: inductive inferences can be *about* aleatory phenomena, but are not framed in terms of them. Epistemic probabilities can be either qualitative (Keynes, Koopman) or quantitative (Carnap, Ramsey, de Finetti); and in either case can be further categorized as being either objective (Keynes, Carnap) or subjective (Koopman, Ramsey, de Finetti); that is, they are either a measure of the rational degree of belief a person *should have* given their state of knowledge (sometimes termed *credibilities*); or they can vary from one person to another. In the later case they are either *personal* or *psychological*, that is, they either satisfy certain rationality constraints (without being uniquely determined), or do not. Personal, subjective, quantitative, epistemic probabilities are natural descriptors of inductive inferences. We refer to them simply as subjective probabilities.

But whether one believes in either, both, or neither epistemic or aleatory probability, we are incomparably better off than our forebears, for today these distinctions are clear. There is a natural reason why such distinctions were not initially noted. In many of the initial examples considered, such as tossing a (fair) coin, or rolling a die, or picking a card from a deck, every interpretation gives rise to the same numerical value. For example, if one tosses a fair coin, the probability is 50–50 that it will come up heads because this is what the classical definition tells us, but also because (a) we think there is an equal propensity for the coin to come up either heads or tails, (b) we find that the coin does come up approximately half the time in a finite number of tosses, (c) our knowledge gives us no reason to think heads any more or less likely to occur than tails, and (d) it is even odds in a bet that one or the other occurs. These represent important conceptual distinctions, but given that they all yield the same numerical value of 1/2, understanding the difference between them may seem less than pressing. It is only when one attempts to extend the ambit of probability to situations where the different meanings have implications for the specific numerical value attached, that the distinctions become important.

1.3. Inductive Inference

The distinction between logical and scientific inference goes back to Aristotle, and are the subjects, respectively, of the *Prior* and *Posterior Analytics*. But inductive, in the modern sense of uncertain, inference fell in yet another category for Aristotle—the realm of rhetoric; the result was belief (*endoxos*) rather than knowledge (*episteme*).

The Academic skeptics later developed a theory of rational decision based on probable knowledge. Carneades in particular developed a scale of conviction, ranging from the "credible," to the "credible and consistent," to the "credible, consistent, and tested." Carneades's theory is in effect an early attempt at developing a theory of qualitative or comparative subjective probability in the modern sense. Witness Cicero's statement in the *Academica* (2.110):

> If a question be put to the wise man about duty or about a number of other matters in which practice has made him an expert, he would not reply in the same way as he would if questioned as to whether the number of the stars is even or odd, and say that he did not know; for in things uncertain there is nothing probable [*in incertis enim nihil est probabile*], but in things where there is probability the wise man will not be at a loss either what to do or what to answer. [Quoted from the Loeb classical edition]

It is interesting to contrast Cicero's statement with that of Nicole Oresme, the Renaissance astronomer and mathematician (c. A.D. 1325–1382), who wrote some 14 centuries later:

> The number of stars is even; the number of stars is odd. One of these statements is necessary, the other impossible. However, we have doubts as to which is necessary, so that we say of each that it is possible. . . . The number of stars is a cube. Now indeed, we say that it is possible, but not, however, probable or likely [*non tamen probabile aut opinabile aut verisimile*], since such numbers are much fewer than others. . . . The number of stars is not a cube. We say that it is possible, probable, and likely. (Oresme 1966)

Some might regard the significance of this passage to be its quasi-numerical assignments of degrees of likelihood to assertions about the number of stars; while others might argue that the significance of the passage lies in its willingness to extend the scope of the theory to propositions such as the parity of the number of stars (we no longer say just that we do not know, but place this on a scale).

1.4. James Bernoulli and the "Art of Conjecture"

Such examples illustrate just how purely qualitative inductive inference was prior to the seventeenth century. This inevitably changed with the birth of mathematical probability in 1654, the date of the famous correspondence between Pascal and Fermat. The first great step in the explication of induction using this "new math" was due to James Bernoulli (1654–1705) in his posthumous *Ars conjectandi* of 1713. In modern terms, Bernoulli proved that in a sequence of n independent trials, if the probability of some event is p, and the event is observed S_n times (the number of "successes"), then for any positive number $\epsilon > 0$, one has

$$\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) = 0.$$

In words, the probability that the empirically observed frequency S_n/n will differ from the true probability p by *any* amount (the "for any $\epsilon > 0$"), will become negligible for sufficiently large sample sizes n.

Although Bernoulli's proof of his theorem cannot be mathematically faulted, its interpretation and relevance to the problem of induction could; something Bernoulli himself was certainly aware of (since he proved his theorem around 1685, but did not publish either it or its applications in his own lifetime); see Hacking (1971).

But while Pascal, Fermat, and their immediate successors (in particular Leibniz and the Bernoullis) were certainly aware of the potential philosophical significance of the mathematics of the "doctrine of chances" for the problem of induction, it was not until an essay by an obscure English cleric that a systematic and principled attack on the problem occurred.

2. Bayes's Essay

In 1764, a paper with the modest title "An essay towards solving a problem in the doctrine of chances" appeared in the pages of the *Philosophical Transactions of the Royal Society of London* (Bayes 1764). Its publication was posthumous; for its author, an obscure English cleric, the Reverend Thomas Bayes, had, like James Bernoulli before him, perished before he was published.

Bayes's essay is a remarkable one. Although Bayes was a member of the Royal Society of London, little was known about him until quite recently. Thanks however to the efforts of modern scholars combing through the appropriate archives, a considerable body of information has now come to light; see in particular Dale (2003) and Bellhouse (2004). The adjective "ingenious" was applied to Bayes by a number of contemporaries who knew him or his work, and indeed the essay itself is not only ingenious but even today rewards close study. Many of the issues that later arose in discussions of inductive probability involved issues that Bayes had already grappled with and attempted to address. Aware of the challenging nature of the question he was considering, Bayes revised his essay in a major way at least once, and even then appears to have been sufficiently uncertain of some of his conclusions that he withheld it from publication for some 12 years, until his death in 1761. It was therefore left to Bayes's friend, fellow dissenting clergyman and executor the Reverend Dr. Richard Price, to see to the publication of the essay.

Price submitted Bayes's essay to the Royal Society on November 10, 1763. In addition to the essay proper, Price included both a cover letter describing the origin of the essay, and an appendix discussing the applications of Bayes's results to the problem of induction. Both of these supplementary documents are of considerable interest in themselves. (The cover letter is in fact our only source of information concerning the path that Bayes took. Bayes had written an introduction to the essay which Price repeatedly refers to in his letter, but Price did not include the introduction when he submitted the essay for publication, and it now appears to be lost.)

2.1. The Beginning of the Essay

The first section of Bayes's essay begins with several definitions; the basic properties of probabilities are then deduced from these. For Bayes the probability of an event is subjective: a ratio of expected gains. Price tells us:

> He has also made an apology for the peculiar definition he has given of the word *chance* or *probability*. His design herein was to cut off all dispute about the meaning of the word, which in common language is used in different senses by persons of different opinions, and according as it is applied to *past* or *future* facts.

This is already impressive. Bayes recognizes the confusion introduced by a failure to give a precise definition of probability, and the need to deduce any properties one uses from such a definition. Shafer (1982) has a very useful discussion of this part of the essay. For our purposes, however, there is little loss in interpreting Bayes's probability for the moment as probability in the modern subjective sense, and accepting that the standard properties of such probabilities can be deduced from its definition. (We return to this question when we come to Ramsey and de Finetti and discuss the precise modern definition.) One of Bayes's results, however, is worth noting here: the formula for a conditional probability. Bayes presents this as a corollary to his third proposition; nowadays it is usually introduced as a definition. It is discussed further on in this chapter.

The Definition of Conditional Probability If A and B are events, and $P(B) > 0$, then the *conditional probability of* A *given* B is defined to be

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

This definition links three quantities, and one can pass from any two to the third. Although this has now been universally adopted, its precise justification has been much debated (see, e.g., Hacking 1967, as will be discussed later).

Let us turn to the heart of Bayes's essay, inductive inference. Price tells us:

> In an introduction which he has writ to this Essay, [Bayes] says, that his design at first in thinking on the subject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumstances, upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times.

Upon supposition that we know nothing concerning it.... This crucial phrase was at the center of much of the subsequent debate surrounding the mathematical model of inductive inference for more than a century and a half; until, that is to say, the writing of Ramsey's essay (itself posthumously published!) in 1926. Precisely what does it mean to say that we are "entirely ignorant of an event" (as Bayes later puts it) except for knowledge of its observed frequency of occurrence? Price tells us

> [Bayes] adds, that he soon perceived that it would not be very difficult to do this, provided some rule could be found according to which we ought to estimate the chance that the probability for the happening of an event perfectly unknown, should lie between two named degrees of probability, antecedently to any experiments made about it.

Provided some rule.... Therein lies the rub: Can one quantify the uncertainty regarding the initial probability of an event, at least in the circumstances described (of complete ignorance)? Several things are happening here. First, there is a passage from the *probability* of an event, say p, to the *chance* that it lies between two values. One would be tempted to take the use of the terms "chance" and "probability" to imply a conceptual distinction (say, between a credibility and a propensity) were it not for the fact that at the beginning of the essay Bayes explicitly states that "by chance I mean the same as probability." Perhaps two different terms are being used to subtly emphasize that although both quantities are probabilities, one is the probability of an event, the other the probability of a probability. (In modern terms, Bayes is referring to the *prior* or *initial* distribution of p. I. J. Good distinguishes between the two by referring to a "type 1" probability and a "type 2" probability.)

What was Bayes's solution to this central problem of inductive inference? Price goes on to tell us

> It appeared to him that the rule must be to suppose the chance the same that it should lie between any two equidifferent degrees; which, if it were allowed, all the rest might be easily calculated in the common method of proceeding in the doctrine of chances.

This is the famous assumption of a *uniform prio*r: The chance that p lies in an interval is uniform throughout: If $\{ x : a \leq x \leq b \}$ is a subinterval of the unit interval, then $P(a \leq p \leq b) = b - a$; the probability of p lying in the subinterval is equal to the interval's length. (For example, if one interval is twice the length of another, it is twice as likely that p falls in the first interval than the second.) Throughout the next century and a half (and beyond), this postulate was debated, praised, reviled, misunderstood, misprepresented, misstated, and misapplied. But what is often overlooked is that Bayes himself had substantial reservations about the assumption (at least as a first principle), and later on went to some lengths to justify it. As Price tells us:

> But [Bayes] afterwards considered, that the *postulate* on which he had argued might not perhaps be looked upon by all as reasonable; and therefore he chose to lay down in another form the proposition in which he thought the solution of the problem is contained, and in a *scholium* to subjoin the reasons why he thought so, rather than take into his mathematical reasoning any thing that might admit dispute.

Thus Bayes did not start from the simplistic assumption of a uniform prior! Indeed, he revised his essay specifically to eliminate it. Instead, the structure of Baye's "ingenious" argument is much more subtle; it proceeds in three carefully considered stages:

- A physical randomization is performed.
- The probabilities of the resulting events are described.
- One passes by analogy from the physical randomization to the actual event of interest.

Let us consider each of these in turn.

2.2. The "Billiards" Table

In his analysis Bayes considered first the case of a ball (W) thrown on a two-dimensional table and coming to rest, there being "the same probability that it rests on any one equal part of the plane as another." A line is then drawn through the ball perpendicular to the lower horizontal edge of the table, bisecting the table. Another ball ("O") is then thrown and the event M is said to occur or not depending on whether the ball falls to the right or left of the line through the original ball W.

Much of Bayes's analysis is given in terms of the resulting positions of the balls when projected onto the lower edge of the table. There is in fact no loss in considering the following simplification of Bayes's original model:

- a point is chosen "at random" (uniformly) from the unit interval;
- n further points $x_1, \ldots, x_n$ are then selected at random from the interval;
- the event is said to occur on the ith trial if $x_i \leq x$.

Why did Bayes choose his two-dimensional model rather than the simpler one-dimensional one? Presumably the concept of choosing a point at random from the plane seemed more natural than choosing a point at random from the line, since the former could be modeled, for example, by rolling a ball on a table, such as one uses to play billiards. (Although Bayes's hypothetical table is often referred to in later literature as a billiards table, the term itself does not occur in his essay.)

In the case of Bayes's simple model of throwing balls onto a table, the answer to his question regarding the probability of an event occurring is a simple and immediate application of the calculus of probabilities (and a standard computation today): If X denotes the value of the unknown probability, then given that the event has occurred k times out of n in the past, the chance that its probability p falls between the limits a and b is

$$P(a \leq X \leq b) = (n+1)\binom{n}{k} \int_a^b p^k (1-p)^{n-k}\, dp.$$

The integral is an example of what is termed an *incomplete beta function*, and is readily computed today using either tables or any of a variety of mathematical software packages. Bayes of course did not have the advantage of these, and his essay concludes with a discussion of the approximate evaluation of such integrals.

In modern terms, Bayes has determined the *posterior* or *final* distribution of the unknown quantity X: it has the probability density

$$(n+1)\binom{n}{k} p^k (1-p)^{n-k};$$

it integrates to one because of the integration formula

$$\int_0^1 p^k (1-p)^{n-k}\, dp = \frac{1}{(n+1)\binom{n}{k}}.$$

Note Bayes's argument does not depend on the billiard ball falling uniformly on the table. Let X denote (the random) the position of the initial ball, let $F_X(x) = P(X \leq x)$, and let $Y = F_X(X)$; note that the range of Y is the unit interval. If the function $F_X(x)$ is continuous, then it can be shown that the random variable Y has a uniform distribution on the unit interval. The value of X corresponds to the throwing of the first ball in Bayes's argument; the probability p that other balls fall to the left of it is $F_X(X)$; thus the distribution of p is uniform provided only that the ball falls continuously on the table. See Fisher (1973, chapter 5, section 6) for some discussion of this point.

2.3. Bayes's Scholium

We now come to the crux of Bayes's argument. He has determined the posterior distribution of the unknown probability p in the case of a specific physical randomization model. Under his model, what is the chance that in n trials one will observe k "successes" (occurrences of the event) and $n - k$ "failures" (nonoccurrences of the event)? Let S_n denote the number of successes in n trials. The answer is readily computed to be

$$P(S_n = k) = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k}\, dp = \frac{1}{n+1}.$$

That is, the number of successes is equally likely to be 0 or 1 or 2 . . . up to n.

Now Bayes turns from his physical randomization model to the case of "an event concerning the probability of which we absolutely know nothing antecedently to any trials made concerning it," and reasons as follows. To get anywhere, one must define exactly what one means by such ignorance, or at least characterize it. Now one could directly take the prior or initial chance distribution for p to be the uniform distribution on the unit interval, just as in the case of the physical randomization model. But Bayes realizes that there are vulnerabilities in such an approach; indeed, vulnerabilities that others would debate for the next two centuries. So he takes a different tack. Let us regard the prior distribution for p, say $d\mu(p)$ as yet unknown, and to be determined by less controversial properties of it that we hope to discover.

Bayes argues that for such a generic event, to "absolutely know nothing antecedently to any trials made concerning it" *means* that in n trials, one thinks that the number of successes is equally likely to be any value between 0 and n.

Bayes's Assumption

$$P(S_n = k) = \frac{1}{(n+1)}, \quad 0 \leq k \leq n.$$

To quote Bayes, "concerning such an event I have no reason to think that, in a certain number of trials, it should rather happen any one possible number of times than another." That is, Bayes adopts this as his *definition* of what it means to be in a state of absolute ignorance. This will be true for each $n \geq 1$;

thus one must have

$$P(S_n = k) = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k}\, d\mu(p) = \frac{1}{n+1}, \quad n \geq 1, \quad 0 \leq k \leq n.$$

But this is true of the uniform prior distribution dp, and therefore, Bayes argues, one must have $d\mu(p) = dp$. In his words:

> And that the same rule [as the one for the physical randomization model] is the proper one to be used in the case of an event concerning the probability of which we absolutely know nothing antecedently to any trials made concerning it, seems to appear from the following consideration; viz. that concerning such an event I have no reason to think that, in a certain number of trials, it should rather happen any one possible number of times than another. For, on this account, I may justly reason concerning it as if its probability had been at first unfixed, and then determined in such a manner as to give me no reason to think that, in a certain number of trials, it should rather happen any one possible number of times than another. But this is exactly the case of the event M [the event involving the ball on the table].

It is in fact possible to make Bayes's intuition rigorous: According to the Hausdorff moment theorem, the moments of a finite measure on a bounded interval uniquely determine the measure. Since in this case one has that

$$\int_0^1 p^n\, d\mu(p) = \frac{1}{n+1} = \int_0^1 p^n\, dp, \quad n \geq 1$$

(the first equality by assumption, the second by calculus), it follows that the two measures have the same moments and therefore must coincide: $d\mu(p) = dp$.

2.4. Price's Appendix

Price correctly recognized both the importance of Bayes's essay, and its application to the problem of inductive inference. He states in his cover letter:

> Every judicious person will be sensible that the problem now mentioned is by no means merely a curious speculation in the doctrine of chances, but necessary to be solved in order to [provide] a sure foundation for all our reasonings concerning past facts, and what is likely hereafter. ... It is certain that we cannot determine, at least not to any nicety, in what degree repeated experiments confirm a conclusion, with the particular discussion of the beforementioned problem [addressed by Bayes]; which, therefore, is necessary to be considered by any one who would give a clear account of the strength of *analogical* or *inductive reasoning*.

2.4.1. Price's Rule of Succession

Price begins by illustrating Bayes's results with some examples. In particular, he considers the case when only successes are observed. In this case the chance that the probability p is found in an interval $a \leq p \leq b$ is

$$P(a \leq p \leq b) = (n+1)\int_0^1 p^n \, dp = b^{n+1} - a^{n+1}.$$

Gillies (1987, 332) calls this *Price's rule of succession*, to distinguish it from *Laplace's rule of succession*, discussed later.

In particular, the "odds that it is somewhat more than an even chance" that the event will occur on the next trial is

$$P\left(\frac{1}{2} \leq p \leq 1\right) = 1 - \left(\frac{1}{2}\right)^{n+1}.$$

For example, if the event has been observed precisely once, then the odds are 3:1 that $p \geq 1/2$.

2.4.2. Alea Jacta Est

Price's next example is quite concrete: the throwing of a possibly biased die having an unknown number of faces.

> Suppose a solid or die of whose number of sides and constitution we know nothing; and that we are to judge of these from experiments made in throwing it.
>
> In this case, it should be observed, that it would be in the highest degree improbable that the solid should, in the first trial, turn any one side which could be assigned beforehand; because it would be known that some side it must turn, and that there was an infinity of other sides, or sides otherwise.

Note Price's strategy: Choose a relatively concrete example (here throwing a die) that seems familiar and draw deductions for that. Others would later refer to drawing from an urn containing a number of colored balls. (If one thinks of each color for a ball as corresponding to a side of Price's die, and the number of balls of a given color as corresponding to the relative likelihood that a given side will come up, the two are seen to be essentially the same example.) Then pass by analogy to more complicated inferences.

At this point Price turns to an issue of great importance often ignored in later discussions. What does it mean to talk about an event of which we are entirely ignorant? Surely the knowledge that the event is even possible implies some prior experience or encounter with it? Price's solution to this issue is to distinguish between the *first* time the event occurs, and all later occurrences. The first occurrence informs us that the event is *possible*, the later occurrences factor into our inductive inferences:

> The first throw only shews that it *has* the side then thrown, without giving any reason to think that it has it any one number of times rather than any other. It will appear, therefore, that *after* the first throw and not before, we should be in the circumstances required by the conditions of the present problem, and that the whole effect of this throw would be to bring us into these circumstances. That is: the turning of the side first thrown in any subsequent single trial would be an event about the probability or improbability of which we could form no judgment, and of which we should know no more than that it lay somewhere between nothing and certainty. With the second trial, then our calculations must begin.

(There is an echo here of Bayes's use of an initial ball W to fix a probability, but Bayes's purpose is very different from Price's.)

Price's proposal, almost universally ignored, gives at least one plausible criterion for when we have absolutely no knowledge of an event: we have seen the event once, so that we are cognizant of it, but otherwise have no experience of it. There is, of course, a natural objection: Even if an event is sui generis, arguing by analogy to other, distinct but similar events, we may form some idea of the likelihood or implausibility of seeing the event again. Nevertheless, Price's formulation is an advance over the reference to such events with no explanation at all, and moreover (as will be seen at the end of this chapter) has a surprisingly modern justification.

Thus it is with the second occurrence that Price begins, and he turns to another example.

2.4.3. The Rising of the Sun

Price now takes his next key step, and passes from the artificial example of a multisided die, to the observation of events in nature. He illustrates the use of Bayes's result by a curious hypothetical:

> Let us imagine to ourselves the case of a person just brought forth into this world, and left to collect from his observation of the order and course of events what powers and causes take place in it. The Sun would, probably, be the first object that would engage his attention; but after losing it the first night he would be entirely ignorant whether he should ever see it again. He would therefore be in the condition of a person making a first experiment about an event entirely unknown to him. But let him see a second appearance, or one *return* of the Sun, and an expectation would be raised in him of a second return. . . . But no finite number of returns would be sufficient to produce absolute or physical certainty.

The image of such a "philosophical Adam" first experiencing the sights and sounds of nature is not original to Price, but was a commonplace of the

Enlightenment; it can be found in one form or another in the writings of Buffon, Diderot, Condillac, and others; see Zabell (1998, n18). But Price clearly has Hume in mind: *His discussion is a direct attack on Hume.* (Gillies [1987] has a particularly good discussion of this point.)

In his *Enquiry Concerning Human Understanding* of 1748, the philosopher David Hume had argued

> Suppose a person, though endowed with the strongest faculties of reason and reflection, to be brought on a sudden into this world; he would, indeed, immediately observe a continual succession of objects, and one event following another; but he would not be able to discover anything further. He would not, at first, by any reasoning, be able to reach the idea of cause and effect.

Note the counterpoint between Hume and Price: Hume denies one can "reach the idea of cause and effect" in this setting; Price illustrates how the calculus of probabilities enables one to "collect from his observation of the order and course of events what powers and causes take place."

And Price's parting shot, that "no finite number of returns would be sufficient to produce absolute or physical certainty," is directed at another assertion of Hume. In his *Treatise* of 1739, Hume had identified a certain species of nonprobabilistic inductive inference:

> In common discourse we readily affirm, that many arguments from causation exceed probability, and may be receiv'd as a superior kind of evidence. One wou'd appear ridiculous, who wou'd say, that 'tis only probable the sun will rise tomorrow, or that all men must dye; tho' tis plain we have no further assurance of these facts, than what experience affords us. (*Treatise*, 124)

This, as will be seen later, was a common point of contention. Price flatly rejects the claim:

> Instead of proving that events will *always* happen agreeably to [uniform experience], there will always be reason against this conclusion. In other words, where the course of nature has been the most constant, we can have only reason to reckon upon a recurrency of events proportioned to the degree of this constancy; but we can have no reason for thinking that there are no causes in nature which will *ever* interfere with the operations of the causes from which this constancy is derived, or no circumstances of the world in which it will fail.

Price then immediately adds (and it is an important caveat):

> And if this is true, supposing our only *data* derived from experience, we shall find additional reason for thinking thus if we apply

> other principles, or have recourse to such considerations as reason, independently of experience, can suggest.

The point is that the calculation does not purport to give the full force of our conviction, only that portion arising from our experience of the number of times the event has and has not occurred.

2.5. Literature

Many papers have been published on the various aspects of Bayes's argument. For further discussion of Bayes's scholium, see Murray (1930), Edwards (1978), Stigler (1982), and Good (1988). For a detailed discussion of Hume's inductive skepticism, see Stove (1973). There have been numerous attempts to understand Hume in probabilistic and Bayesian terms since Price; one interesting example is Salmon's use of Bayes's theorem (in its modern sense) to interpret Hume's arguments in his *Dialogues Concerning Natural Religion* (Salmon 1978). The nature of such interpretations has, of course, been subject to considerable debate. Gower (1991), for example, believes Hume's probability of causes to be essentially non-Bayesian in nature; see Mura (1998) for an interesting rebuttal. Indeed Mura believes that "Hume developed—albeit informally—an essentially sound system of probabilistic inductive logic that turns out to be a powerful forerunner of Carnap's systems" (p. 303).

3. Laplace

> The curve described by a molecule of air or of vapour is following a rule as certainly as the orbits of the planets: the only difference between the two is due to our ignorance. Probability is related, in part to this ignorance, in part to our knowledge. (Laplace, *Essai philosophique sur les probabilités*)

Although Bayes is the eponymous founder of Bayesian statistics, Pierre-Simon, the Marquis de Laplace transformed the subject from a single brilliant, if sui generis, analysis, to an imposing edifice. There are several reasons for this.

First, Laplace was a mathematician of the first order, and in his hands mathematical probability was transformed from a specialized set of tools in the case of a finite number of outcomes (albeit a very impressive sets of tools in the hands of de Moivre), to a much richer theory applicable in both the discrete and continuous case. Second, Laplace was interested in the applications of probability, and both his papers and *Théorie analytique des probabilités* give numerous examples of serious applications of the calculus of probabilities to practical questions ranging from the analysis of observational errors to that of human populations. In the hands of Laplace, the calculus of probabilities became an important tool of the working scientist.

Laplace's conception of probability is subjective; probability relates in part to our knowledge and in part to our ignorance. In the Laplacean conception, one's probabilities evolve over time as our knowledge of the world changes, and we pass from prior or initial probabilities to posterior or final probabilities as information is received. Laplace's very first paper on probability is in fact devoted, in part, to the same problem considered by Bayes, but his approach is quite different. First, he takes as an axiomatic principle what is today referred to as "Bayes's theorem" (but in fact does not appear in Bayes's essay at all).

3.1. Bayes's Theorem

The standard modern definition of conditional probability, and the one used by Laplace, is the same as Bayes's: If an event B has positive probability, then the probability of an event A, given that B is known to have occurred, is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0.$$

There are two immediate consequences of the definition of conditional probability that are quite useful. Recall that a finite sequence of subsets $B_1, B_2, \ldots, B_n$ is said to be a *partition* of the sample space if the subsets are mutually exclusive and exhaustive. That is, they are pairwise disjoint (if $i \neq j$, then $B_i \cap B_j = \emptyset$) and $\Omega = \bigcup_{i=1}^n B_i$ (so that every $\omega \in \Omega$ lies in one and only one of the sets).

The Theorem of Total Probability If $B_1, B_2, \ldots, B_n$ is a partition of Ω, and $P(B_i) > 0$ for $1 \leq i \leq n$, then

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \mid B_i)\, P(B_i).$$

This is an immediate consequence of the additivity of probability, and the definition of conditional probability.

Bayes's Theorem If $B_1, B_2, \ldots, B_n$ is a partition, and $P(B_i) > 0$ for $1 \leq i \leq n$, then

$$P(B_j \mid A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A \mid B_j)\, P(B_j)}{\sum\limits_{i=1}^{n} P(A \mid B_i)\, P(B_i)}.$$

This is an immediate consequence of the theorem of total probability and the definition of conditional probability. It states that one can pass from probabilities of the form $P(A \mid B_i)$ to probabilities of the form $P(B_i \mid A)$ provided one knows the initial or prior probabilities $P(B_i)$. There is an alternative version of Bayes's theorem that is of particular interest. If $0 < P(A) < 1$, then the *odds in favor of* A are

$$O(A : A^c) = \frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}.$$

Given this definition we can now state the following.

Odds Version of Bayes's Theorem

$$\frac{P(B_i \mid A)}{P(B_j \mid A)} = \frac{P(A \mid B_i)}{P(A \mid B_j)} \cdot \frac{P(B_i)}{P(B_j)};$$

that is, the posterior odds equals the *likelihood ratio* times the prior odds. Put another way, the likelihood ratio is precisely the factor that transforms the initial odds into the final odds.

In his 1774 paper (Laplace 1774), Laplace states Bayes's theorem in both forms in the case that the hypotheses B_i are equally likely.

3.2. The Rule of Succession

Although he went far beyond the simple binomial situation, Laplace did consider it, proposing what is sometimes termed the *rule of succession*. Bayes had considered, given an unknown value of a probability p, what our degree of belief regarding it should be given k successes and $n-k$ failures have been observed in n trials. Laplace considered instead the *predictive probability* that the event itself would recur on the next trial. This is

$$P(X_{n+1} = 1 \mid S_n = k) = \int_0^1 p \cdot (n+1)\binom{n}{k} p^k (1-p)^{n-k}\, dp = \frac{k+1}{n+2}.$$

Thus, if in n trials one observes k successes, then the probability of a success on the next trial is $(k+1)/(n+2)$.

From the outset, Laplace's rule of succession has been widely criticized. A lightning rod for much of this is his use of the rule in the example of the rising of the sun:

> Thus we find that an event having occurred successively any number of times, the probability that it will happen again the next time is equal to this number increased by unity divided by the same number, increased by two units. Placing the most ancient epoch of history at five thousand years ago, or at 1826213 days, and the sun having risen constantly in the interval at each revolution of twenty-four hours, it is a bet of 1826214 to one that it will rise again tomorrow.

Few statements have been more often quoted or more often misinterpreted! First, as we have seen, this example does not originate with Laplace, but already appeared in Price's appendix to Bayes's essay, being in turn a response to Hume and an Enlightenment fascination with a human tabula rasa. Nor did Laplace seriously think this calculation gave the force of persuasion. As quoted as the foregoing passage is, few go on to Laplace's caveat:

> But this number is incomparably greater for him who, recognizing in the totality of phenomena the regulatory principle of days and

> seasons ("connaissant par l'ensemble des phénomenès le principe régulateur des jours et des saisons") sees that nothing at the present moment can arrest it.

(Note the similarity to Price's comment quoted earlier.)

3.2.1. The Finite Rule of Succession

One of the weaknesses of Laplace's rule in the form originally stated is its appeal to the infinite. Laplace initially talks in terms of drawing from an urn having a finite number of balls, but in his discussion of Bayes's problem, he refers to an urn having an infinite number of balls (or tickets) that are black and white. Strictly speaking this makes no sense, and Laplace must regard this a shorthand for something else (for example, the limiting case of an urn with a very large number of balls).

There is a natural move to make. Consider an urn having n balls, some black and some white. If a sample of size m is drawn from the urn and all m balls are black, what is the probability that the next ball is also black? To apply the Bayesian approach, one must make some assumption regarding the probabilities of the different possible constitutions. Let X denote the unknown number of black balls in the urn. The natural assumption to make, parallel to Bayes's postulate, is that $P(X = j) = 1/(n + 1)$, because then all possible ratios j/n are equally likely. One can then compute, using Bayes's theorem, that

$$P(X_{m+1} = 1 \mid S_m) = \frac{1}{n-m} \frac{\sum_{j=m+1}^{n} j(j-1)(j-2)\cdots(j-m)}{\sum_{j=m}^{n} j(j-1)(j-2)\cdots(j-m+1)}.$$

If the sum in the denominator is denoted $S_{m,n}$, then one can prove in a number of ways (see Zabell 1989b) that

$$S_{m,n} = \frac{(n+1)!}{(m+1)(n-m)!}$$

and therefore

$$P(X_{m+1} = 1 \mid S_m) = \frac{1}{n-m} \frac{S_{m+1,n}}{S_{m,n}} = \frac{m+1}{m+2};$$

exactly the same answer as in the infinite version of the law of succession!

This concordance, properly understood, turns out to be less surprising than it might seem at first sight. The reason will be revealed in a later section, but for now let us briefly summarize how this version of the rule of succession substantially improves on the infinite version.

- It eliminates all reference to the infinite in the analysis.
- It eliminates references to an objective chance p.
- It introduces another form of enumerative induction (based on sampling from a finite population).

3.3. Hume on Miracles

In his *Essai philosophique* of 1814, Laplace discussed a wide range of applications of the calculus of probabilities to science and society. One of these was of particular interest, illustrating the utility of the theory in cutting through the confusions attending Hume's celebrated attack against the belief in miracles. In his *Essay on Miracles*, Hume had argued that the credibility of testimony depended both on the reliability of the witness *and* the inherent plausibility or implausibility of the fact attested to. This is easily understood using Bayes's theorem. Let E represent the event in question, T the testimony under consideration. Let p be the prior probability regarding the event ($P(E)$), $a = P(T \mid E)$ the the conditional probability that if the event occurred, the witness would so testify, and $b = P(T \mid E^c)$ the corresponding conditional probability given E did not occur. Then by Bayes's theorem:

$$P(E \mid T) = \frac{pa}{pa + (1-p)b}.$$

As usual, this is more transparent in terms of odds: The posterior odds in favor of the event are:

$$\frac{P(T \mid E)}{P(T \mid E^c)} \left(\frac{p}{1-p} \right).$$

The first factor, the likelihood ratio, is a function of the reliability of the witness (but depends on two facets of that reliability, the competing likelihoods of how likely he is to tell the truth under the two scenarios that the event did or did not occur); the second, the prior odds, reflects the initial plausibility or implausibility of the event in question.

Although Bayes's theorem makes clear the appositeness of Hume's observation, the latter was questioned almost from the start. One of Hume's contemporaries, George Campbell, for example, noted we often hear reports about events that are inherently unlikely, yet we never question them on this ground. (One of Campbell's examples: If we know that a ferry has crossed the river successfully 1000 times, we usually do not question the veracity of someone we previously thought trustworthy, but who now tells us that the ferry has just sunk.) In his *Essai philosophique*, Laplace deftly illustrates the confusion underlying this objection.

Consider two simple cases. In the first, a ball is drawn from an urn containing 1000 balls numbered from 1 to 1000; in the second, a ball is drawn from an urn containing 999 black balls and 1 white ball. Suppose the witness correctly reports the result 9 out of 10 times. Now consider two questions: in the first case, if the witness announces that the ball drawn is numbered 79, what is the probability that it actually is 79; in the second, if the witness announces that the ball is white, what is the probability that it actually is white?

Laplace's point is that in both cases, there is a 1 in 1000 chance of the event, but the two posterior probabilities are very different—there is an important difference that the purely verbal discussion has glossed over. In the first case, if the ball drawn is *not* 79, but some other number, then for the witness to announce the ball is a 79, two things must happen: (a) The witness must not announce the correct number (which happens 1 in 10 times) *and* (b) choose 79 from among the remaining 999 possibles. Assuming this choice is equally likely to be any of the 999 candidates, the posterior odds the ball is 79 become:

$$\frac{9/10}{(1/10)(1/999)}\,\frac{1}{999} = \frac{9}{1}.$$

In contrast, in the second case, if the ball is black, then no choice is involved and the color white necessarily announced. Thus the posterior odds in this case are

$$\frac{9/10}{1/10}\,\frac{1}{999} = \frac{9}{999} = \frac{1}{111}.$$

That is, it is still odds *against* the ball being white, but the odds have decreased from 999 to 1 against, to only 111 to 1 against. The key distinction is that the first case is a simplified version of Campbell's ferry crossing, but the second corresponds to the case of a miracle. (It is much easier to think of reasons why one might lie about seeing a miracle than about a more mundane event.)

The example illustrates the utility of the subjective theory in giving a framework to reason about belief, even if one only thinks of subjective probabilities as being approximate or qualitative. (Anyone who doubts this last statement is invited to trudge through the vast literature written after Hume—on either side!—that did not have the benefit of the Bayesian framework.)

3.4. Literature

The literature on Laplace—interpretations of his work, the background of his day, his continuing influence up to the present—is vast. One classic paper on the interaction between the Laplacean program and its uses in the political and social realms is Gillispie (1972); Baker (1975) explores this theme in great depth. For the transmission of Hume's inductive skepticism to our time via Laplace and his school, see Stove (1973, Chapter 8), Zabell (1989b).

There is an extensive literature discussing the Bayesian interpretation of Hume's argument against the belief in miracles; see for example Owen (1987), Sobel (1987, 1991), Dawid and Gillies (1989), Gower (1990), Langtry (1990), Earman (1993), Holder (1998), Levine (1998). John Earman's *Hume's Abject Failure: The Argument against Miracles* (Earman 2000) is of particular interest.

4. The Frequency Theory

The Laplacean edifice was a formidable one, but after Laplace passed from the scene some spirited critics emerged. One central point of attack was the assertion that probability is a measure of frequency of occurrence, not a degree of belief. This position was championed by Robert Leslie Ellis, Antoine

Augustin Cournot, and Jacob Friedrich Fries in the 1840s, and later, by John Venn, Charles Sanders Peirce, and others.

There is a puzzling dichotomy here, one that is advanced often without apparent recognition of the important implicit underlying assumption. It is to be found in much of the literature from this period. And that is the facile assumption that *probability must be one or the other*. Once recognized, this type of attack is easily defanged. It is instructive, for example, to see how easily Frank Ramsey (in 1926) brushes aside nearly a century of frequentism: He concedes it to be a common linguistic usage, that it provides a natural application of the mathematical theory, he is even willing (even though he makes it clear he does not actually believe this) to concede that it may be the most important application of the theory:

> Suppose we start with the mathematical calculus and ask, not as before what interpretation of it is most convenient to the pure mathematician, but what interpretation gives results of the greatest value to science in general, then it may be that the answer is again an interpretation in terms of frequency; that probability as it is used in statistical theories, especially in statistical mechanics—the kind of probability whose logarithm is the entropy—is really the ratio of two numbers, of two classes, or the limit of a ratio. I do not myself believe this, but I am willing for the present to concede to the frequency theory that probability as used in modern science is really the same as frequency.

This passage is quoted at length because it makes an important point. It is one thing to argue or believe or establish that the frequency theory of probability is a valid or important one, it is quite another to conclude that a logic of partial belief is thereby somehow ruled out. Richard Leslie Ellis, for example, argued that whenever a person judges one event to be more likely to happen than another, introspection ("an appeal to consciousness") will reveal the concomitant "belief that on the long run it will occur more frequently" (Ellis 1844); see Zabell (1991, 212–215). This ignores the fact that in many instances we have intuitions about the likelihood of events that can only occur once.

Nevertheless, it must at the same time be conceded that the proponents of the subjective theory also had their own house in serious disorder as well. If probability is a numerical measure of partial belief, exactly what does such a number mean? And what is the justification for the axioms of probability?

Consider for example the analysis of Augustus De Morgan, one of the subjective theory's more thoughtful nineteenth-century proponents. After discussing the use of probability as a degree of belief, and the choice of a scale from 0 to 1 for convenience, De Morgan is faced with the justification of the assumption of additivity. He begins by stating it as a postulate:

> When any number of events are disjunctively possible, so that one of them may happen, but not more than one, the measure of our

> belief that one out of any some of them will happen, ought to be the amount of the measures of our separate beliefs in each one of those some. (De Morgan 1847, 179)

After three pages of discussion, De Morgan candidly admits that he cannot derive this postulate from any more basic assumption:

> And I cannot conceive any answer except that it is by an assumption of the postulate. That such an assumption will finally be knowingly made, on the fullest conviction, by every one who studies the theory, I have no doubt whatever: nor that it has been made, no matter in what words, nor with what clearness of avowal, by everyone who has studied that theory. (De Morgan 1847, 182)

Thus such questions had to wait until the twentieth century for a satisfactory answer.

5. The Rule of Succession

> A word of apology may be offered here for the introduction of a new name. The only other alternative would have been to entitle the rule one of *Induction*. But such a title I cannot admit, for reasons which will be almost immediately explained. (Venn 1866, 190, 3rd ed.)

Absent a cogent explication of the meaning of numerical epistemic probability, the opponents of the subjective theory had two obvious strategies. One, which we have seen, was to deny that belief could be numerically measured. The other was to examine the applications of the theory and ridicule them. One of the prime objects of this was the rule of succession. There were two prime targets here, each reflecting a weakness in the subjective position. One was the principle of indifference, the other the analogy to the urn of nature.

5.1. The Principle of Indifference

As we have seen earlier, Bayes was certainly aware of the vulnerability of the assumption that, absent any knowledge concerning it, our degree of belief regarding the probability of an event should be uniformly distributed; Bayes went to considerable lengths to avoid directly invoking it. During the nineteenth-century debate Bayes's own rather cautious position was ignored. Given the dominance of the Laplacean position, it was natural to start from Laplace rather than Bayes, and Laplace did not employ Bayes's reasoning. Laplace, moreover, freely used uniform priors in many other situations, ones in which Bayes's reasoning could not be invoked.

The example of the rising of the sun served as a touchstone for attack, although, as we have seen, Laplace's qualified use of the example was based on its historical interest, as well as his desire to correct an error in the

French literature. In England, De Morgan's adoption of the rule of succession provoked a sustained attack by his compatriot, the logician John Venn. But Venn's criticism, verging on ridicule, employed examples that clearly violated the obvious assumptions underlying the rule of succession, and even R. A. Fisher, who was largely sympathetic to Venn's position, felt compelled to take exception to Venn's criticisms. (See Zabell 1989a.)

The problem with the defense by proponents of the rule that its application requires no prior knowledge of the event in question is that it is difficult to in fact advance appropriate and realistic examples. Jevons, for example, whose enthusiasm outran his judgment, suggested the proposition that "a platythliptic coefficient is positive." Keynes (1921, 42–43) points out that the force of the example depends on our total ignorance regarding the meaning of the word "platythliptic," but that the example is flawed because we do possess information regarding the words "coefficient" and "positive." Furthermore, if someone knows no Arabic, is it equally likely that every statement in Arabic has an equal probability of being either true or false? Such examples avoid the previous difficulty but appear either absurd or useless.

Logical theories of probability hope to circumvent such difficulties by taking the logical syntax of language as a starting point, the atoms of language representing equipossible alternatives. If Wittgenstein's program of logical atomism had succeeded, then logical probability might be possible, but the failure of the former underlies the futility of the later. Logical probability reached its high point under Carnap, but Carnap's program necessarily retains an ultimate element of subjectivism, both in its choice of language and its assumption that the alternatives under consideration are equiprobable.

Ultimately it came to be realized, even in the benighted nineteenth century, that the determination that a certain set of alternatives is equally likely in fact represents a *positive* state of knowledge. Von Kries referred to a judgment of equiprobability under such circumstances as the *principle of cogent reason*. Take, for example, a six-sided die and let X represent the number that comes up when one tosses the die. If one has no reason to believe that one side is more likely to come up than another, then $P(X = i)$ must have the same value for $1 \leq i \leq 6$. Because these probabilities must add to one, $P(X = i) = 1/6$ for every i. This is not an argument concerning an event "about which we absolutely know nothing"; it represents instead one in which a considerable amount of information is available, but information that distributes itself equally among the alternatives. (Laplace would not have been impressed. He considered the multinomial as well as the binomial case, and famously stated that "probability is relative, in part to our knowledge and in part to our ignorance.")

The weakness of this approach is that it is only applicable when we are in a situation where the alternatives are judged to be equally likely, that is, what is sometimes termed a "fundamental probability set." This represents a considerable restriction on the applicability of the theory, and initial attempts to finesse this point did not inspire confidence.

5.2. The Urn of Nature

Indeed, such attempts to skirt the issue go back to the earliest attempts at realistic application of the mathematical theory. James Bernoulli, in his *Ars conjectandi* (published posthumously in 1713), implicitly assumed one can always reduce to the case of equally likely outcomes by breaking down unequal possibilities into other, more basic equipossible cases. But even the simplest realistic examples raise intractable difficulties. Take the case of the sex ratio (considered by Arbuthnot 1710): If a child is born, what are the odds the child is male? The natural partition is $\Omega = \{H, T\}$, but it is known empirically that a male child is slightly more likely than a female (typically on the order of 53 percent).

Price, in his appendix to Bayes's essay, passed from the case of the sides of a die to events of nature with no attempt at justification, and others took a similar tack. Jevons (1874, 150, 1877 edition), for example, states:

> Nature is to us like an infinite ballot box, the contents of which are being continually drawn, ball after ball, and exhibited to us. Science is but the careful observation of the succession in which balls of various character present themselves. . . .

Note that a subtle shift from Price's argument has taken place here. Price likened the event to a side on a die having an unknown number of sides. In effect each of the possible outcomes has (or is judged to have) an equal propensity of occurrence, and this is the justification for giving each equal probability. But for Jevons (and others), one is sampling from a population each of whose members is equally likely to be chosen; this is the *urn of nature* (see, e.g., Strong 1976). But how does one analyze such a situation? One is back to the flat prior of Bayes and Laplace! (Some assumption about the contents of the urn has to be made.)

Venn's criticisms of the classical theory were ultimately toned down in the third edition of his *Logic of Chance*, due in part to the influence of Edgeworth. (Venn's *Logic of Chance* went through three editions. The second and third editions both saw major shifts in Venn's position; it is unfortunate that no systematic study of these exists at present.) Edgeworth, who also wrote a review of the third edition of Venn's book, may be regarded as the informed Laplacean response to Venn's attack.

In the end, Edgeworth's approach is pragmatic; the flat prior is appropriate because it is found to hold, at least approximately, in nature. This assertion seems both too facile and totally unsupported, but it reflects nevertheless an important advance: The prior represents our knowledge, not our ignorance. It is only one step from Edgeworth's defense to an acknowledgment that in general our knowledge may not be appropriately described by a flat prior. But then what do we do?

5.3. The Family of Beta Priors

One alternative that was subsequently advanced is the use of the *beta family* of probability distributions. The beta family has the attractive and mathematically convenient feature that if one observes k successes in n trials, and the probability of success p has a prior distribution that is beta with parameters α, β, denoted $f_{\alpha,\beta}(p)$, then the posterior distribution of p is

$$f_{\alpha+k,\beta+n-k}(p) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+k)\,\Gamma(\beta+n-k)} p^{\alpha+k-1}(1-p)^{\beta+n-k-1}, \quad 0 \le p \le 1.$$

In other words, the posterior distribution lies in the same family as the prior; the beta distribution is the *conjugate prior* for the binomial.

Note that if $\alpha = \beta = 1$, then this reduces (as it must) to Bayes's posterior distribution:

$$f_{k+1,n-k+1}(p) = \frac{(n+1)!}{k!\,(n-k)!} p^k (1-p)^{n-k}, \quad 0 \le p \le 1.$$

Just as in the case of the uniform prior, it is a simple matter to derive the predictive probabilities for the general beta prior. If the prior distribution of p is $f_{\alpha,\beta}(p)$, then

$$P(S_n = k) = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\, dp.$$

Thus in particular

$$P(S_1 = 1) = P(X_1 = 1) = \frac{\alpha}{\alpha+\beta};$$

the prior probability of a success is the ratio $\alpha/(\alpha+\beta)$.

Given these "cylinder set" probabilities, it is simple matter to compute the rules of succession (or *predictive probabilities*):

$$\begin{aligned}
P(X_{n+1} = 1 \mid S_n = k) &= \frac{P(X_{n+1} = 1, S_n = k)}{P(S_n = k)} \\
&= \frac{\int_0^1 p^{\alpha+k}(1-p)^{\beta+n-k-1}\, dp}{\int_0^1 p^{\alpha+k-1}(1-p)^{\beta+n-k-1}\, dp} \\
&= \frac{\Gamma(\alpha+k+1)\,\Gamma(\beta+n-k-1)}{\Gamma(\alpha+\beta+n+1)} \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+k)\,\Gamma(\beta+n-k-1)} \\
&= \frac{\alpha+k}{\alpha+\beta+n}.
\end{aligned}$$

Note that the predictive probability can be instructively expressed as a *weighted average*

$$P(X_{n+1} \mid S_n = k) = \left(\frac{n}{n+\alpha+\beta}\right)\frac{k}{n} + \left(\frac{\alpha+\beta}{n+\alpha+\beta}\right)\frac{\alpha}{\alpha+\beta}$$

of the *sample frequency* k/n and the *prior probability* $\alpha/(\alpha+\beta)$.

The impact on the posterior odds of choosing different values of α and β is now apparent: The *ratio* α/β determines the *prior odds*; the *sum* $\alpha+\beta$ determines the *relative weight* (in conjunction with the sample size n) assigned to the prior odds versus the sample frequency.

The Expectation and Variance of the Beta How does one assign values to α and β? There is a useful trick, but first one must compute the moments of the prior. By definition, if X is a random quantity, then $E[X^n]$ is the *nth moment* of X. For the beta distribution one has

$$E[p^n] = \int_0^1 p^n \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\,dp = \frac{\Gamma(n+\alpha)\,\Gamma(\beta)}{\Gamma(n+\alpha+\beta)}\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}.$$

In particular,

$$\begin{aligned} E[X] &= \frac{\alpha}{\alpha+\beta}, \\ Var[X] &= E[X^2] - E^2[X] \\ &= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 \\ &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}. \end{aligned}$$

This gives two equations in two unknowns; if one summarizes one's prior knowledge by estimating the expectation and variance of p (in terms of the center and spread of one's initial distribution of belief), then one has two equations and two unknowns, and one can solve for α and β.

The Device of Imaginary Results James Bernoulli titled his book the *Ars conjectandi*, the "art of conjecture"; there is certainly an element of art in the actual practical use of subjectivist methods. One tool among many, for purposes of illustration, is I. J. Good's *device of imaginary results*.

As will be stressed later, the subjective theory of probability is a theory of *consistency*; it tells us that various quantities must satisfy certain constraints. Although it is usual to think in terms of Bayes's theorem as passing from prior to posterior as new information is received, one could just as well think of it as linking prior and posterior, without prejudging which is given and which derived. In Good's approach, one elicits information about the prior by varying the values of α and β, seeing which give posterior probabilities consistent with our beliefs.

For example, suppose a friend claimed to have ESP (extrasensory perception), and attempted to demonstrate it by having you toss a coin and he guesses the outcome. How many successes in a row would your friend have to achieve before you thought it odds of 9:1 that he would correctly guess the outcome of the next toss of the coin? If your prior distribution was symmetric (so that $\alpha = \beta$), then the answer to this single question would uniquely determine the value of α.

An Urn Model Interpretation for the Beta To be useful, one needs some guides for using the beta distribution as a family of priors. Here is one that is both attractive and simple: the *Polya urn model.* Suppose one has an urn containing α red and β black balls, α and β being some pair of positive integers. Each time a ball is chosen, it is replaced *together with a ball of the same color.* Thus if in n trials, one chooses k reds and $n-k$ blacks, the urn has a total of $\alpha+\beta+n$ balls, $\alpha+k$ red and $\beta+n-k$ black. Thus the probability of choosing a red ball on the next trial is

$$P(X_{n+1} = 1 \mid S_n = k) = \frac{\alpha + k}{\alpha + \beta + n},$$

that is, exactly the same probability as the predictive probability arising from a beta prior with parameters α, β.

In general the probability of a sequence of events can be built up from a succession of conditional probabilities. For example,

$$P(X_1 = e_1,\, X_2 = e_2,\, X_3 = e_3)$$

is the same as

$$P(X_1 = e_1) \cdot P(X_2 = e_2 \mid X_1 = e_1) \cdot P(X_3 = e_3 \mid X_1 = e_1, X_2 = e_2).$$

(The probabilities $P(X_1 = e_1)$ can be regarded either as trivial instances of conditional probabilities based on null information.) This is one reason rules of succession are of such importance: They can be used to compute the probability of *any* possible sequence of outcomes. It is an immediate corollary that whenever two (apparently different) descriptions of two processes give rise to the same rules of succession, the two processes are in fact identical.

The application to the case at hand is an exceedingly interesting one. For concreteness, consider the case $\alpha = \beta = 1$ (the uniform prior, giving rise to the Bayes-Laplace process). At first blush the Polya urn process and the Bayes-Laplace process are two very different entities. In one case one is choosing n balls (say) at random from an urn according to a rule; in the other, one chooses a probability p at random from the unit interval and then tosses a p-coin n times in succession. But because the Polya process and the Bayes-Laplace process have the same rules of succession, the two are in fact stochastically indistinguishable. Thus the use of a beta prior having integer parameters α and β is tantamount to acting as if the entire impact of one's prior knowledge is that one judges it to be as likely to observe a success as to choose a red from an urn containing α red and β black balls.

5.4. The Confirmation of Universal Generalizations

It can be shown that the beta priors have the following apparently paradoxical property: No matter how long an unbroken string of n_1 successes one has observed in the past, the posterior probability of a subsequent string of n_2 unbroken successes in the future goes to zero as n_2 becomes arbitrarily large. It is particularly easy to see this in the case $\alpha = \beta = 1$, because in this case

$$\begin{aligned} P(X_{n_1+1} = X_{n_1+2} = \cdots = X_{n_1+n_2} \mid S_{n_1} = n_1) \\ = \frac{n_1+1}{n_1+2} \cdot \frac{n_1+2}{n_1+3} \cdots \frac{n_1+n_2}{n_1+n_2+1} \\ = \frac{n_1+1}{n_1+n_2+1}, \end{aligned}$$

and clearly this tends to zero as $n_2 \to \infty$ (n_1 remaining fixed). The resolution of this apparent paradox had to wait until the modern subjective theory.

6. W. E. Johnson

The ingenious Bayes had avoided many of the later objections to the principle of indifference by focusing on the predictive probabilities of discrete and observable events (the number of successes in a fixed number of future trials), rather than about an unknown (and unobservable) continuous parameter p. In the 1920s, the English philosopher and logician William Ernest Johnson formulated a generalization of this approach applicable to the multinomial setting.

Johnson accomplished this in two stages. First, in an appendix to volume 3 of his treatise on logic (Johnson 1924), Johnson presented a derivation of the uniform multinomial prior by introducing two assumptions, the *permutation* and *combination* postulates. In a later paper, published posthumously, Johnson (1932) replaced the combination postulate by another, the *sufficientness* postulate (to use the later terminology of I. J. Good), thereby deriving the family of symmetric Dirichlet priors. Each of these three assumptions involved major advances in understanding.

6.1. The Permutation Postulate

First, some terminology. Suppose there are t types or categories representing the possible outcomes in a sequence of observations or trials, say $c_1, c_2, \ldots, c_t$ ($t \geq 2$). These might be, for example, the letters in an alphabet or the species in a region. Let $X_1, \ldots, X_n$ denote a sequence of n observations, and $e_1, \ldots, e_n$ the resulting sequence of outcomes. The notation $X_i = e_j$ means that on the ith trial one observes e_j. (Thus X_i denotes the random quantity prior to its being determined, e_j the specific outcome observed, and $X_i = e_j$ means that on the ith trial the outcome e_j is observed.)

Associated with any sequence of observations, say $X_1 = e_1, \ldots, X_n = e_n$, is a vector of *frequency counts* $n_1, \ldots, n_t$, so that n_j is the number of times outcomes of type j occur in the sequence (that is, n_j records the number of times in the sequence that the jth type has occurred). Formally,

$$n_j = |\{\, i : e_i = c_j, 1 \leq i \leq n \,\}|.$$

Johnson's permutation postulate states that any two sequences having the same frequency counts are equally likely to occur; that is, in modern terminology, that they are *exchangeable.* This is one of the earliest instances of the use of this assumption, and it has great philosophical importance. Let us try to understand it.

First, there is a simple alternative characterization of (finite) exchangeability: The probability distribution of a sequence of random variables is finitely exchangeable if it is invariant under permutation of the time index; that is, if for every possible sequence of outcomes $e_1, \ldots, e_n$, and every permutation σ of $\{1, \ldots, n\}$, one has

$$P(X_1 = e_1, \ldots, X_n = e_n) = P(X_1 = e_{\sigma(1)}, \ldots, X_n = e_{\sigma(n)}).$$

(To see this equivalence, just note that the original and permuted sequences obviously have the same frequency counts $n_1, \ldots, n_t$; and—only slightly less obviously—two sequences having the same frequency counts are merely permuted versions of each other.)

Thus, if a sequence is exchangeable, then any two sequences having the same frequency counts have the same probability. Suppose $X_1, \ldots, X_n$ is a random sequence having frequency counts $n_1, \ldots, n_t$. There are

$$\binom{n}{n_1 n_2 \cdots n_t} = \frac{n!}{\prod_{i=1}^{t} n_i!}$$

such possible sequences (given $n_1, \ldots, n_t$), and each of these has equal probability. Thus

$$P\big(X_1 = e_1, \ldots, X_n = e_n \mid (n_1, \ldots, n_t)\big) = \frac{\prod_{i=1}^{t} n_i!}{n!}.$$

This observation has two important consequences.

First, the fact that the conditional probability distribution is uniform means that the frequency counts are *sufficient statistics* for the sequence $X_1, \ldots, X_n$. (In statistical inference, sufficient statistics summarize the relevant information contained in a set of data for purposes of estimation.)

Second, there is a simple urn model for these conditional probabilities. Consider an urn, containing n_j balls of color c_j, for $j = 1, \ldots, t$, Now pick successive balls at random until the urn is empty. The resulting probability distribution coincides with the one above. Thus this probability distribution corresponds to the one arising from *sampling without replacement* from a finite population. This is sometimes termed a *hypergeometric distribution.*

The different possible frequency vectors $\mathbf{n} = (n_1, \ldots, n_t)$ partition the possible sequences into disjoint subsets. Thus, by the so-called theorem of total probability, one has

$$P(X_1 = e_1, \ldots, X_n = e_n) = \sum_{\mathbf{n} \in S_n} P(X_1 = e_1, \ldots, X_n = e_n \mid \mathbf{n}) \cdot P(\mathbf{n}),$$

where S_n denotes the possible different partitions of n into a sum of the form $n_1 + \cdots + n_t$. Thus the probability function P has been decomposed into a sum, each term being the product of a distribution arising from sampling without replacement from an urn, weighted by $P(\mathbf{n})$. The overall probability can be realized or simulated by picking an urn at random, according to the weights $P(\mathbf{n})$, and then sampling without replacing from the urn until its contents are exhausted. Such a process will generate a finitely exchangeable sequence, and every finitely exchangeable sequence arises in this way.

The importance of this observation is that *the assumption of exchangeability permits us to dispense with the urn of nature.* As we have seen, starting with Price's argument and continuing throughout the nineteenth century, a key element in the probabilistic analysis of induction was the analogy between a natural process and some concrete chance set-up such as a many-sided die (as in Price's appendix), or an urn, finite or otherwise, of unknown composition. The concept of exchangeability permits us to dispense with such questionable analogies. Instead one asks: Is my information such that I judge that any two sequences having the same frequency counts are equally likely to occur? If the answer is "yes," then the probability distribution summarizing my beliefs is necessarily exchangeable, and therefore *mathematically equivalent* to a weighted average of hypergeometric distributions.

6.2. Multinomial Sampling

Laplace also considered the above case of *multinomial sampling*, that is, the situation where there are a discrete number of outcomes, not necessarily limited to just two. Let us consider multinomial setting from Laplace's perspective.

As before, let $X_1, X_2, \ldots, X_n$ denote a random sample drawn from a population the members of which fall into one of t possible categories types ($t \geq 2$). Each type has a probability p_i of occurring, so that $p_i \geq 0$ and $p_1 + \cdots + p_t = 1$. The vector of probabilities lies in the $t-1$ dimensional probability simplex

$$\Delta_t := \left\{ (p_1, \ldots, p_t) : p_j \geq 0, \sum_{j=1}^{t} p_j = 1 \right\}.$$

Given the random sample $X_1, \ldots, X_n$, the associated frequency counts $n_1, \ldots, n_t$ have a *multinomial distribution*

$$\frac{n!}{n_1!\, n_2! \cdots n_t!} p_1^{n_1} p_2^{n_2} \cdots p_t^{n_t}.$$

Now suppose that the true probability vector is unknown. Let $d\mu$ denote a prior distribution on Δ_t, representing either an objective, aleatory mechanism giving rise to the vector $(p_1, \ldots, p_t)$, or a subjective, epistemic degree of belief regarding the possible values of the vector $(p_1, \ldots, p_t)$. Suppose $e_1, \ldots, e_n$ is a possible sequence with frequency counts $n_1, \ldots, n_t$. Then, just as in the binomial case, one can invoke the theorem of total probability to see that

$$P(n_1, n_2, \ldots, n_t) = \int_{\Delta_t} \frac{n!}{n_1!\, n_2! \cdots n_t!} p_1^{n_1} p_2^{n_2} \cdots p_t^{n_t} \, d\mu(p_1, \ldots, p_t).$$

In principle $d\mu$ can be any probability measure on the simplex, but it is mathematically attractive to take a member of the *Dirichlet family*. To proceed we need the following *basic fact*: If $\alpha_1, \ldots, \alpha_t > 0$ and $\alpha := \alpha_1 + \cdots + \alpha_t$, then

$$\frac{\prod_{j=1}^t \Gamma(\alpha_j)}{\Gamma(\alpha)} = \int_{\Delta_t} \prod_{j=1}^t p_j^{\alpha_j - 1} \, dp_1 dp_2 \cdots dp_{t-1}.$$

Every integration formula involving a nonnegative integrand gives rise to a probability density by normalization. Here one has

$$\int_{\Delta_t} \frac{\Gamma(\sum_{j=1}^t \alpha_j)}{\prod_{j=1}^t \Gamma(\alpha_j)} \prod_{j=1}^t p_j^{\alpha_j - 1} \, dp_1 dp_2 \cdots dp_{t-1} = 1.$$

The integrand is a probability distribution on Δ_t, a Dirichlet distribution with parameters $\alpha_1, \ldots, \alpha_t$.

Suppose as before one observes frequencies $n_1, \ldots, n_t$ in a sequence of n trials. Then it is not difficult to see (see the appendix) that

$$P(X_{n+1} = c_j \mid n_1, \ldots, n_t) = \frac{n_j + \alpha_j}{n + \alpha},$$

a rule of succession that is a natural generalization of the one discussed in the binomial setting. In this case,

$$P(X_1 = e_1, \ldots, X_n = e_n) = \int_{\Delta_t} \frac{\Gamma(\sum_{j=1}^t \alpha_j)}{\prod_{j=1}^t \Gamma(\alpha_j)} \prod_{j=1}^t p_j^{n_j + \alpha_j - 1} \, dp_1 dp_2 \cdots dp_{t-1}.$$

In the special case that $\alpha_1 = \alpha_2 = \cdots = \alpha_t = 1$, the rule of succession reduces to

$$P(X_{n+1} = c_j \mid n_1, \ldots, n_t) = \frac{n_j + 1}{n + t}.$$

6.3. Johnson's Combination Postulate

The Laplacean analysis employs in an essential way probabilities about probabilities, and raises questions about the differences between the two. Johnson was able to avoid this, noting

> I substitute for the mathematician's use of Gamma Functions and the α-multiple integrals, a comparatively simple piece of algebra, and thus deduce a formula similar to the mathematician's, except that, instead of for two, my theorem holds for α alternatives, primarily postulated as equiprobable. (Johnson 1932, 418) [Johnson's α corresponds to our t.]

Johnson's initial approach was to assume that all possible frequency counts are equipossible.

Johnson's Combination Postulate Every ordered t-partition of n is equally likely. (Note that if $t = 2$, then Johnson's combination postulate reduces to Bayes's; for in this case the ordered partitions are of the form $(k, n - k)$, and the equiprobability assumption reduces to the assumption that all values of k are equally likely.)

There are some rather attractive combinatorics at play here. There are a total of

$$A_{n,t} =: \binom{n+t-1}{t} = \frac{(n+t-1)!}{t!\,(n-1)!}$$

ordered t-partitions of n (Feller 1968, 38), and each such partition is assigned a probability of $1/A_{n,t}$ under the combination postulate. (In statistical mechanics the frequency counts are sometimes termed *occupancy numbers*, and this particular probability distribution on them *Bose-Einstein statistics*.) It is not difficult to show (as Johnson noted) that the combination postulate in conjunction with the permutation postulate uniquely characterizes the rules of succession: one obtains the Laplacean result

$$P(X_{n+1} = c_j \mid n_1, \ldots, n_t) = \frac{n_j + 1}{n + t}.$$

(Note that this again represents a generalization of Bayes's approach: If $t = 2$, then one obtains the classical rule of succession $(k + 1)/(n + 2)$.)

6.4. Johnson's Sufficientness Postulate

Although its mathematics is attractive, Johnson's combination postulate is by no means compelling (and for this reason was soon after criticized by C. D. Broad). Later on Johnson arrived at a more general postulate, published posthumously in 1932. His *sufficientness postulate* assumes that

$$P(X_{n+1} = c_i \mid n_1, \ldots, n_t) = f_i(n_i, n);$$

that is, conditional on the observed frequency counts $n_1, \ldots, n_t$ in a sample of size n, the predictive probability of observing an outcome in the ith category is a function of the number of elements n_i in the sample previously observed in that category, and the total sample size n, but is *independent of how the other counts distribute themselves.* (Strictly speaking, Johnson assumed that the function $f_i(n_i, n) = f(n_i, n)$, that is, that the rule of succession does not depend on the particular category under consideration, but his analysis in fact easily extends to the more general case where such dependence is permitted.) Johnson's postulate is certainly true for the predictive probabilities arising from the Dirichlet prior. However, Johnson discovered a surprising fact: This qualitative assumption actually determines the quantitative values of the predictive probabilities (up to a finite set of parameters)!

Theorem 1 Let $X_1, X_2 \ldots,$ be a t-multinomial sequence such that $X_1, \ldots, X_n$ is exchangeable for every $n \geq 1$. If

1. $P(X_1 = e_1, \ldots, X_n = e_n) > 0$ for every possible sequence $e_1, \ldots, e_n$;
2. $t > 2$;
3. $P(X_{n+1} = c_i \mid n_1, \ldots, n_t) = f_i(n_i, n)$;

then either (1) the outcomes of the sequence are independent of one another, or (2) there exist positive constants $\alpha_1, \ldots, \alpha_t$ such that (setting $\alpha = \alpha_1 + \cdots + \alpha_t$)

$$P(X_{n+1} = c_i \mid n_1, \ldots, n_t) = \frac{n_i + \alpha_i}{n + \alpha}$$

for every $n \geq 0$, and set of frequency counts $n_1, \ldots, n_t$.

In the special case $n = 0$ the theorem tells us that the initial, unconditional probabilities are

$$P(X_1 = c_i) = \frac{\alpha_i}{\alpha}.$$

Thus the α_i encode the relative initial likelihood of the outcomes, and α quantifies the relative weight given to past experience versus the information provided in the sample. Note that the initial probabilities and the one-step predictive probabilities determine the probabilities of every cylinder set. For example, the cylinder set probability

$$P(X_1 = e_1,\, X_2 = e_2,\, X_3 = e_3)$$

is equal to

$$P(X_1 = e_1)\, P(X_2 = e_2 \mid X_1 = e_1)\, P(X_3 = e_3 \mid X_1 = e_1,\, X_2 = e_2).$$

Because the initial probabilities and rules of succession are the same as those arising from a Dirichlet prior, it follows that the "cylinder set" probabilities of the sequence (that is, the probabilities of finite sequences, such as e_1, e_2, e_3)

are the same as those arising from a mixture of multinomial probabilities using a Dirichlet prior. That is, not only do the predictive probabilities arising from such a mixture satisfy Johnson's sufficientness postulate, but the postulate *characterizes* such mixtures (in the sense that all relevant probabilities are determined).

Of course the justification for the sufficientness postulate can itself be questioned, but nevertheless this remarkable result has several important consequences. First, it replaces an essentially mathematical device (the assumption of a Dirichlet prior) by a purely *qualitative* assumption on the predictive probabilities. It is an assumption that could summarize a state of knowledge, and is therefore in the spirit of an epistemic approach to probability (as opposed to a quantitative assumption made purely for mathematical convenience.) Second, by characterizing the Dirichlet mixtures of multinomial probabilities, Johnson's theorem tells us precisely when their use is appropriate: when the sufficientness postulate accurately describes our state of knowledge. Finally, when the sufficientness postulate is deemed to be an appropriate description of our state of knowledge, the determination of probabilities is reduced from an infinite dimensional problem to one of determining a finite number of parameters $\alpha_1, \ldots, \alpha_t$.

7. Ramsey and de Finetti

Thus, in 1924 (when the initial portion of Johnson's work appeared), despite much progress, serious gaps in the subjective theory of probability remained. All this was to change with the work of Frank Plumpton Ramsey and Bruno de Finetti.

7.1. Ramsey

Frank Plumpton Ramsey (1903–1930) was a remarkable figure by any standard. Despite his early death at the age of 27, the slim volume of his posthumously collected papers (Braithwaite 1931) contains groundbreaking work in a number of areas: mathematical logic, probability, economics, and pure mathematics. One example of this is Ramsey's 1926 essay *Truth and Probability*.

This remarkable paper remains worthwhile reading even today. Lucid and profound, Ramsey disposes in rapid succession of competing theories, gives in what seems almost an obvious manner the first complete discussion of the foundations of quantitative subjective probability, seriously considers the problem of the "dynamic assumption of Bayesianism," and discusses the foundations of inductive inference. Here we give only a brief discussion of the paper, focusing on its relevance to the problem of induction. For a more detailed discussion of the paper as a whole, see Zabell (1991).

7.1.1. The Meaning of Probability

Ramsey began with a brief description of some of his predecessors and their competing theories, in particular the frequency theory. He dismisses the latter very simply, noting that even if it provides a reasonable interpretation of the mathematical structure, and even if its plays a useful role in scientific application, this does not preclude the additional possibility of an epistemic interpretation. Turning instead to the possibility of such an epistemic interpretation, Ramsey argued for an operational definition: "The degree of a belief is just like a time interval; it has no precise meaning unless we specify more exactly how it is to be measured."

Ramsey's greatest advance is in fact to present an operational definition of epistemic probability, and to demonstrate that the standard properties of probability can be deduced from such a definition.

Ramsey discusses two ways in which this can be done. One is the time-honored method of betting odds. If we are unsure about an outcome, the amount we are willing to bet on it occurring is a numerical measure of the strength of our belief. In some form this can already be found in Bayes, is clearly stated by some of Ramsey's predecessors, and Ramsey makes no claim of originality (indeed, his wording suggests the exact opposite). But such an approach does have some drawbacks (such as accounting for the diminishing utility of money), and Ramsey almost immediately passes to another system, his celebrated simultaneous axiomatization of probability and utility.

In this approach, it is assumed that one can consistently order preferences for different outcomes. Given outcomes α and β, to pass from a qualitative statement that one is more likely than the other to a quantitative statement regarding their respective probabilities, it is necessary to interpolate between a continuum of possibilities. Ramsey achieves by introducing the *ethically neutral* proposition, the philosophical equivalent of tossing a coin. (These are propositions regarding whose truth values we are entirely indifferent.) But of course there are many propositions concerning whose truth values we are indifferent, and it is necessary to introduce a criterion for when such propositions correspond to the outcome of tossing a *fair* coin. Ramsey's solution is simple:

> The subject is said to have belief of degree $\frac{1}{2}$ in such a proposition p if he has no preference between the options (1) α if p is true, β if p is false, and (2) α if p is false, β if p is true, but has a preference between α and β simply.

The introduction of this device, together with appropriate axioms of consistency, enables us—up to affine transformation—to assign both numerical utilities to the options, as well as numerical probabilities. Ramsey then shows that the probabilities so derived must satisfy the usual axioms of the probability calculus, and that one option is preferred to another if and only if its expected utility is greater. The argument is only sketched, but its outline is clear, and many of the subsequent axiomatization schemes follow the same general approach. One generalization of particular interest is due to Richard Jeffrey; in Jeffrey's system the utility function is determined only up to a fractional linear transformation (see Jeffrey 1983).

Providing a detailed operational definition for epistemic probability was in itself already a major advance, but Ramsey went on in his next section to list several important additional advantages.

- "It gives us a clear justification for the axioms of the calculus."
- "The principle of indifference can now be altogether dispensed with."
- The existence of probable knowledge: "I think I perceive or remember something but am not sure."

All three of these were substantial contributions to the theory. The first two are by now familiar, but the third raises new issues that deserve separate discussion.

7.2. The Dynamic Assumption of Bayesianism

The consistency or coherence assumptions resulting in a numerical measure of belief satisfyings the axioms of probability are essentially *static* in nature. That is, they constrain our beliefs at a fixed instant in time. Key to Ramsey's program was the recognition that it is fruitless to attempt to derive our current degrees of belief by a process of starting out in a state of primeval ignorance, and then tracking how these evolve over time with the receipt of new information.

But at the same time, it is certainly true that our degrees of belief do change over time as new information is acquired. The traditional method of incorporating such changes into the theory is via the use of conditional probabilities. The operational definition of these in the subjective approach is via the appeal to *conditional bets*. Thus de Finetti describes the conditional probability $P(A \mid E)$ as "the probability that we would regard as fair for a bet on A to be made immediately, but to become operative only if E occurs" (de Finetti 1972, 193); similar language is used by Ramsey (1931, 180). It is then a *consequence* of the Dutch book argument that $P(A \mid E) = P(A \cap E)/P(E)$.

Suppose now that we actually learn that the event E has occurred (or that the corresponding proposition is true). It common to assume that the new probability, call it $P^*(A)$, should be the same as the conditional probability $P(A \mid E) = P(A \cap E)/P(E)$. Ian Hacking calls this the *dynamic assumption of Bayesianism*. Ramsey himself recognized that such an identification is subject to question:

> [The degree of belief in p given q] is not the same as the degree to which [a subject] would believe p, if he believed q for certain; for knowledge of q might for psychological reasons profoundly alter his whole system of beliefs. (Ramsey 1931, 180)

There is an extensive literature discussing the possible justifications for the identification; one classic reference is Freedman and Purves (1969).

7.2.1. Probable Knowledge

It is impressive that Ramsey recognized not only the limitations in the dynamic assumption but also the problem of probable knowledge. Here is an example of this phenomenon:

> Suppose we are about to hear one of two recordings of Shakespeare on the radio, to be read by either Olivier or Gielgud, but are unsure of which, and have a prior with mass $\frac{1}{2}$ on Olivier, $\frac{1}{2}$ on Gielgud. After the recording, one might judge it fairly likely, but by no means certain, to be by Olivier. The change in belief takes place by direct recognition of the voice; all the integration of sensory stimuli has already taken place at a subconscious level. To demand a list of objective features that we condition on to affect the change would be a logician's parody of a complex psychological process. (Diaconis and Zabell 1982, 823)

7.2.2. Jeffrey's Rule of Conditioning

In the 1960s, Richard Jeffrey put forward a framework for understanding certain types of belief changes in response to probable knowledge. Suppose that P represents initial beliefs, and P^* final beliefs. Jeffrey posits that there exists a partition $\{E_1, E_2, \ldots, E_n\}$ of the sample space that captures the totality of our belief change, in the sense that

$$\begin{aligned} P(E_i) &\to P^*(E_i), \quad 1 \leq i \leq n; \\ P(A \mid E_i) &= P^*(A \mid E_i), \quad \text{all } A,\, i. \end{aligned}$$

Thus, any change is possible on the elements E_i of the partition, but conditional probabilities relative to members of the partition do not change. (A simple example would involve a finitely exchangeable sequence $X_1, X_2, \ldots, X_n$, with $S_n = X_1 + \cdots + X_n$. Receipt of new information might cause the change $P(S_n = i) \to P^*(S_n = i)$, $1 \leq k \leq n$, but if one continued to adhere to an exchangeable assignment, then

$$\begin{aligned} P(X_1 = e_1, \ldots, X_n = e_n \mid S_n = i) &= \frac{1}{\binom{n}{i}} \\ &= P^*(X_1 = e_1, \ldots, X_n = e_n \mid S_n = i), \end{aligned}$$

provided $e_1 + \cdots + e_n = i$ (and zero otherwise).

The rejection of the principle of indifference (and the development of a theory that made it unnecessary) freed the theory from defending a principle that, although reasonable in some settings, was unreasonable to assert as a universal truth. But the abandonment of the principle of indifference also raised a difficulty for the theory: How does one justify the inductive process? The answer to this question had to await the contributions of another pioneer, Bruno de Finetti.

7.3. de Finetti

de Finetti's writings on probability span his lifetime and cover a wide variety of topics. His 1937 essay, *La prevision: ses lois logiques, ses source subjectifs* (de Finetti 1937), summarizes his prewar work; his two books on probability in English (de Finetti 1972) summarize much of his later work. His interest in subjective probability reflects his more general subjectivist view of science; see de Finetti (1931). We focus here on his 1937 essay.

Like Ramsey, de Finetti considered more than one way of defining epistemic probability and justifying its axioms. In *La prevision* his definition is that of betting odds, and his tactic for deriving the axioms is the appeal to the *Dutch book argument*: Whatever one's odds, they should satisfy the minimal consistency requirement that a clever bettor not be able to place a bet against you ensuring your loss of money in all circumstances; see Armendt (1993) for a lucid discussion. But for the analysis of the problem of induction, the centerpiece of de Finetti's theory was his concept of *exchangeability* and the *representation theorem.*

7.3.1. The de Finetti Representation Theorem

As we have already seen, exchangeability per se was already known to Johnson, termed by him the permutation postulate. But whereas Johnson's analysis was purely finite and predictive, de Finetti was able to go further and derive an extremely important result basic to the subsequent development of the theory.

Theorem 2 Let $X_1, X_2, \ldots$ be an infinitely exchangeable sequence of 0s and 1s, and let $S_n := X_1 + \cdots + X_n$. Then

1. The limit $Z := \lim_{n\to\infty} \dfrac{S_n}{n}$ exists almost surely;
2. If $\mu(A) = P(Z \in A)$ denotes the distribution of Z, then for every $n \geq 1$ and $0 \leq k \leq n$,

$$P(S_n = k) = \binom{n}{k} \int_0^1 p^k (1-p)^{n-k}\, d\mu(p).$$

This remarkable result has several major philosophical consequences for the theory of induction. First, contrary to popular belief, subjectivists need not reject the existence of infinite limiting frequencies (at least to the extent that they are willing to discuss infinite sequences of outcomes). *To the contrary*, if a subjectivist believes a sequence $X_1, X_2, \ldots$ to be infinitely exchangeable, then they necessarily believe that limiting frequencies exist "almost surely" (that is, with probability one). (Although they might deny that such limiting frequencies have any "objective" meaning.) Thus one has a subjective explanation for chance (objective probabilities). Second, the representation theorem tells us that the probability of seeing k successes in n trials can be expressed as a mixture of binomial probabilities

$$\binom{n}{k} p^k (1-p)^{n-k},$$

where the mixing measure $d\mu$ represents our degree of belief regarding the value of the limiting frequency of 1s in the sequence.

Note the way the theorem has been stated. In some cases, just the second part is cited: There exists a probability measure $d\mu$ on the unit interval such that the representation holds. The problem with this more limited statement is that, put this way, the mixture measure $d\mu$ appears to be merely a mathematical object or device, whereas in reality it represents a degree of belief: our judgment regarding the different possible values the limiting frequency p may assume, and the relative likelihood of each. The reader should be warned that de Finetti himself had a much more nuanced view; see (Cifarelli and Regazzini 1996).

The importance of the result cannot be understated. First, it gives a subjective explanation of chance. Second, it reduces the determination of the distribution of an infinite exchangeable sequence to that of a single probability measure on the unit interval. Finally, it gives a "principled" derivation of the standard Bayesian procedure of mixing binomials with respect to p. (It is easy to show that conditional on the value of $Z = p$, the distribution of $X_1, X_2, \ldots$ is that of a sequence of independent coin-tosses of a "p-coin.")

But the key importance of the representation theorem is its implications for inductive inference. Like Ramsey, de Finetti had abandoned the program of determining a unique descriptor of degree of belief, of finding a unique prior to describe a fanciful state of total ignorance. But given his knowledge of the representation theorem, de Finetti was also able to go further and describe induction as something other than a "useful habit." In principle, any prior on the unit interval is possible. But except for certain extreme, pathological priors, one can easily prove that the posterior distribution of p will tend to peak around the observed sample frequency $\hat{p}$ for large values of the sample size n.

Why should the future resemble the past? *If* the outcomes are judged exchangeable, then this is a simple consequence of the representation theorem. And if the outcomes are *not* exchangeable, then it is far from clear that it should. *Now* the inappositeness of the example of the rising of the sun is apparent: Successive risings of the sun are *not* exchangeable; the probability that the sun will rise today, but not rise tomorrow is not, for most people, the same as the probability that the sun will not rise today, but will rise tomorrow. (For this author the probability of the first is small, but not zero, the probability of the second is zero.)

Intersubjective Agreement The representation theorem also has consequences for intersubjective agreement. Two people may disagree initially about the likelihood of an event, but as each receives more information their posterior distributions will converge to a delta function about the sample frequency, and therefore converge to each other. The *speed* at which such convergence takes place will depend on the nature of the separate individual distributions, but not the fact of the convergence itself.

Finite Forms of de Finetti's Theorem It is possible to dispense with the appeal to infinite sequences in the above. First, as has been seen earlier, there is a finite representation theorem, but of course this does not, by itself, give the applications to induction just discussed. However, Diaconis and Freedman (1980b) have shown that if an exchangeable sequence of length m can be extended to an exchangeable sequence of length n, then the greater the value of n, the more closely probability statements regarding the initial segment of length m can be approximated by an integral representation of the de Finetti type. In this sense the infinite representation theorem can be regarded by a finitist as a purely mathematical device to give a simpler approximate form for the probabilities.

Multinomial Versions of de Finetti's Theorem More general versions of de Finetti's theorem also exist. For example, if $X_1, X_2, \ldots$ is a sequence taking t values instead of just two, and $N_1, \ldots, N_t$ represent the frequency counts for the different categories in a sample of size n, then there exists a probability measure $d\mu$ on the t-simplex Δ_t such that for every $n \geq 1$ and partition $n_1 + \cdots + n_t = n$, one has

$$P(N_1 = n_1, \ldots, N_t = n_t) = \binom{n}{n_1 \cdots n_t} \int_{\Delta_t} p_1^{n_1} \cdots p_t^{n_t} \, d\mu(\mathbf{p}).$$

7.3.2. Partial Exchangeability

If a sequence is not exchangeable, then we do not just throw in our hat. Perhaps there are other forms of symmetry present, in which case these may be exploited to derived corresponding representation theorems and rules of succession. One example is called *Markov exchangeability*; the sufficient statistics are (assumed to be) the initial state and for each pair of states i and j, the number of transitions from i to j. There is a de Finetti type representation theorem in these cases stating that such sequences (subject to a recurrence condition) are mixtures of Markov chains (Diaconis and Freedman 1980b), and a Johnson type characterization for their rules of succession (Zabell 1995). The study of this and other generalized forms of exchangeability goes back to de Finetti himself; see Diaconis and Freedman (1980a). Although de Finetti considered a few special cases, a very general theoretical framework has now been developed; see Diaconis and Freedman (1985) for the basic mathematical theory, and Jeffrey (1988) for a very readable discussion of it.

The work of Diaconis and Freedman can be thought of as a general program. As we have seen, the de Finetti representation theorem provides a subjective means of interpreting the coin-tossing parameter p in binomial sampling: Coin tossing is described in terms of an infinite exchangeable sequence of 0s and 1s; and infinite exchangeability ensures that the fraction of heads $\hat{p}$ observed in the sample converges almost surely to some (random) value p. This is a subjective interpretation of the parametric binomial sampling model. Diaconis and Freedman have identified a number of other classical parametric statistical models that have a corresponding subjectivist interpretation, using the tools of partial exchangeability and de Finetti type representations.

8. Carnap and His Successors

No discussion of the relation between probability and induction would be complete without some reference to the work of Rudolph Carnap. That our discussion here is as brief as it is reflects our focus on the technical mathematics of Ramsey and de Finetti; for a detailed appreciation of Carnap's work the reader is referred to another chapter of the author (Zabell, 2009).

It is ironic that in the decades after Johnson's death, Carnap and some of his followers would, unknowingly, reproduce much of Johnson's work. In 1945 Carnap, working in the multinomial setting, introduced the function c^* $(= P(X_{n+1} = c_i \mid \mathbf{n}))$ and proved that it had to have the (by now familiar) form $(n_i + 1)/(n + t)$ under the assumption that all "structure-descriptions" (our partitions $(n_1, n_2, \ldots, n_t)$) were equally likely; see Carnap (1945), Carnap (1950, appendix).

But just as Johnson grew uneasy with his combination postulate, so too Carnap would later introduce the "continuum of inductive methods" $\{\, c_\lambda : 0 \leq \lambda \leq \infty \,\} = (n_i + \alpha/t)/(n + \alpha)$. But while Johnson derived such expressions from his sufficientness postulate, Carnap initially *assumed* both, although his collaborator John G. Kemeny was soon after able to demonstrate their equivalence for $t > 2$. Subsequently Carnap was able to generalize these results, showing that Johnson's rule of succession follows in the case $t = 2$ from a simple linearity condition (Carnap and Stegmüller 1959), and later, in his last and posthumously published work on the subject, dropping the condition that the rules of succession be independent of the category i, and replacing it by the more general assumption that

$$P(X_{n+1} = c_i \mid n_1, \ldots, n_t) = f_i(n_i, n);$$

see Carnap (1980, section 19); see generally Kuipers (1978). For the historical evolution of this work, see Schilpp (1963, 74–75, 979–980), Carnap and Jeffrey (1971, 1–4, 223), Jeffrey (1980, 1–5, 103–104).

In the decades after Carnap's magisterial book *Foundations of Logical Probability* (1950) and his technical monograph *The Continuum of Inductive Methods* (1952) appeared, a school sprang up that systematically exploited his approach. Among the most important members of this are Hintikka, Niiniluoto, and Kuipers; see Hintikka (1966), Kuipers (1973), Hintikka and Niiniluoto (1980). Kuiper's beautiful monograph (Kuipers 1978) summarizes much of this work up to the date of its publication.

9. The Sampling of Species Problem

There is an important generalization of exchangeability that has been developed in recent decades having immediate and fascinating implications for inductive inference. This deals with the so-called sampling of species problem, in which, contrary to the classical models considered thus far, it is not assumed that the possible species are known ahead of time. A number of definitions are needed, but the reader should be reassured to learn that the final result is a theory exactly parallel to the classical one of de Finetti. The following discussion is necessarliy brief; for a more leisurely treatment, see Zabell (1992, 1998). Aldous (1985) gives a careful and systematic development of the mathematics.

9.1. Exchangeable Random Partitions

Let $n \geq 1$ be a positive integer and let $I_n = \{1, \ldots, n\}$. An *ordered partition* π of I_n is a sequence of disjoint subsets $\langle A_1, \ldots, A_t \rangle$ of I_n such that $I_n = \bigcup_i A_i$. If $n_i := n(A_i)$ is the number of elements in A_i, then the *frequency vector* corresponding to the partition is the vector

$$\mathbf{n} = \mathbf{n}(\pi) := \langle n_1, \ldots, n_t \rangle.$$

Let a_j denote the number of times the frequency j appears in the vector $\mathbf{n}$; the *partition vector* (or "allelic partition") is the vector

$$\mathbf{a} = \mathbf{a}(\pi) := \langle a_1, \ldots, a_n \rangle.$$

A *random partition* is a random quantity whose values are partitions of I_n for some fixed $n \geq 1$. A random partition Π_n of I_n is an *exchangeable random partition* if all partitions $\langle A_1, \ldots, A_t \rangle$ having the same partition vector have the same probability of occurrence; that is, if π_1 and π_2 are ordered partitions of I_n, then

$$\mathbf{a}(\pi_1) = \mathbf{a}(\pi_2) \Rightarrow P(\Pi_n = \pi_1) = P(\Pi_n = \pi_2).$$

The partition vector consists of the sufficient statistics of the partition in the case of an exchangeable random partition, just as the frequency vector consists of the sufficient statistics in the case of an exchangeable random sequence. This will be the first of many parallels between the two theories.

9.1.1. Consistent Sequences of Random Partitions

Let Π_n be a random partition of I_n. For each $1 \leq m < n$, Π_n induces in a natural way a random partition of I_m (consider the ordered partition of

I_m arising from $A_i \cap I_m, 1 \leq i \leq t$); call this induced partition $\Pi_{m,n}$. Let $\Pi_1, \Pi_2, \ldots$ be a sequence of exchangeable random partitions such that Π_n is a random partition of I_n. The sequence is said to be *consistent* if for every $m \geq 1$,

$$\Pi_m = \Pi_{m,n}$$

for every $n \geq m$.

If $\{\Pi_n : n \geq 1\}$ is an infinite consistent sequence of exchangeable random partitions, then the resulting induced random partition Π of the integers is said to be *exchangeable.*

9.1.2. Paintbox Processes

The simplest such exchangeable partitions are Kingman's *paintbox processes.* Let $Z_1, Z_2, \ldots$ be an infinite sequence of real-valued independent, identically distributed random variables. For each element $\omega \in \Omega$ of the sample space, given the sequence $Z_1(\omega), Z_2(\omega), \ldots$, one can group those times k when $Z_k(\omega)$ share a common value. For example, if one observes

$$Z_1(\omega) = 3,\ Z_2(\omega) = 1,\ Z_3(\omega) = 4,\ Z_4(\omega) = 1,\ Z_5(\omega) = 5,$$
$$Z_6(\omega) = 9,\ Z_7(\omega) = 2,\ Z_8(\omega) = 6,\ Z_9(\omega) = 5,\ Z_{10}(\omega) = 4,$$

then the *first* number observed is 3, and because this only occurs once in the sequence, at time 1, one has $A_1 = \{1\}$. Similarly, the *second* number observed is 1, and since this is observed at times 2 and 4, one has $A_2 = \{2, 4\}$. Thus, continuing in this way, one obtains the partition

$$A_1 = \{1\},\ A_2 = \{2,4\},\ A_3 = \{3,10\},$$
$$A_4 = \{5,9\},\ A_5 = \{6\},\ A_6 = \{7\},\ A_7 = \{8\}.$$

9.1.3. The Kingman Representation Theorem

During the 1970s and 1980s the English mathematician J. F. C. Kingman developed a theory of exchangeable random partitions entirely parallel to the one crafted by de Finetti (see Kingman 1975, 1978a,b, 1980). It turns out that the general infinite exchangeable random partition can be expressed as a mixture of paintbox processes, just as the general infinite exchangeable sequence can be expressed as a mixture of independent and identically distributed sequences of random variables. This is the *Kingman representation theorem.*

This representation theorem provides an amusing insight. Every probability measure on the real line can be uniquely decomposed into a *discrete* and *continuous* part. Suppose, for example, that P_1 denotes the random variable corresponding to choosing 0 or 1 with equal probability, and P_2 the probability measure that corresponds to choosing a number uniformly from the interval $(2, 3)$; and let

$$P = \frac{1}{2}P_1 + \frac{1}{2}P_2.$$

This probability measure describes a random sampling situation where the number 0 occurs 25 percent of the time, the number 1 occurs 25 percent of the time, and *distinct* numbers between 2 and 3 occur 50 percent of the time. (There is a zero probability that any two independent observations of a continuous random variable have the same value.) If each of the numbers arising as a result of the sampling are interpreted as corresponding to different species, then one species occurs 25 percent of the time, another occurs 25 percent of the time, and 50 percent of the time one sees a species that occurs *once and only once*. How can one determine whether a species corresponds to the discrete or the continuous component? (That is, whether the number representing it lies in the support of the discrete or the continuous component.) The answer is simple: As soon as one observes a member of the species a *second* time, it corresponds to the discrete component. (There is an analogy with recursive enumerability here: If the species corresponds to the discrete component, then this is ultimately discovered, because if an outcome has a positive probability of occurring, then with probability one it occurs more than once.) But if the species belongs to the continuous component, then one can never determine this for certain: If a species has been observed only one time so far in a very large sample, then this might be due to the fact that it belongs to the discrete component but has a very small probability of occurring (and therefore, with probability one, will ultimately be seen again), or that it belongs to the continuous component (and therefore, with probability one, will never be seen again).

Now we can appreciate Richard Price's insight in his appendix to Bayes's essay! Inductive inference only becomes possible on the second observation because it is only then that we know that *it corresponds to an element of the discrete component*.

But it is not in fact necessary to develop Kingman's entire theory of exchangeable random partitions to obtain counterparts of the Johnson–Carnap continua. It turns out that there is a natural analog to Johnson's axiomatic approach that gives simple and easily interpreted rules of succession.

9.2. The Pitman Continuum

Consider the following three axioms, that parallel (in two cases) or extend (in one case) those of Johnson. The first axiom is:

1. $$P(\Pi_n = \pi_n) > 0 \quad \text{for all ordered partitions } \pi_n \text{ of } I_n;$$

that is, all possible partitions are assumed possible.

Let $Z_{n+1} \in S_i$ denote the event that the $(n+1)$st observation turns out to a member of the ith species already observed. Then the second axiom is:

2. $$P(Z_{n+1} \in S_i \mid < n_1, \ldots, n_t >) = f(n_i, n), \quad 1 \leq i \leq t.$$

This states that the predictive probability of observing the ith species on the next trial is function only of the number of times that species has already been observed, and the total sample size n.

The third and final assumption is of an entirely new type, and addresses the fact that new species are always possible:

3. $$P(Z_{n+1} \in S_{t+1} \mid < n_1, \ldots, n_t >) = g(t, n).$$

That is, the probability of observing a new species (one other than the t to date already observed) is a function only of the number of species already observed and the sample size.

It is a remarkable fact that if the infinite consistent exchangeable random partition $\Pi_1, \Pi_2, \ldots$ satisfies the above three hypotheses, then one can prove that the functions $f(n_i, n)$, $g(t, n)$ are members of a three-dimensional continuum described by three parameters α, θ, γ.

The Continuum of Inductive Methods for the Sampling of Species

Case 1: If $n_i < n$ for some i, then

$$f(n_i, n) = \frac{n_i - \alpha}{n + \theta}, \quad g(t, n) = \frac{t\alpha + \theta}{n + \theta}.$$

Note that if $n_i < n$, then $t > 1$, there are at least two species, and the universal generalization is *dis*confirmed.

Case 2: If $n_i = n$ for some i, then

$$f(n_i, n) = \frac{n_i - \alpha}{n + \theta} + c_n(\gamma), \quad g(t, n) = \frac{t\alpha + \theta}{n + \theta} - c_n(\gamma);$$

here

$$c_n(\gamma) = \frac{\gamma(\alpha + \theta)}{(n + \theta)\left[\gamma + (\alpha + \theta - \gamma)\prod_{j=1}^{n-1}\left(\frac{j-\alpha}{j+\theta}\right)\right]}$$

represents the increase in the probability of seeing the ith species again due to the confirmation of the universal generalization. Not all parameter values are possible: One must have

$$0 \le \alpha < 1; \quad \theta > -\alpha; \quad 0 \le \gamma < \alpha + \theta.$$

There is a simple interpretation of the three parameters θ, α, γ. The first, θ, is related to the likelihood of new species being observed; the larger the value of θ, the more likely it is that the next observation is that of a new species. Observation of a new species has a double inductive import: It is a *new* species, and it is a *particular* species. Observing it contributes (via α) both to the likelihood that a new species will again be observed and, if a new species is not observed, that the species just observed will again be observed (as opposed to another of the species already observed). Finally, the parameter γ is related to the likelihood that only one species will be observed. If ϵ is the initial probability that there will only be one species, then $\gamma = (\alpha + \theta)\epsilon$. The special case $\alpha = \gamma = 0$ is of particular interest. In this case the probability of an allelic partition has a particularly simple form: Given a sample of size n,

$$P(a_1, a_2, \ldots, a_n) = \frac{n!}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{r=1}^{n} \frac{\theta^{a_r}}{r^{a_r} a_r!};$$

this is the *Ewens sampling formula.* There is a simple urn model for such a process in this case, analogous to the Polya urn model (Hoppe 1984). Suppose we start out with an urn containing a single black ball: the *mutator.* The first time we select a ball, it is necessarily the black one. We replace it, together with a ball of some color. As time progresses, the urn contains the mutator and a number of colored balls. Each colored ball has a weight of one, the mutator has weight θ. The likelihood of selecting a ball is proportional to its weight. If a colored ball is selected, it is replaced together with a ball of the same color; this corresponds to observing a species that has already been observed before (hence balls of its color are already present). If the mutator is selected, it is replaced, *together with a ball of a new color*; this corresponds to observing a new species. It is not difficult to verify that the rules of succession for this process are

$$f(n_i, n) = \frac{n_i}{n+\theta}; \quad g(n) = \frac{\theta}{n+\theta}.$$

Note that in this case the probability of a new species does not depend on the number observed. Such predictive probabilities arguably go back to De Morgan; see Zabell (1992).

10. Conclusion

The quantitative theory of inductive inference only became possible after the rise of mathematical probability. Initially the technical developments in probability ran ahead of its careful philosophical analysis (with the exception of Bayes's essay), but in the twentieth century, thanks to luminaries such as William Ernest Johnson, Frank Plumpton Ramsey, and Bruno de Finetti, a coherent theory and analysis of the inductive process became possible. Eight decades of discussion, analysis, and criticism of that theory has led to a rich structure in which a solution to Hume's problem is finally possible.

This is not to say that the theory has not had its critics, most notably Karl Popper and his school. But the theory has had a resilience surpassing each of its frequentist, propensity, and credibilist competitors. Although every theory must introduce assumptions, simplifications, and idealizations, the modern theory, properly understood, is a useful and impressive contribution to the philosophical clarification of the justification of one of our most basic forms of reasoning.

Appendix: Gamma and Beta

The mathematics underlying the evaluation of certain integrals needs to be briefly recalled.

The Gamma Function

First, one has the *gamma function*

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt, \quad x > 0.$$

The gamma function satisfies the *duplication formula*

$$\Gamma(x+1) = x\,\Gamma(x), \quad x > 0,$$

and is therefore completely determined by its values in the interval $0 < x \leq 1$. In particular, because $\Gamma(1) = 1$, one has

$$\Gamma(n+1) = n!$$

for all integers $n \geq 0$. It is therefore an extension of the integer factorial function to all positive reals.

The Beta Function

A close relative and friend of the gamma function is the (two-parameter) *beta function*:

$$B(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\, dt, \quad \alpha, \beta > 0.$$

The key to its computation is the formula

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

The Beta Distribution

Because every integration formula for a nonnegative function enables one to define a probability density, we can now define the important two-parameter *beta family* of densities:

$$f_{\alpha,\beta}(p) := \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}, \quad 0 \leq p \leq 1 \quad (\alpha, \beta > 0).$$

Note that if $\alpha = \beta = 1$, this reduces to the uniform distribution on the unit interval.

References

Aldous, David. 1985. Exchangeability and related topics. In *École d'Été de Probabilités de Saint-Flour XIII—1983*, ed. P. L. Hennequin, 1–198. Lectures Notes in Mathematics, vol. 1117. New York: Springer-Verlag.

Arbuthnot, John. 1710. An argument for divine providence, taken from the constant regularity observed in the births of both sexes. *Philosophical Transactions of the Royal Society of London* 27: 18–190.

Armendt, Brad. 1993. Dutch books, additivity, and utility theory. *Philosophical Topics* 21: 1–20.

Baker, Keith Michael. 1975. *Condorcet: From Natural Philosophy to Social Mathematics*. Chicago: University of Chicago Press.

Bayes, Thomas. 1764. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London* 53: 370–418. (Reprinted 1958, *Biometrika* 45: 292–315; page citations in the text are to this edition.)

Bellhouse, David. 2004. The Reverend Thomas Bayes, FRS: a biography to celebrate the tercentenary of his birth. *Statistical Science* 19: 3–43.

Braithwaite, R. B., ed. 1931. *The Foundations of Mathematics and Other Logical Essays*. London: Routledge and Kegan Paul.

Carnap, Rudolph. 1945. On inductive logic. *Philosophy of Science* 12: 72–97.

Carnap, Rudolph. 1950. *Logical Foundations of Probability*. Chicago: University of Chicago Press.

Carnap, Rudolph. 1980. A basic system of inductive logic, part ii. In *Studies in Inductive Logic and Probability*, ed. Richard C. Jeffrey, vol. 2, 7–155. Berkeley and Los Angeles: University of California Press.

Carnap, Rudolph, and Richard C. Jeffrey. 1971. *Studies in Inductive Logic and Probability*, vol. 1. Berkeley and Los Angeles: University of California Press.

Carnap, Rudolph, and W. Stegmüller. 1959. *Inductie Logik und Wahrscheinlichkeit*. Vienna: Springer-Verlag.

Cifarelli, Donato Michele, and Eugenio Regazzini. 1996. De Finetti's contribution to probability and statistics. *Statistical Science* 11: 253–282.

Dale, Andrew I. 2003. *Most Honourable Remembrance: The Life and Work of Thomas Bayes*. Sources and Studies in the History of Mathematics and Physical Sciences. New York: Springer-Verlag.

Dawid, Philip, and Donald Gillies. 1989. A Bayesian analysis of Hume's argument concerning miracles. *Philosophical Quarterly* 39: 57–65.

de Finetti, Bruno. 1931. Probabilismo. *Logos* 14: 163–219. English translation, *Erkenntnis* 31 (1989): 169–223.

de Finetti, Bruno. 1937. La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré* 7: 1–68. (Translated in H. E. Kyburg, Jr. and H. E. Smokler [eds.], *Studies in Subjective Probability*, New York: Wiley, 1964, pp. 93–158).

de Finetti, Bruno. 1972. *Theory of Probability*, vols. 1 and 2. New York: Wiley.

De Morgan, Augustus. 1847. *Formal Logic: Or the Calculus of Inference Necessary and Probable*. London: Taylor and Walton.

Diaconis, Persi, and David Freedman. 1980a. De finetti's generalizations of exchangeability. In *Studies in Inductive Logic and Probability*, ed. Richard C. Jeffrey, vol. 2, 233–249. Berkeley and Los Angeles: University of California Press.

Diaconis, Persi, and David Freedman. 1980b. De Finetti's theorem for Markov chains. *Annals of Probability* 8: 115–130.

Diaconis, Persi, and David Freedman. 1985. Partial exchangeability and sufficiency. In *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Institute Golden Jubilee International Conference*, eds. J. K. Ghosh and J. Roy, 205–236. Calcutta: Indian Statistical Institute.

Diaconis, Persi, and Sandy L. Zabell. 1982. Updating subjective probability. *Journal of the American Statistical Association* 77: 822–830.

Earman, John. 1993. Bayes, Hume, and miracles. *Faith and Philosophy* 10: 293–310.

Earman, John. 2000. *Hume's Abject Failure: The Argument against Miracles*. New York: Oxford University Press.

Edwards, A. W. F. 1978. Commentary on the arguments of Thomas Bayes. *Scandinavian Journal of Statistics* 5: 116–118.

Ellis, Richard Leslie. 1844. On the foundations of the theory of probabilties. *Transactions of the Cambridge Philosophical Society* 8: 1–6.

Feller, William. 1968. *An Introduction to Probability Theory and Its Applications*, vol. 1, 3rd ed. New York: Wiley.

Fisher, Ronald A. 1973. *Statistical Methods and Scientific Inference*, 3rd ed. New York: Hafner Press.

Freedman, David, and Roger Purves. 1969. Bayes method for bookies. *Annals of Mathematical Statistics* 40: 1177–1186.

Gillies, Donald. 1987. Was Bayes a Bayesian? *Historia Mathematica* 14: 325–346.

Gillispie, Charles Coulston. 1972. Probability and politics: Laplace, Condorcet, and Turgot. *Proceedings of the American Philosophical Society* 116: 1–20.

Good, I. J. 1988. Bayes's red billiard ball is also a herring, and why Bayes withheld publication. *Journal of Statistical Computation and Simulation* 29: 335–340.

Gower, Barry. 1990. David Hume and the probability of miracles. *Hume Studies* 16: 17–31.

Gower, Barry. 1991. Hume on probability. *British Journal for the Philosophy of Science* 42: 1–19.

Hacking, Ian. 1967. Slightly more realistic personal probability. *Philosophy of Science* 34: 311–325.

Hacking, Ian. 1971. James Bernoulli's "Art of Conjecturing." *British Journal for the Philosophy of Science* 22: 209–229.

Hintikka, J. 1966. A two-dimensional continuum of inductive methods. In *Aspects of Inductive Logic*, eds. J. Hintikka and P. Suppes, 113–132. Amsterdam: North-Holland.

Hintikka, J., and I. Niiniluoto. 1980. An axiomatic foundation for the logic of inductive generalization. In *Studies in Inductive Logic and Probability*, ed. R. C. Jeffrey, vol. 2, 157–181. Berkeley: University of California Press.

Holder, Rodney D. 1998. Hume on miracles: Bayesian interpretation, multiple testimony, and the existence of God. *Journal of the Royal Statistical Society B* 49: 49–65.

Hoppe, Fred. 1984. Polya-like urns and the Ewens samapling formula. *Journal of Mathematical Biology* 20: 91–94.

Jeffrey, Richard C., ed. 1980. *Studies in Inductive Logic and Probability*, vol. 2. Berkeley: University of California Press.

Jeffrey, Richard C. 1983. *The Logic of Decision*, 2nd ed. Chicago: University of Chicago Press.

Jeffrey, Richard C. 1988. Conditioning, kinematics, and exchangeability. In *Causation, Chance, and Credence*, eds. W. L. Harper and B. Skyrms, vol. 1, 221–255. Dordrecht: Kluwer.

Jevons, William Stanley. 1874. *The Principles of Science: A Treatise on Logic and Scientific Method*. London: Macmillan (2nd ed. 1877; reprinted 1958, New York: Dover).

Johnson, William Ernest. 1924. *Logic, Part III: The Logical Foundations of Science*. Cambridge: Cambridge University Press.

Johnson, William Ernest. 1932. Probability: The deductive and inductive problems. *Mind* 41: 409–423.

Keynes, John Maynard. 1921. *A Treatise on Probability*. London: Macmillan.

Kingman, J. F. C. 1975. Random discrete distributions. *Journal of the Royal Statistical Society B* 37: 1–22.

Kingman, J. F. C. 1978a. Random partitions in population genetics. *Proceedings of the Royal Society A* 361: 1–20.

Kingman, J. F. C. 1978b. The representation of partition structures. *Journal of the London Mathematical Society A* 18: 374–380.

Kingman, J. F. C. 1980. *The Mathematics of Genetic Diversity*. Philadelphia: SIAM.

Kolmogorov, Andrei Nikolaevich. 1933. *Grundbegriffe der Wahrscheinlichkeitrechnung*. Berlin: Springer-Verlag.

Kuipers, Theo A. F. 1973. A generalization of Carnap's inductive logic. *Synthese* 25: 334–336.

Kuipers, Theo A. F. 1978. *Studies in Inductive Probability and Rational Expectation*. Dordrecht: Reidel.

Langtry, Bruce. 1990. Hume, probability, lotteries and miracles. *Hume Studies* 16: 67–74.

Laplace, Pierre Simon Marquis de. 1774. Mémoire sur la probabilité des causes par les évenements. *Mémoires de l'Académie royale des sciences presentés par divers savans* 6: 621–656. (Reprinted in *Oeuvres complètes de Laplace* (1878–1912), vol. 8, pp. 27–65, Paris: Gauthier-Villars).

Levine, Michael. 1998. Bayesian analyses of Hume's argument concerning miracles. *Philosophy and Theology* 10: 101–106.

Mura, Alberto. 1998. Hume's inductive logic. *Synthese* 115: 303–331.

Murray, F. H. 1930. Note on a scholium of Bayes. *Bulletin of the American Mathematical Society* 36: 129–132.

Oresme, Nicole. 1966. *Die proportionnibus proportionum and Ad pauca respicientes*. University of Wisconsin Press. (E. Grant ed. and trans.)

Owen, David. 1987. Hume versus price on miracles and prior probabilities: Testimony and the Bayesian calculation. *Philosophical Quarterly* 37: 187–202.

Ramsey, Frank Plumpton. 1931. Truth and probability. In *The Foundations of Mathematics and Other Logical Essays*, ed. R. B. Braithwaite, 156–198. London: Routledge and Kegan Paul.

Salmon, Wesley C. 1978. Religion and science: A new look at Hume's *Dialogues*. *Philosophical Studies* 33: 143–176.

Schilpp, P. A., ed. 1963. *The Philosophy of Rudolph Carnap*. La Salle, Ill.: Open Court.

Shafer, Glenn. 1982. Bayes's two arguments for the rule of conditioning. *Annals of Statistics* 10: 1075–1089.

Sobel, Jordan Howard. 1987. On the evidence of testimony for miracles: A Bayesian interpretation of David Hume's analysis. *Philosophical Quarterly* 37: 166–186.

Sobel, Jordan Howard. 1991. Hume's theorem on testimony sufficient to establish a miracle. *Philosophical Quarterly* 41: 229–237.

Stigler, Stephen M. 1982. Thomas Bayes's Bayesian inference. *Journal of the Royal Statistical Society Series A* 145: 250–258.

Stove, David C. 1973. *Probability and Hume's Inductive Skepticism.* Oxford: Clarendon Press.

Strong, John V. 1976. The infinite ballot box of nature: De Morgan, Boole, and Jevons on probability and the logic of induction. In *PSA 1976: Proceedings of the Philosophy of Science Association*, vol. 1, 197–211.

Venn, John. 1866. *The Logic of Chance.* London: Macmillan. (2nd ed. 1876, 3rd ed. 1888; reprinted 1962, New York: Chelsea).

Zabell, S. L. 1989a. R. A. Fisher on the history of inverse probability. *Statistical Science* 4: 247–263.

Zabell, S. L. 1989b. The rule of succession. *Erkenntnis* 31: 283–321.

Zabell, S. L. 1991. Ramsey, truth, and trobability. *Theoria* 57: 211–238.

Zabell, S. L. 1992. Predicting the unpredictable. *Synthese* 90: 205–232.

Zabell, S. L. 1995. Characterizing Markov exchangeable sequences. *Journal of Theoretical Probability* 8: 175–178.

Zabell, S. L. 1998. The continuum of inductive methods revisited. In *The Cosmos of Science: Essays of Exploration*, eds. John Earman and John D. Norton, 351–385. Pittsburg-Konstanz Series in the Philosophy and History of Science. University of Pittsburgh Press/Universitätsverlag Konstanz.

Zabell, S. L. 2009. Carnap and induction. In volume 10 (Inductive Logic) of the "Handbook of the History of Logic", to appear.

16

Logic and Linguistics in the Twentieth Century

Alessandro Lenci and Gabriel Sandu

1. Introduction

A crucial aspect of the revolution that affected logic at the beginning of the twentieth century concerns the severance of its traditional dependence on the form and structure of natural language. Such a breakdown has had enormous consequences not only for the development of formal logic but also for the opening of new perspectives in the study of language. This peculiar relationship between mathematical logic and language inquiry is best illustrated by Willard V. O. Quine (1961, 1):

> Mathematicians expedite their special business by deviating from ordinary language. Each such departure is prompted by specific considerations of utility for the mathematical venture afoot. Such reforms may be expected to reflect light on the ordinary language from which they depart.

As a major consequence of its "reforms," the new mathematical logic has been able to revivify and boost the notion of *philosophical* and *logical grammar*, typical of the seventeenth- and eighteenth-century rationalist tradition. New life has therefore been given to the idea that there exists a common grammatical core shared by every language and determinable a priori, with respect to which diversity and variation are just prima facie features of natural language, hiding its universal logical structure.

The dominant and almost exclusive role of logic in the quest for the universal principles of human language has however been radically challenged in the course of the century by the arising and fast growth of *generative* linguistics, which has claimed the investigation of universal grammar as the major goal of linguistic science. In contrast with the descriptive and empiricist approach of structural linguistics—mostly focused on the taxonomic study of particular languages—Noam Chomsky has reaffirmed the need for a rationalist perspective on language analysis. The focus on the formal properties of language structure, the central role occupied by the study of syntax and of language creativity, the attention to the relation between syntactic composition and meaning, the inquiry into the universal principles that form the conditions of the possibility of human language have thus become the core areas of research in generative theoretical linguistics, thereby granting an important convergence with the research and the aims pursued in the field of logical grammar.

The Chomskian revolution has therefore deeply affected the relationship between logic and linguistics, the latter being intended as a naturalistic scientific enterprise subject to the same methodological requirements of other *Naturwissenschaften* like physics or chemistry. On one hand, the rationalist turn in linguistics has actually allowed for an unprecedented convergence of linguistics with important areas of mathematical logic. On the other hand, the generative paradigm has also set constraints on the study of natural language and formulated hypotheses on its architecture that have often dramatically conflicted with the logicomathematical approach. Thus, the history of the relationship between logic and theoretical linguistics in the past decades is rather a deeply dialectic one. It is a history of profound and synergic efforts toward the common aim of understanding the nature and universal principles of human language and its formal structure, but it is also an history of harsh conflicts and divergences on the nature of universal grammar itself. At the core of this confrontation lies the issue of the relationship between grammatical form and logical form, that is to say, the possibility itself of carving out the natural language syntactic and semantic space as a logical space. This is also primarily related to the role of meaning in the architecture of language, that is, its methodological function in guiding the discovery of language universals, the proper nature of a semantic theory, and the relationship between syntactic descriptions and semantic interpretation. Such questions have reached their peak with the debate on the *principle of the autonomy of syntax*, which was proposed by Chomsky since the very outset of the generative enterprise, and has largely dominated and oriented the discussion throughout this period.

Logic and generative linguistics have partly been divided by the inherently cognitive and psychological orientation of the rationalist approach boosted by Chomsky. While mathematical logic has mostly focused on a mind-independent, speaker-independent notion of language, generative linguistics is, in fact, ultimately interested in the principles of language intended as the description of a particular cognitive faculty of speakers. Moreover, the tradition of logical grammar—with the major landmark of Richard Montague's contribution—is largely dominated by the hypothesis that actually no substantial difference exists between natural language and formal languages (Montague 1974). Differently, the major constraint imposed by Chomsky on the principles of universal grammar is rather an empirical one, that is, whether they provide a real explanation of speakers' knowledge of language and of men's innate capability to acquire language. Therefore, the history of the interactions between these disciplines is also deeply related to a fully theoretical issue, namely, how much is language amenable to a treatment as a formal language. In this way, reconstructing the difficult and contrasted history of the interactions between logic and contemporary linguistics leads us to set and investigate crucial philosophical questions concerning the nature of language and the way to face its complexity.

The interaction between logic and linguistics on the nature of universal grammar can be roughly divided into three phases, which will be analyzed in details in the next sections. The first phase (beginning of the century up to the 1960s) was characterized by an extremely intense work in the field of logical grammar (section 2), with the rising of categorial grammar within the Polish School in the first decades of the century and its extensive application to ordinary language by Yehoshua Bar-Hillel in the early 1950s. Besides, the work on truth-conditional semantics by Tarski, Quine, and Davidson (section 3) provided the necessary background to the model-theoretic analyses of natural language and to Montague grammar in the 1970s. On the linguistics side, in 1957 the transformational generative paradigm made its first steps out of the banks of American post-Bloomfieldian structuralism and behaviorism (section 4). Chomsky's critique of the inadequacies of phrase structure grammars had a strong impact on the tradition of logical grammar, by revealing the limits and problems of categorial models. Moreover, in these years the generative architecture of the universal grammar received its first shaping, accompanied by the initial steps of the debate on the role and nature of semantic theory.

The second phase (the late 1960s, throughout the 1970s) began with the crisis of the semantic models developed in the early period of generative grammar and the rise and fall of the generative semantics enterprise (section 5). One of the major events of this period was the explosion of Montague grammar and the subsequent breakthrough made in the linguistic community by the development of model-theoretic semantics (section 6). More generally, these years were characterized by a great debate on the proper position of semantics within the theory of grammar, and by the first attempts to carry out extensive comparisons and integrations between generative linguistics and logical grammar.

The third phase (starting from the beginning of the 1980s) is best illustrated by referring to the central role acquired in the Government and Binding version of Chomsky's theory by the notion of logical form (LF) (section 7), resulting in an intense work in linguistics on topics like quantification, coreference, and so on, with the consequent constant readjustment of the border between logic and formal linguistics.

Before starting, let us make a final general remark. Talking about logic and linguistics in the twentieth century unavoidably entails presenting theories and models which often not only belong to the history of these disciplines but also form the daily working tools for militant researchers and scholars. This is exactly what happens with Montague grammar and generative linguistics. Therefore, the following sections are not intended to represent general introductions or descriptions of any of these frameworks, for which very good manuals and summarizations exist at various degrees of difficulty and for various audiences. The discussion will rather be focused on trying to reconstruct the dynamics and interactions between these approaches in logic and in linguistic theory, which represent the major landmarks in the quest for the individuation of the universal structure of language.

2. Logical Grammar

2.1. Frege and Russell

For Frege, the most important concept of logic is truth. It is the analysis of this notion that forced him to create a theoretical framework in which sentences are broken into parts which in turn are related to entities in the universe in a systematic way. By taking truth as the focal point of his considerations, Frege was the initiator of a semantic project that was to dominate the logical study of language up to the present day. Even if semantics proper was established as a discipline much later by Tarski (who also introduced the Polish analog of the name in 1936), Frege saw probably more clearly than any other how an analysis of the truth of a sentence compels one to introduce meaning (referential) relations between expressions of the sentence and extralinguistic entities. Frege's views about the logical analysis of language constituted a complete break with the tradition. Before him it was customary to write things like this:

> Every categorial proposition has a subject, a predicate, a copula, a quality and a quantity. Subject and predicate are called "terms." For example, in "the pious man is happy," "the pious man" and "happy" are terms of which "the pious man" is the subject, "happy" is the predicate, and "is" is the copula. The "quality" of the proposition is affirmation or negation ... the "quantity" of a proposition is its universality or particularity. (Leibniz, *Opuscules et fragments inedits de Leibniz*, 77–78)

Nothing could be further from Frege. He rejected explicitly the distinction between subject and predicate (something that will find an echo later on in Chomsky's redefinition of grammatical subject; see section 4.2):

> A distinction between *subject* and *predicate* finds *no place* in my representation of a judgment. Now all those features of language that results only from the interaction of the speaker and listener ... have no counterpart in my formula language, since here the only thing that is relevant in a judgement is that which influences

> its possible consequences. [In this I closely follow the formula language of mathematics, in which subject and predicate can also be distinguished only by violating it]. (Frege 1879, §3)

Instead, he introduced (Frege 1891) the distinction much more familiar to him from his training as a mathematician between *object* and *function.* According to it, the sentence "John is tall" is to be analyzed into a concept word "tall" and the proper name "John." The latter designates an object, the bearer of the proper name. The former designates a concept, that is, a function that for Frege is an unsaturated entity whose arguments are objects and whose values are the truth values *True* or *False.* Thus in our example, the concept-word "tall" designates the concept which, for every object as its argument yields the truth value *True* if and only if that object is tall; the whole sentence designates *True* if and only if the individual designated by "John" is tall.

The rejection of the subject-predicate distinction can be seen even more clearly in the case of relational expressions. The statement "3 is greater than 2" ("$3 > 2$") is not to be analyzed into the subject "3" and the predicate "is greater than 2" but into the relation symbol "is greater than" and the proper names "2" and "3" (Frege 1891, 154).

Frege's analysis of quantifiers ("signs of generality") constituted another break with the tradition. His predecessors regarded the sentences

(∗) "Socrates is mortal"

and

(∗∗) "Everyone is mortal"

as having the same logical complexity, that is to say, (∗) was regarded as equivalent to "Every Socrates is mortal," where "Socrates" is a term denoting one single object. Frege saw things in a completely different way. (∗) is for him an atomic sentence built up from the proper name "Socrates" and the concept-word "mortal," while (∗∗) is a statement of generality in which the (first-level) concept designated by "mortal" is the argument of the (second-level) concept designated by the sign of generality "everyone."

> It must be here remarked that the words "all," "any," "no," "some," are prefixed to concept-words. In universal and particular affirmative and negative sentences, we are expressing relations between concepts; we use these words to indicate the special kind of relations. They are thus, logically speaking, not to be more closely associated with the concept-words that follow them, but are to be related to the sentence as a whole. (Frege 1892, 187)

In Frege's perspicuous notation, introduced in his *Begriffsschrift*, the position and role of every expression as well as their level are clearly specified. Thus the general form of statements of generality like (∗∗) is

$$\underbrace{\;a\;}\; f(a),$$

where "f" indicates the place of the first-level function which is its argument (Frege 1891, 153). The second-level concept designated by "every" is defined as the function that takes the value *True* for a first-level concept as its argument (to be inserted in the position indicated by "f") if and only if the first-level concept takes the value *True* for every object in the universe; otherwise it has the value *False*.

It is worth emphasizing that Frege's views on logic together with the conceptual notation that went along with it opened the door for possibilities undreamt of by his predecessors, which found their place into both logic and linguistics once and for all. Let us focus on those which are relevant for the present study.

First, we have in Frege for the first time the idea of the derivational history of a sentence with the engendered possibility of determining its truth or falsity in stages, beginning with the atomic stage. Frege's procedure is not entirely compositional with respect to *truth*, that is, the truth of a compound sentence cannot be obtained from the truth of the *compounds*, and this for the simple reason that the compounds are not always sentences. Since for Frege variables are empty places which indicate the positions where arguments must be filled in, and not terms receiving meaning through an assignment, he could not and did not have the notion of satisfaction available, which had to wait for Tarski's work. But he still could define, and in fact he did, as Dummett has observed, the *truth* of a compound sentence *in stages*. Thus the sentence "everyone is mortal" is *True* if and only if "George is mortal" and "John is mortal," and so on, that is, if and only if the first-level concept word "mortal" yields the value *True* when we run through all the (names of) objects in the universe which are persons.

Second, Frege's categorical distinction between objects and concepts, and the syntactical distinction between complete and incomplete expressions that goes along with it gave rise to a hierarchy of levels which, in turn, yields a theory of signification for natural language sentences. In other words, he was able to explain why certain sentences in natural language, although grammatical, are completely meaningless or paradoxical. As Dummett pointed out, the failure of significance of such sentences is accounted for by the impossibility of constructing a corresponding sentence in the symbolic language. We find essentially the same idea later on in Russell's theory of types as well as in the categorical languages of the Polish school (Ajdukiewicz and Lesniewski) which grew out of Husserl's work. The explanation goes shortly like this.

At the basis (level 0) of Frege's hierarchy of levels, we have complete names, that is, proper names and sentences, while all the expressions situated at higher levels are incomplete. Thus at the first level we have one-place predicate expressions of first-level which are incomplete expressions obtained from sentences by removing one or more occurrences of a proper name. Frege used the notation

$$\dots\ \xi\ \dots$$

to denote a first-level one-place predicate which has been formed from a sentence

$$\ldots\ a\ \ldots$$

by removing one or more occurrence of some proper name "a" and leaving a gap indicated by the Greek letter ξ to mark the argument place of the predicate. On the same level we also have two-place, three-place, and so on relational expressions of first-level and also one-place, two-place, and so on functional expressions of first-level, that is, incomplete expressions obtained from proper names by removing one or more occurrences of other proper names.

At the next level we find second-level predicates, that is, incomplete expressions obtained from sentences by removing one or more occurrences of first-level predicates. Typical examples are quantifiers, like our earlier example "$\underbrace{\ a\ }\ f(a)$," where "f" shows the gap left by the removal of the first-level predicate, and the expressions "a" in brackets shows the initial argument of the removed first-level predicate.

Suppose now that we want to insert a new first-level predicate in the argument place of the second-level relation. We will have to put that predicate in the place of f and then insert a as its argument. This mechanism shows why we cannot insert a proper name in the argument place of a second-level predicate: There is no place for the argument left by the removal of the first-level predicate to go to. For similar reasons, we do not obtain a sentence when we insert a first-level predicate into the argument place of another first-level predicate, for the resulting expression still contains a gap ξ to be filled in.

We are now able to understand what is paradoxical about natural language sentences like "The concept horse is not a concept" if interpreted as saying something about a concept or, to borrow an example from Dummett, why certain natural language sentences like "Chairman Mao is rare" are perfectly grammatical, yet meaningless. The reason is that "is not a concept" denotes a second-level concept although it appears in the grammar of English as a first-level one. And the same goes for "is rare" in the second example. In other words, what in the grammar of natural language appears like a first-level predicate, is not so in logic:

> The concept of a function must be a second-level concept, whereas in language it always appears as a first-level concept. While I am writing this, I am well aware of having again expressed myself imprecisely. Sometimes this is just unavoidable. All that matters is that we know we are doing it, and how it happens. In a conceptual notation, we can introduce a precise expression for what we mean when we call something a function (of the first level with one argument). (Frege, letter to Russell, 1902)

Dummett has drawn attention to the fact that Frege's hierarchy of levels is essentially the same as Russell's theory of simple types formulated in terms of Frege's notion of incomplete expressions. Although Russell does not explicitly make the distinction between complete and incomplete expressions, we find in his notion of "ambiguity" essentially the same notion of incomplete expression and the same criteria of significance as in Frege:

> Thus "$(x).\varphi x$," which we have already considered, is a function of φx; as soon as φx is assigned, we have a definite proposition, wholly free from ambiguity. But it is obvious that we cannot substitute for the function something which is not a function: "$(x).\varphi x$" means "φx in all cases," and depends for its significance upon the fact that there are "cases" of φx, i.e., upon the ambiguity, which is characteristic of a function. This instance illustrates the fact that, when a function can occur significantly as argument, something which is not a function cannot occur significantly as argument. But conversely, when something which is not a function can occur significantly as argument, a function cannot occur significantly. (Whitehead and Russell, *Principia Mathematica*, 48).

Later on we shall regain Frege's hierarchy of levels in the Lesniewski–Ajdukiewicz grammar of semantic categories. As we said at the beginning of this section, truth was for Frege the main concept of logic, and truth is a property of sentences and thoughts. Thoughts and sentences were thus primary for Frege, but this did not prevent him to realize the combinatorial and compositional power of language, a methodological credo that was to remain constantly transparent in his writings and after him was going to mark, if not even demarcate (see section 4.1), the project of logical grammar of the Polish school, Carnap, Davidson, and Montague from other developments in the study of language. Indeed, Frege wrote:

> It is astonishing what language can do. With a few syllables it can express am incalculable number of thoughts, so that even if a thought has been grasped by an inhabitant of the Earth for the very first time, a form of words can be found in which it will be understood by someone else to whom it is entirely new. (Frege 1923–1926, 390)

2.2. Husserl's Theory of Meaning Categories

Husserl was directly concerned with the question of what makes natural language expressions significant. The answer he gave to this question is essentially the same as that given by Frege before him: in virtue of these expressions obeying the principles of combination and substitution governing the meaning categories they belong to. Thus like Frege, Husserl makes categorical distinctions and states explicitly the connections between expressions belonging to different categories. These connections are codified in the so-called *meaning connection rules*, which state the mode of combination and substitution of different expressions into more complex ones. These rules allow Husserl to explain why combinations of certain strings in language are nonsensical. The purely *logical grammar* is the set of a priori laws common to all languages.

To understand Husserl's meaning connection rules, we have first to understand his distinction between *form* and *matter*, that is, between expressions signifying forms and expressions signifying matters. In the sentence

(1) This house is green.

the words *this* and *is* do not have an independent meaning: They are *syncategorematic* expressions, that is, expressions that become meaningful only after completion with other expressions. For Husserl, *syncategorematic* expressions signify *forms*, in contradistinction to nominal expressions, like *house* and adjectival expressions like *green* which signify *matters*, that is, things and entities in the world, and so on. He perspicuously observed that in (1) we can substitute nominal matters for *house* and adjectival matters for *green* and the result is an expression which is still "well formed," or, in Husserl's words, it has a unitary meaning. So in (1) we can discern an underlying propositional form

(2) This S is p

which yields unitary meaning only if substitute for the variables S and p expressions belonging to the same *Bedeutungskategorien* (*meaning category*, as distinguished from the term *semantic category* used later by the Polish school). (See also Casadio 1988.)

Each such form has associated with it a *meaning connection rule* which states to which meaning categories the expressions substitutable for the variables of the form must belong. In Husserl's words:

> each primitive form adheres to a certain a priori ... law stating that every meaning connection obeying that form effectively gives rise to a unitary meaning, provided that the terms (the underdetermined elements, the variables of the form) belong to certain meaning categories. (Husserl 1913, 330)

In the case of (2), the meaning connection law states that any nominal matter may be substituted for S and any adjectival matter may be substituted for p.

If in a form we violate the meaning connection rule by substituting for the variables in the form words belonging to inappropriate categories, the resulting expression turns out to be nonsignificant or *nonsense* (Unsinn). This happens, for instance, if in (2) we substitute for S an adjective like *careless* and for p an adjective like *green* (which is appropriate). However, even if we obey the meaning connection rules, we may get an absurd expression like

(3) This quadrilateral has 5 vertices.

which does not denote a possible state of affairs (Husserl 1913, 327). This is a case of *countersense* (Widersinn). The distinction between nonsense and countersense justifies Husserl to introduce two kinds of laws: laws of avoiding nonsense and laws of avoiding formal countersense (Husserl 1913, 334–335). We see that for Husserl, nonsense is prevented by the meaning-connection rules.

Bar-Hillel has made the interesting observation that Husserl's distinction between nonsense and countersense is an anticipation of Carnap's distinction between *formation rules* and *transformation rules* (Bar-Hillel 1970, 93; Casadio 1988, 116). In Carnap's *Logical Syntax of Language* (1937), the former define the well-formed expressions (sentences) of a language, and the latter define the set of sentences that are consequences of a system of axioms. In this setting, the sentence (3) cannot be true, whereas the sentence *This careless is green* is not a sentence at all. According to Bar-Hillel, Husserl's insight that the rules of avoiding nonsense are logically prior to the rules of avoiding countersense is nothing else that the Carnapian requirement that the statement of the rules of formation has to precede its rules of transformation (Bar-Hillel 1970, 93–94).

The Husserlian distinction between form and matter reminds one of the Fregean distinction between complete and incomplete expressions and the ontological distinction between objects and concepts which goes along with it. Some of the incomplete expressions became later in the Polish school the *functorial* (*operator*) categories.

2.3. The Polish School

The Polish school gathered philosophers and logicians who worked in Lwow, Warsaw, and Krakow between the two wars. However, the history of the group starts much earlier with Twardowski, who attended Brentano's lectures. Twardowski taught in Lwow and so did his pupil, Lukasiewicz. Among the students of the latter one could find Leśniewski, Ajdukiewicz, and Kotarbinski. Leśniewski and Kotarbinski moved later to Warsaw where a new generation of logicians was raised, including Lindebaum, Sobocinski, and Tarski.

One of the main problems considered by many of the logicians in this group was to give an adequate answer to the question raised by Husserl, namely, "the specification of the condition under which a word pattern, constituted of meaningful words, forms an expression which itself has a unified meaning.... A word pattern of this kind is syntactically connected" (Ajdukiewicz 1935, 1). We pointed out earlier that Frege's categorical distinctions and Russell's theory of types were intended as an answer to the same question. However, for reasons we cannot go into here, Russell's theory of types was found dissatisfactory, and many logicians in the group adopted instead the theory of semantic categories expounded by Leśniewski in *Grundzüge eines neuen Systems der Grundlagen der Matematik* (1929). Leśniewski made a distinction between language and metalanguage, which was later explored by Tarski. He was also the first to point out that every language which contains its own semantics cannot obey the laws of classical logic, and if those laws are to be preserved, one has to reconstruct the language through hierarchical levels, where each level is interpreted in the next one.

2.4. Lesńiewski and Ajdukiewicz: The Grammar of Semantic Categories

Lesńiewski's theory of semantic categories, which he formulated around 1922, was deeply influenced by Husserl's theory of meaning categories and by Russell and Whitehead's theory of types. For Lesńiewski, too, any expression, understood as a finite sequence of inscriptions, belongs exactly to one semantic category. Lesńiewski himself did not have an explicit classification of categories into kinds, like Russell before him, but such a classification was built up inside his system later by Ajdukiewicz (Ajdukiewicz 1935). According to it, there are *basic* categories and *functor* categories, which reminds one of the Fregean distinction between saturated and unsaturated expressions. Moreover, like in Frege's hierarchy of levels, and in Russell's simplified type theory, one finds only two basic categories in Lesńiewski's system: sentences and names. All the other categories are functor categories.

Lesńiewski's system forms a ramified ascending hierarchy of functor categories which are characterized in two ways: by the number and the semantic categories of the arguments and by the semantic category of the whole expression formed by the functor together with its arguments. Lesńiewski's theory remained largely unknown outside Poland until 1935, when Ajdukiewicz gave it a more elegant formulation. It was intended to be applied to formal (constructed) rather than natural languages. Although Ajdukiewicz was more sensitive, at least in principle, to the latter, when he constructed his logical system, like Lesńiewski, he limited his attention only to languages having two basic semantic categories: singular names (names of individuals) and general names (names of universals).

Ajdukiewicz added to Lesńiewski's system an indexicalization of the semantic categories. To the basic categories of names and sentences he assigned the indices "n" and "s," respectively. To the functor categories he assigned a fractionary index consisting of a numerator and a denumerator. The former is the index of the semantic category of the value of the functor for its arguments. The latter is a sequence consisting of the indices of the semantic categories of the arguments.

Ajdukiewicz's categories are few in number and selected so that they fit the language of mathematics. He notices that the number of categories in ordinary language is much bigger, and there one has a fluctuation in meaning that renders the design of the system much more difficult. However, he points out that "In simple and favorable cases, however, the index apparatus cited above will be quite suitable for linguistic usage" (1935, 211).

We are now in a position to return to the initial question: What are the necessary and sufficient conditions for an expression to have unitary meaning? The necessary condition is for the expression to be *articulated throughout* (1935, 213). This means, first, that the expression may be divided into a *main functor* and its arguments. Ajdukiewicz is well aware that in ordinary language the order of the arguments in the main functor is not the same as its sequential ordering. Second, one has to check that each argument is also analyzable into a main functor and its arguments, and so on. Again, he points out that "ordinary language often admits elliptical expressions so that sometimes a significant composite expression cannot be well articulated throughout on the sole basis of the words explicitly contained in it. But a good overall articulation can be easily established by introducing the words omitted but implicit" (ibid., 213). The sufficient condition is that, after the division into functors and arguments, there must be a perfect fit between the number of arguments required by each functor and its actual arguments, which in addition must belong to the appropriate categories. An expression that fulfills both the necessary and sufficient condition has a unitary meaning, or, as Ajdukiewicz calls it, is *syntactically connected*. The matching of the functors' arguments with the semantic categories of the functors is checked mechanically by an algorithm that we now describe by way of an example. The sufficient condition is met if the result of this procedure is a simple index.

Ajdukiewicz gives the following simple sentence of mathematics (using parentheses instead of dots), where we write below each of its symbols the index of its category:

$$\begin{array}{ccccc} (p & \vee & p) & \rightarrow & p. \\ s & \frac{s}{ss} & s & \frac{s}{ss} & s \end{array}$$

We then arrange the parts of the expression into a main functor and its arguments:

$$\begin{array}{ccccc} \rightarrow, & p & \vee & p, & p. \\ \frac{s}{ss} & s & \frac{s}{ss} & s & s \end{array}$$

We apply the same procedure to any subexpression that can still be decomposed into a main functor and its arguments:

$$\begin{array}{ccccc} \rightarrow, & \vee, & p & p, & p. \\ \frac{s}{ss} & \frac{s}{ss} & s & s & s \end{array}$$

We next detach the sequence of indices of the expression:

$$\frac{s}{ss} \ \frac{s}{ss} \ s \ s \ s.$$

In the sequence thus obtained, we try, starting from left to right, to find a combination of indices so that we have a fractional index followed immediately by a sequence of indices that occur in the denominator of the fractional index. We cancel the sequence (if there are several, we cancel the first one), and replace it by the numerator of the fractional index. In our particular example, the combination we are looking for consists of the second, third, and fourth members of the sequence. The result is:

$$\frac{s}{ss} \ s.$$

We apply the same operation once more, and we get s.

This last index is the *exponent* of the expression. Because it is simple (and not fractionary), and all the others conditions have been fulfilled, our initial sentence is syntactically connected.

2.5. The Categorial Grammar of Bar-Hillel

Ajdukiewicz's theory was considerably developed by Bar-Hillel in a series of papers in the 1950's and 1960's (Bar-Hillel 1964, 1970). He shaped the concept of categorial grammar and popularized it to the English speaking world. As his predecessors, Bar-Hillel was interested in the question of the "unitary meaning" of a string of words. This problem was perceived even more acute in the 1950s, a period that sees the rise of computers and addresses the question of the feasibility of translation. As contrasted to his predecessors, Bar-Hillel was much more interested in the application of the tools of logic to ordinary language. He very much deplored the attitude of his teacher Rudolf Carnap, who, on one hand, developed very sophisticated mathematical tools to be applied to the study of language in general in *The Logical Syntax of Language* (Carnap 1937), but on the other hand found natural language too complicated to be studied with these tools. Carnap's attitude resumed in his *Introduction*,

> In consequence of the unsystematic and logically imperfect structure of the natural word-languages (such as German or Latin), the statement of their formal rules of formation and transformation would be so complicated that it would hardly be feasible in practise,

was, as we saw, symptomatic for most of the logicians working on the foundations of language (including Frege and the Polish school) and was regrettable for at least one reason. Carnap's work was what many linguists (including Zellig Harris and Noam Chomsky) read when they wanted to get acquainted with what logicians said about language. Comments like the ones just quoted would have and did eventually discourage them from seeing the relevance of some of the tools developed by logicians for solving problems in their own field. The Carnapian distinction anticipated by Husserl between formation and transformation rules would have been, as Bar-Hillel pointed out, highly relevant for studying the relation between, say, active and passive constructions in natural language undertaken much later by the generativists, especially if we recall that in Carnap's system both of them were formulated in *syntactic* terms.

One of Bar-Hillel's most important insights was that the theory of semantic categories as developed by Leśniewski and Ajdukiewicz was too rudimentary to be applied to the syntax of an ordinary language. For that purpose he improved Ajdukiewicz's theory in several directions. He noticed that Ajdukiewicz's notation α/β (this is the way he rewrote Ajdukiewicz's fractional index) for the functor categories makes it explicit that the functor is intended to apply only to an argument which occurs to its right. This was very clearly seen in the preceding section where Ajdukiewicz had only a right cancellation rule, which can be explicitly formulated as the following.

C1: Replace a string of two category symbols α/β, β by α. In symbols: $\alpha/\beta,\ \beta \to \alpha$.

Bar-Hillel pointed out that this kind of rule makes the Leśniewski–Ajdukiewicz theory applicable only to formal languages that have explicitly that sort of structure, like the formal languages expressed in the parentheses-free Polish notation. In these languages, one has expressions like "$\cdot + abc$" (i.e., "$(a+b)\cdot c$" in the notation which uses parentheses) and "$+a \cdot bc$" (i.e., "$a + (b \cdot c)$"). But that system does not apply to natural language like English. For instance, in a very simple English sentence like *John died* the natural order is that in which the nominal *John* precedes the functor expression *died*: n, s/n.

But then the cancellation rule C1 is not applicable to it. The system would work only if we rewrite the above sentence as *Died John.* So one of the shortcomings of the Leśniewski–Ajdukiewicz grammar was the unidirectionality of its semantic categories, that is, the functor had to appear only on the left of the argument.

Bar-Hillel overcame this limitation by adding a new kind of functor category of the form $\alpha\backslash\beta$ where the functor operates now on arguments to its left. The new categories will now be more sensitive to the natural language syntax. These are the main categories used by Bar-Hillel (1964, 76):

Basic categories

Nominals: n

Sentences: s

Functor categories

Intransitive verbals: $n\backslash s$

Adjectivals: n/n

Intransitive verbal adverbials: $(n\backslash s)\backslash(n\backslash s)$

Binary operators: $s\backslash s/s$

And so on.

Corresponding to this, he also introduced a left-cancellation rule.

C2: Replace a string of two category symbols α, α/β by β. In symbols: $\alpha,\ \alpha/\beta \to \beta$.

Another limitation in the Leśniewski–Ajdukiewicz grammar was the fact that to each expression there was assigned only one category. Consequently, each sentence had only one structural derivation. Such a limitation may be justified for artificial languages. But as pointed out by Carnap in his *Logical Syntax of Language*, in more complex languages one expression may belong to more categories (homonymy), and an expression may be ambiguous, that is, it may have more than one derivation. Accordingly, another improvement made by Bar-Hillel was to have expressions belonging to more than one category.

Applying these rules to a natural language sentence yields a derivation of that sentence. However, given that each expression may belong to several categories, its set of derivations may be rather large. An expression is well formed if it has at least one correct derivation. For example, consider the sentence:

(4) Little John slept soundly.

The dictionary will give us first the categories to which every word belongs. Thus we shall have *Little* (n/n), *John* (n), *slept* ($n\backslash s$), and *soundly* ($(n\backslash s)\backslash(n\backslash s)$). The next stage is to resolve the constituent structure of the sentence by the same mechanical procedure that Ajdukiewicz had used. The only difference is that now there are two cancellation rules that may be applied to a string of indices (Bar-Hillel 1964, 77). Let us illustrate how this procedure works in the case of (4).

We start with the sequence of indices of the subexpressions of (4):

(5) $$n/n, \quad n, \quad n\backslash s, \quad (n\backslash s)\backslash(n\backslash s).$$

We notice that there are three different ways to perform a cancellation, each of them resulting in one of the following sequences:

(6) a. $n, \quad n\backslash s, \quad (n\backslash s)\backslash(n\backslash s).$

b. $n/n, \quad s, \quad (n\backslash s)\backslash(n\backslash s).$

c. $n/n, \quad n, \quad n\backslash s.$

The sequence (6b) cannot be continued. The sequence (6a) can be continued by applying a cancellation rule to the first two members, after which we are in a blind alley, or by applying a cancellation rule to the second and the third member, the result being

(7) $$n, \quad n\backslash s.$$

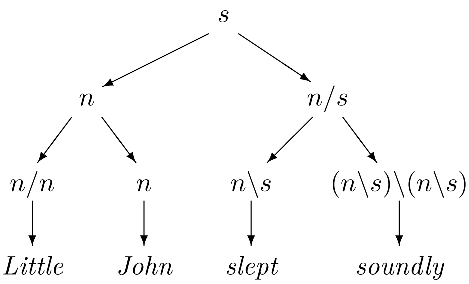
By applying a cancellation rule to this sequence, we reach the exponent s. The sequence (6c) can be continued, by applying a cancellation rule to its second and third members, after which we are in a blind alley, or by applying a cancellation rule to its first and second members, the result being

(8) $$n, \quad n\backslash s.$$

Finally, applying a cancellation rule once more, we reach the exponent s. Let us write down the two "successful" derivations.

1. $n/n,\ n,\ n\backslash s,\ (n\backslash s)\backslash(n\backslash s)$	1. $n/n,\ n,\ n\backslash s,\ (n\backslash s)\backslash(n\backslash s)$
2. $n,\ n\backslash s,\ (n\backslash s)\backslash(n\backslash s)$	2. $n/n,\ n,\ n\backslash s$
3. $n,\ n\backslash s$	3. $n,\ n\backslash s$
4. s	4. s

These two derivations differ from each other just in the fact that the cancellation step that occurs on the left side at stage two occurs on the right side at stage one. They are therefore equivalent, as can be seen from the fact that they give rise to the same tree expansion:



Things do not work so well, however, for more complex sentences like

(9) Paul thought that John slept soundly.

In this case, we get two derivations that are not equivalent. We are not going to exhibit the rather detailed analysis here (for the details, see Bar-Hillel 1964, 78–79). The important thing is that there are reasons to regard one of the two resulting derivation trees as unacceptable. So either the categorization is ill-chosen or the whole model is inadequate.

Problems are increased when we remember that for Bar-Hillel an expression may belong to several categories. For instance, *that* is sometimes a nominal (n) and sometimes an adjectival (n/n). *Thought* belongs to the categories n, $n\backslash s$, and $(n\backslash s)/(s/s)$ (*Paul thought John was asleep*). Thus the list of the category entries that the dictionary provides for some words may be rather long. In this case, the grammaticalness of some of the resulting derivations is highly dubious. In addition, the computational complexity of the process of constructing all the possible derivations is very high. The feasibility of the model decreases even more if we remember that the number of fundamental categories was very small. If we go on and add singular and plural, animate and inanimate, and so on, then the complexity becomes much bigger. Considerations of this sort made Chomsky (*Syntactic Structures*) very skeptical about the Bar-Hillel model. Actually Bar-Hillel himself came to the same conclusion (see section 4.1). A better linguistic model, according to him, is the transformational model of Harris (1957), and Chomsky (1957).

3. Truth-Conditional Semantics

The development of truth-conditional semantics starting in the works of Frege, Russell, Wittgenstein, and reaching the founding fathers of the field, Carnap and Tarski, is detailed in chapter 13. Here we only resume its main conclusions.

For the Polish logician Alfred Tarski, a semantic theory took the form of a theory of truth for a given language. The essential features of such a theory are laid down in his seminal paper, *The Concept of Truth in Formalized Languages* (1956). In designing a theory of truth for a given language, Tarski wanted to apply the methodology of semantic categories developed by his teacher, Lesńiewski (see previous discussion). This methodology is nothing else than the compositional method that requires that the syntax (grammar) of the language be specified in terms of explicit rules that dictate how expressions of appropriate categories combine themselves to form more complex expressions and finally sentences. The semantics, in this case the truth of a sentence, is then given by a set of semantical rules that mirror the appropriate rules of the syntax. But Tarski did not see any hope in his time to have a precise formulation of natural language ("colloquial language," as he called it) syntax, which had still to wait for the generative turn. In addition, he was fully aware that being semantically closed, natural language is beset by semantic paradoxes, and so he explicitly gave up the task of formulating a theory of truth for a natural language fragment. He believed that only a theory of truth for formalized languages is scientifically attainable. For Tarski, a formalized language is an interpreted one, like the language of arithmetic, and it can be given a precise syntactic representation.

For a formal language *L*, Tarski defined in the metalanguage *ML* (set theory) the predicate *truth-in-L*. For such a definition to be possible, for each expression of the object-language *L* there has to be an expression in *ML* which has the same meaning or translates it. The definition of *truth-in-L* is defined via the notion of satisfaction by induction on the complexity of formulas of the object-language *L*, as shown in chapter 9. The important thing to emphasize is that Tarski's theory presupposes the notion of translation or meaning.

In the late 1960s, Donald Davidson phrased Tarski's definition of the truth-predicate for a language *L* as an empirical theory, that is, a theory like any other in empirical sciences, with theoretical terms and axiomatic laws, from which logical consequences are to be derived which are then empirically tested. The purpose of such a theory is to give an answer to the question "What do we know that enables us to interpret the words of others?" (Davidson, *Radical Interpretation*, 125). Frege gave an analysis of the meaning of sentences, and Tarski a semantic analysis of the concept of truth for a formalized language. None of them was much interested in relating this semantic analysis to the actual behavior of the language users. Davidson is looking for much more: a theory that shows what one knows when one understands a language. The switch is clearer toward the active use of language and its interpretation. The details of Davidson's theory are described in chapter 13. Here is enough to point out that Davidson merged, in an ingenious way, an empirical setting that he had inherited from his teacher Quine with the Tarskian theory of truth: The language to be investigated is identified with Tarski's object language, the language of the investigating linguist with Tarski's metalanguage, and the correlation of the sentences of the former with those of the latter plays the role of the T-sentences. The major difference with Tarski is that the truth-predicate is now a primitive notion, not one to be defined. Tarski's requirement that for each expression in the object-language there be an expression in the metalanguage which is its translation, or has the same meaning, is not any longer the starting point of the inquiry but its outcome. The result of Davidson's undertaking under his reinterpretation of Tarski's theory of truth is broader than Quine's: not only a translation of the sentences of the target language into the sentences of the home language but also a systematic procedure that shows how the translation (meanings) of the former depends on their structure. This extra payoff was possible to achieve thanks to the Tarskian compositional definition of the notion of satisfaction.

Davidson wants his theory of truth to be a theory of meaning for a natural language, or a fragment of it. In laying the bases of his program for semantics, Davidson criticizes linguists and philosophers for having "exaggerated the difficulties in the way of giving a formal theory of natural language" (Davidson 1984, 55). In particular, what Davidson mostly refuses is the common conclusion that "there are two kinds of language, natural and artificial." This alleged difference would, in fact, correspond to the existence of some inherent features in natural language, which would act as insurmountable barriers forbidding a formal definition of truth in the same rigorous terms as the one provided by Tarski (1936) for logical languages. Not only is the existence of such barriers totally unwarranted, but the effort of pursuing a formal semantics theory of language is worthwhile because "in so far as we succeed in giving such a theory . . . , we see natural language as a formal system; and . . . we can think of linguists and analytic philosophers as co- workers" (ibid.).

The Davidsonian connection between truth and meaning of the kind Tarski has shown us how to construct, and which, as we have seen, finds its roots in Frege's work, has left a long-standing legacy in the interplay between logic and language inquiry in the last century. It is based on the compositional method of defining semantics on a rule-by-rule basis in tandem with a recursively defined syntax. It is, as we are going to see, the legacy embraced by Montague when he embarked on his project on "English as a formal language." It is, according to Montague, precisely this concern with truth-functional semantics that radically separates his project from the language paradigm emerging from the generative school, to which we now turn. However, at the end of our study, we are going to signal some interesting endeavors of combining the two (e.g., Higginbotham 1985, 1986).

4. Chomsky's Revolution in Linguistics

Prima facie the claim that logic and linguistics in the twentieth century would not have met without the development of generative grammar might easily be taken as an overstatement. Nonetheless, a closer analysis of what went on in the study of language in the late 1950s reveals that Chomsky's generative enterprise has actually been the precondition for any real and effective dialogue between new twentieth-century linguistics on the one hand and logic and philosophy on the other. This emerges with particular evidence if we take into account the character and methodological assumptions of pregenerative linguistics. Therefore, if among linguists it is still a matter of debate whether it is appropriate to regard generative grammar as a real revolution with respect to the past, from the point of view of logic and philosophy, using this term is not an overestimation. The breakthrough of *Syntactic Structures* (Chomsky 1957) goes far beyond the contribution to formal grammar provided by the theory of transformations. The rise of generative grammar actually represents the first appearance of contemporary linguistics into the debate on the universal structure of language. Actually, the Chomskian turn has ultimately resulted in an altogether new way of looking at language, different with respect to both traditional linguistics and logicophilosophical grammar.

4.1. The Birth of Transformational Generative Grammar

Linguistics experienced the first major change from the nineteenth-century tradition with the development of *structuralism*. Although this approach to the study of language stemmed more or less directly from the *Cours de linguistique générale* of Ferdinand de Saussure (1916), what is usually termed as structural linguistics should actually better be described as a family of linguistics schools, which, notwithstanding a common methodological overlapping, greatly differed in their conception of linguistic inquiry. What is mostly typical of structural linguistics is (i) the Saussurian distinction between *langue* and *parole*; (ii) the clear separation of the diachronic approach to language from the synchronic one, and the legitimation of the latter as an autonomous field of inquiry; and finally (iii) a conception of language as a structural system of signs—intended as arbitrary relations between a form (*significant*) and a content (*signifié*)—whose elements receive a value from their position in the system and from the reciprocal interrelations with the other parts of the system. Thus, structural linguistics essentially established itself as a taxonomic and paradigmatic inquiry, mostly consisting in the analysis of elements and the structure of the system of a specific language. Besides, a crucial feature of the structural approach to language in the first half of the century was the central role occupied by phonology and morphology with respect to syntax.[1]

Structural linguistics developed into various schools on the two sides of the Atlantic, but no real and significant relationship with logic blossomed in either cases. On the one hand, the Prague school of Troubetzkoy and Jacobson, and the Copenhagen circle of Hjelmslev, the leading representatives of European structuralism, mainly focused on developing respectively the Saussurrian notion of phoneme and its theory of linguistic signs. Moreover, they firmly regarded the study of language as part of human sciences, totally beyond the domain of any naturalistic, scientific, or formal inquiry. On the other hand, American structuralism with Edward Sapir and Leonard Bloomfield as its main representatives, was deeply oriented toward an anthropological study of language. Besides, Bloomfield's *Language* (1933) firmly established linguistics as an irreducibly empiricist scientific enterprise (thus differing from the mentalist approach of Sapir). By 1950s, the so-called post-Bloomfieldian school outnumbered the followers of Sapir's structuralism, so that the pre-Chomskian American linguistic environment was totally dominated by the empiricist paradigm. In contrast to European linguistics, Bloomfieldian structuralism tried to provide linguistics with the same scientific status as natural sciences. It was the particular conception of natural science that characterized the post-Bloomfieldian school that divorced the goals of empirical linguistics from the goals and the tradition of logical and philosophical grammar.

In fact, being deeply influenced by the empiricist philosophy of science of the Vienna circle,[2] Bloomfield himself and his followers regarded linguistics as an empirical science, to be studied with a strictly inductive and physicalist method. The purpose of linguistic inquiry was to "discover" the grammar of a particular language that emerges out of the stream of physical sounds produced by its speakers. Every abstract construct or generalization that could not be traced back to an empirical observation was to be ignored. Linguistics was thus merely *descriptive* and *taxonomical*, and linguistic investigation was strictly intended as the description of a *given* language and not of language *qua* language. Therefore, metatheory was limited to the formulation of a series of prescriptive rules that had to guide the discovery procedure of the grammarian. Linguistic inquiry was to be based on a merely external corpus of data, consisting mainly of physical records of the flow of speech. As a consequence, the judgments of the speakers were completely disregarded because of their alleged scientific unreliability. The analysis of a given language consisted in the discovery of four ordered levels of grammatical descriptions: phonemics, morphemics, syntax, and discourse. Similarly to European structural linguists, post-Bloomfieldian analyses almost exclusively concentrated on the first two levels. The inductive method of American structuralism, as well as its focusing on phonological and morphological descriptions of particular languages, also reveal the complete lack of any interest in the combinatorial nature of language and its syntactic creativity. The acknowledgment of the fundamental fact that language allows speakers to express an indefinite number of thoughts by combining finite resources—an issue already regarded as crucial by von Humboldt and Frege in the third of his *Logische Untersuchungen*[3] and which plays such a central role in logical grammar—was completely lacking in Bloomfieldian linguistics, as was the awareness of the recursive mechanisms of natural language. Consequently, there was no real interest in addressing the question of the general laws of syntactic combination of linguistic expressions.

This empiricist orientation also affected the way American structuralism approached the question of meaning. Semantics was excluded from the domain of scientific explanation in linguistics, and analyses grounded on semantic considerations were firmly denied. This represents another major difference between pregenerative linguistics and the work in the logical grammar tradition: From the point of view of Bloomfieldian methodology, any role assigned to semantic categories in deriving language composition would represent a dangerous and misleading confusion between formal and semantic considerations.

To summarize, by the middle of the 1950s, logic and the new science of linguistics lived in a "splendid mutual isolation" (Bar-Hillel 1969, 182). Harris deplored this attitude in a representative passage: "Whereas the logicians have avoided the analysis of existing languages, linguists study them" (Harris, 1951, 16 n17).

This situation dramatically changed with the publication of *Syntactic Structures* in 1957, which put the issue of the combinatorial nature of language and linguistic creativity at the core of linguistic inquiry, thereby laying the preconditions for the relevance of the work done in mathematical logic for linguistic theorizing. With a revolutionary move, Chomsky rejected the taxonomic and descriptive approach of post-Bloomfieldian linguistics and claimed "the linguist's task to be that of producing a device of some sort (called a grammar) for generating all and only the sentences of a language" (Chomsky 1957, 85). As a direct consequence of this methodological innovation, the domain of linguistics was enlarged from the description of a specific language to the theory of language structure itself. In Chomsky (1957, 50) a *condition of generality* is stated, according to which,

> we require that the grammar of a given language be constructed in accord with a specific theory of linguistic structure in which such terms as "phoneme" and "phrase" are defined independently of any particular language.

The overall goal of linguistics is thus the quest for the *universal principles* that make up the possibility of human language. This also led to a radical modification of the adequacy conditions for linguistic descriptions. In the Bloomfieldian tradition, a language description was adequate only in so far as it respected the methodological prescriptions that guaranteed its empirical and totally inductive nature. This view is, instead, firmly rejected by Chomsky (1957, 106): "The theory of linguistic structure must be distinguished clearly from a manual of helpful procedures for the discovery of grammars." In the Chomskian framework, a language description has to pass different levels of adequacy, the observational one being just the first. The top level of the adequacy conditions of a grammar, is what Chomsky calls its *explanatory adequacy*:[4]

> A linguistic theory that aims for explanatory adequacy is concerned with the internal structure of the device [*the generative grammar*]; that is, it aims to provide a principled basis, independent of any particular language, for the selection of the descriptively adequate grammar of each language. (Chomsky 1964, 63)

Linguistic description does not have to face only the tribunal of data, but also and more crucially the higher tribunal of the explanation of the general principles which lay at the basis of human creativity. Some of the direct consequences of this shift of perspective are the acceptance of speakers' grammaticality judgments as the fundamental empirical evidence, the introduction of abstract levels of representation of linguistic information, and the centrality of the investigation of the formal structure of grammars and linguistic rules.

The introduction—or rather the reintroduction—of the theme of linguistic creativity led to a direct connection between generative grammar and the rationalist, Cartesian tradition.[5] The rationalist turn not only switched the attention of linguists to the combinatorial nature of language, it also gave meaning a new and important role in linguistics. Linguistic creativity is, in fact, defined as the ability of understanding and generating an indefinite number of linguistic expressions, and the combinatorial power of language, identified with the generative nature of the syntactic component, is subservient to the goal of pairing a potentially infinite sound and meaning patterns.

Besides the breakthrough in linguistics, *Syntactic Structures* significantly contributed to the debate on the form of natural language grammar by showing the inadequacy of phrase structure grammars, and by introducing the first version of the *transformational generative* model. Chomsky's argument is focused on the claim that a theory of language structure based on phrase structure grammars "will be extremely complex, *ad hoc*, and 'unrevealing' " (Chomsky 1957, 34). Some of the most interesting examples brought by Chomsky to illustrate these faults concern the analysis of the auxiliary system in English and of the relations between active and passive sentences. The generation of all the possible combinations of auxiliaries verbs (e.g., *has taken*, *is taking*, *has been taking*, *is being taken*, etc.)—and the exclusion of all the impossible ones, represented an incredibly hard task for the phrase structure grammars available at the time. Much of the difficulty is due to the co-occurrence relation between the auxiliary and the morphological affix:

(10) a. *have -en* (perfect tenses)[6]
b. *be -en* (passive)
d. *be -ing* (progressive form)

Chomsky claims that while phrase structure grammars pay a very high price to capture such relations, the auxiliary distribution can easily be accounted for by assuming phrase rules that generate the discontinuous elements—auxiliary plus affix—as unit constituents, and then by positing a transformation rule that permute affix and verb to their surface position. This argument based on the simplification of the theory obtained by augmenting phrase structures with transformations is also applied to the analysis of passive sentences. In fact, formulating a proper rule that generates passive sentences requires taking into account a whole series of restrictions, such as the type of the auxiliary, the (in)transitivity of the verb, the type of the subject and of the object, and so on, which were extremely complex to express in terms of the existing context-free phrase structure formalisms. More generally, phrase structure grammars are judged to be inadequate to account for the context-sensitivity aspects of natural language, and last but not least, they are incapable to capture the fact that the application of a particular rule may require to look back to past stages of the derivation. For instance, to correctly describe the phenomenon of subject-verb agreement and thus generate the grammatically correct *The man runs*, while excluding the ungrammatical **The man run*, the rule that decides which affix to add to the verb must necessary look back to the stage of the derivation that had generated the subject noun phrase, and check whether it is singular or not.

The theory of grammatical description proposed by Chomsky to overcome the difficulties of phrase structure grammars is an abstract system of representations, made of three different layers: the *phrase structure component*, the *transformational component*, and the *morphophonemic component*. The phrase structure component is composed of phrase structure rules[7] that generate abstract sequences of constituents. The transformational rules, then, apply to some of these strings and convert them into other abstract strings with a different analysis of constituents. For instance, the following is the definition of passive transformation (Chomsky 1957, 43):

(11) If S_1 is a grammatical sentence of the form

$$NP_1 - Aux - V - NP_2,$$

then the corresponding string of the form

$$NP_2 - Aux + be + en - V - by + NP_1$$

is also a grammatical sentence.

The derivation of a sentence like *The book was taken by John* goes as follows:

(12) a. *John* – past – *take* – *the book.*
b. *The book* – past + *be* + *en* – *take* – *by* + *John.*
c. *The book* – *be* + past – *take* + *en* – *by* + *John.*
d. *The book was taken by John.*

The phrase structure rules generate the first level of representation formed by the sequence of phrase markers in (12a); then the passive transformation applies, yielding (12b), where the passive auxiliary + affix complex is inserted. The application of the auxiliary transformation distributes the participle affix to the main verb (12c). Finally, the morphophonemic rules apply to (12c), producing the sentence in (12d).

The introduction of derivations containing abstract terms, whose form and order can be quite far from the final output, and the important innovation of transformations that permute the elements during the derivation represented a radical departure from phrase structure grammars. Bar-Hillel et al. (1960) proved the equivalency with respect to their generative capacity between categorial grammars and context-free phrase structure grammars. In this way, the attack brought by Chomsky against the latter automatically extends to categorial grammars, whose inadequacy as a formal theory for natural language was ultimately recognized by Bar-Hillel himself:

> As a matter of fact, I had already noticed six years ago that the model did not work too well for complex sentences, but had rather hoped that this was due only to lack of refinement that could be partially remedied by increasing the number of fundamental categories, partly by additional rules. I have now come to realize that its failure in the more complex cases has a much deeper cause: the linguistic model on which this model was based is just not good enough. (Bar-Hillel 1964, 83)

It is important to remark that a large part of the innovative power of the arguments for transformations in *Syntactic Structures* also lies in the relevance assigned to certain types of syntactic phenomena for the evaluation of formal theories of grammar. Syntactic dependencies between discontinuous elements and constructions involving displaced or "moved" elements (e.g., passive and interrogative sentences) became core facts of natural language, and their characterization and proper treatment began to be regarded as necessary conditions for any theory aiming to provide an adequate description and explanation of the universal principles of language structure.

In the years that followed the publication of *Syntactic Structures*, the new generative paradigm rapidly conquered many spaces once dominated by Bloomfieldian linguistics. At the same time, the first transformation model underwent important modifications, especially thanks to an intense work directed to make explicit the exact nature of transformations and their classification. Finally, a more stabilized version of the transformational generative syntax—the so-called standard theory—emerged in 1965 with Chomsky's *Aspects of a Theory of Syntax*. The standard theory includes a syntactic component made of two abstract layers of representation: *deep structure* and *surface structure*. The level of deep structure is generated by the application of three sets of rules (base rules): *phrase structure rules*, *subcategorization rules*, and *lexical insertion rules*, the latter taking lexical items from the lexicon and inserting them into the phrase structure tree. In contrast to the 1957 model, where the recursive capacity of language was provided by generalized transformations—a particular type of transformations that take representations generated by phrase structure rules and embed one into the other (e.g., to form relative clauses or complex sentences)—in the *Standard Theory* the recursive power lies in the base rules, so that each derivation produces a single formal object that then enters the transformational component. Transformations produce the surface structure, which is then given as input to the phonological rules to derive the phonetic representation. We will see in section 5 how in the late 1960s, the standard theory became the starting point of an intense debate involving linguists, logicians, and philosophers on the relation between syntax and semantics, as well as concerning the nature of the deep structure representations.

4.2. The Autonomy of Syntax

With his critique of phrase structure grammars, Chomsky showed how linguistic theory in the form of a transformational generative theory can make its contribution to the analysis of the conditions of the possibility of human language. In that, he went much beyond the generative and explanatory resources reached by categorial grammars. In the same time, Chomsky's proposal represented a radical departure from the main tenets underlying the study of language universals in the logical grammar tradition by severing the link that related the analysis of syntactic structure to the semantic categories of grammatical terms. Almost at the end of *Syntactic Structures*, summing up the general view presented and defended throughout his book, Chomsky claims that: "Grammar is best formulated as a self-contained study independent of semantics. In particular, the notion of grammaticalness cannot be identified with meaningfulness" (Chomsky 1957, 106). Chomsky states here the well-known *principle of the autonomy of syntax*, which stands in deep contrast both to Husserl's idea that the principles of syntactic connection can be explained in terms of meaning connection rules (see section 2.2), and to Ajdukiewicz's principle according to which the well-formedness conditions of linguistic expressions depend on the "specification of the conditions under which a word-pattern, constituted of meaningful words, forms an expression that itself has a unitary meaning" (Ajdukiewicz, see section 2.3). Chomsky attacks these views by criticizing the equivalence of meaningfulness with grammaticalness. The notorious example (3) is intended to show that sentences without a unitary meaning can well be judged to be grammatical:

(13) Colorless green ideas sleep furiously.

And here are examples of ungrammatical sentences having a unitary meaning:[8]

(14) a. Have you a book on modern music?
b. *Read you a book on semantic music?
c. The book seems interesting.
d. *The child seems sleeping.

Chomsky argues that there is no semantic reason to prefer (14a) to (14b) and (14c) to (14d), besides the fact that there is surely a sense in which even the ungrammatical sentences have a unitary meaning exactly as their grammatical equivalents. The consequence was that only the independence from semantic considerations was seen to grant a reliable foundation for the search for the formal structure of language. It is, however, of crucial importance that independence neither means nor entails irrelevance. In other words, the autonomy of syntax is not a way to declare the structure and nature of the semantic component irrelevant with respect to the tasks of linguistic theory or to the architecture of the grammar of a language. On the contrary, the principle is regarded by Chomsky as the precondition of any attempt to tackle this issue seriously. There is no doubt that syntactic descriptions feed the semantic component and guide the composition of the interpretation of complex expressions:

> much of our discussion can be understood as suggesting a reformulation of parts of the theory of meaning that deals with so-called "structural meaning" in terms of the completely nonsemantic theory of grammatical structure. (Chomsky 1957, 103 n10)

Nonetheless, it is essential that structural descriptions must be defined independently of meaning and of general semantic considerations, that is, in purely formal terms, to be the formal, effective machinery for the derivation of semantic compositions.

The main aim of Chomsky's principle is therefore to provide a definition of syntax which is *nondependent* on semantics. Chomsky is actually always ready to accept the fact that semantics is one of the main guides for linguists' analysis. However, once a given analysis is suggested and enlightened by some semantic insight, it must ultimately be shaped in purely syntactic terms, that is, in terms of nonsemantic elements and rules. The revolutionary strength of the autonomy of syntax can hardly be questioned when we look at some of the examples in *Syntactic Structure*. For instance, Chomsky attacks the traditional idea that notions like grammatical subject or grammatical object have to be defined respectively in terms of the semantic notions of agent of an action and patient of an action. To this purpose, he brings striking counterexamples, like *John received a letter*, or *The fighting stopped*, where the grammatical subjects do not satisfy such semantic requirements. Chomsky's alternative proposal is to define subject and object in purely syntactic and formal terms, that is, by means of particular configurations of syntactic descriptions. Later on, notions like agent or patient also entered the theory as *thematic roles*,[9] and played a crucial part at the *Government and Binding* stage of Chomsky's theory.

As we pointed out in the section on Frege, there is a striking similarity between Chomsky's redefinition of grammatical subject, and what Frege says in his *Begriffsschrift* about the same notion. In the same way as Frege wanted to set logic free of its old grammatical connotations, thereby dismissing the logical relevance of problematic and jeopardized concepts like that of subject, in 1957 Chomsky wants to define the crucial grammatical notion of subject independently of any semantic considerations, for that would prevent one, according to him, from reaching a rigorous definition.

Syntactic rules drive semantic compositionality, but the critical point is whether this can be explained by a priori positing the uniformity between syntactic and semantic processes, and especially by defining the former in terms of the latter. Chomsky's view is rather that syntactic and semantic features match to the point of allowing language to do its job—that is, carrying complex thoughts by means of finite combinatorial resources—but this match is not complete. The truly empirical issue is therefore to evaluate the extension of the correspondence between syntax and semantics, or better the extension of the mismatch that still makes language possible. Chomsky's claim that the relation between form and meaning in language must be tackled by assuming syntactic representations whose form and conditions feed meanings but do not depend on them was actually repeatedly challenged in the late 1960s and 1970s. Stemming directly from within the generative enterprise, the generative semantics movement brought a strong attack on the thesis of the autonomy of syntax, by defining "deep syntax" as actually a logico-semantic level. Not very differently, Montague presented a new model of logical grammar according to which—thanks to a proper pairing of semantic and syntactic operations—sentences are analyzed in such a way as to exhibit the logical form directly on their syntactic sleeves.

4.3. Semantic Theory in Early Generative Grammar

The problem of the relations between grammatical description and their interpretation had a central role in generative linguistics since its very beginning. Linguistic creativity is, in fact, defined as the ability of understanding sentences, and the combinatorial power of language is subservient to the goal of pairing sound and meaning patterns:

> The grammar as a whole can thus be regarded, ultimately, as a device for pairing phonetically represented signals with semantic representations, this pairing being mediated through a system of abstract structures generated by the syntactic component. (Chomsky 1964, 52)

The standard theory is actually a model for the cognitive architecture of the language faculty, as composed of three independent modules, that is, *the syntactic component*, *the phonological component*, and *the semantic component*. Generative linguistics, thus, reestablished the centrality of language as a system of sound-meaning pairs, and thereby recovered De Saussure's main insight, which was lost in the empiricist version of American structuralism due to its incapability to integrate the role of meaning in a scientific theory of language in other than behaviorist stances. Chomsky's negative remarks on current theories of meaning that he blamed for using the term meaning as a "catch-all term to include every aspect of language that we know very little about" (Chomsky 1957, 104), was certainly directed against the behaviorist approaches to meaning framed in terms of stimulus-reaction patterns, as well as against any approach making use of intensions, sentential truth conditions, conditions for nondeviant utterances, distribution, and rules of use (Katz and Fodor 1963, 480). Thus from the very beginning, the approach to the study of meaning in generative grammar is characterized by a strenuous opposition against every externalist characterization of the semantic component, be it formulated in terms of behaviorist Skinnerian models,[10] or truth-conditional semantics so typical of the logical grammar tradition. In fact, according to Chomsky (1964, 77), "explanatory adequacy for descriptive semantics requires ... the development of an independent semantic theory ... that deals ... with the question: what are the substantive and formal constraints on systems of concepts that are constructed by humans on the basis of presented data?"

The burden of giving a shape to the semantic component within the standard theory was taken by J. J. Katz and J. Fodor with *The Structure of a Semantic Theory* (1963), which represents the first attempt to develop a semantic theory consistent with the generative approach to language:

> A semantic theory describes and explains the interpretative ability of speakers: by accounting for their performance in detecting the number and content of the readings of a sentence; by detecting semantic anomalies; by deciding upon paraphrase relations between sentences; and by marking every other semantic property or relation that plays a role in this ability. (Katz and Fodor 1963, 486)

Describing the interpretive ability of speakers requires tackling what Katz and Fodor call the *projection problem*, that is, determining the compositional procedure by which speakers are able to interpret an infinite number of linguistic expressions by combining a finite repository of meaningful expressions through an equally finite set of rules. The projection problem is now supposed to be solved by the descriptions associated with the syntactic component of the transformational generative grammar which is specified independently and is autonomous of the semantic module. Thus, according to Katz and Fodor (1963, 484), "linguistic description minus grammar equals semantics," and, as a consequence, the role of semantics is purely *interpretive*, that is, it *merely* provides an interpretation of the syntactic descriptions. This approach, a direct corollary to the principle of the autonomy of syntax, stands in deep contrast to the approach of categorial grammars, where the semantic categories of lexical items, with their basic distinction between basic and functor categories, are supposed to drive the "syntactic connexity" of the complex linguistic expression. In the architecture of the standard theory, the syntactic composition is guaranteed by the syntactic component, while the semantic module assigns meanings to the lexical items and then projects from them, up along the syntactic tree, the unitary meaning to be assigned to the complex expression.

The semantic theory (*KF*) proposed in Katz and Fodor (1963), further developed in Katz (1972), has two components, the *dictionary* and the *projection rules*. The former assigns to every lexical item an entry consisting of its grammatical category and a semantic part describing its possible *senses*. Each sense is in turn described by means of a list of *semantic markers* (e.g., [Human], [Male], etc.) and a *distinguisher*. Together these elements "decompose the meaning of a lexical item (or one sense) into its atomic concepts, thus enabling us to exhibit the semantic structure in a dictionary entry and the semantic relations among the various senses of a lexical item" (Katz and Fodor 1963, 496). Semantic markers are intended to express those aspects of a sense that are systematic with respect to the language and give rise to basic semantic oppositions among senses. In this way, their role becomes similar to that of features in phonology, from which they are actually inspired. On the other side, distinguishers represent the idiosyncratic and unsystematic aspects of the sense of a lexical item. Once the lexical items that appear as leaves in a syntactic tree are assigned a dictionary entry, the projection rules compose these entries along the paths of the syntactic tree. This projection does not take place according to the function-argument schema typical of the Fregean tradition, but through a process of unification of the corresponding clusters of semantic markers and distinguishers, thus producing all the possible readings to be associated to the complete sentence.[11]

The notion of projection rules actually represents a common element of both *KF* and of the semantic analysis of meaning to be proposed in *Montague grammar*: In both cases, semantic rules pair syntactic formation rules. On the other side, the conception of the lexicon in *KF* is completely different from the corresponding notion in logical semantics and logical grammar, and is surely one of the weakest and most attacked part of the theory. Where in categorical grammars the lexical organization takes place according to the function-argument distinction, the dictionary has a fairly standard lexicographic organization, as also remarked by Bar-Hillel (1970, 185), who criticizes the theory for its "identification of semantics with lexicology." As Katz (1972) claims, *KF* is actually intended to be a theory of sense and not a theory of reference, and truth and truth-conditions have thus no role to play in it. Semantic interpretation and meaning representation are rather achieved through a process of semantic decomposition, by assuming a fairly traditional "chemical view on concepts"[12]—typical of the rationalist analyses in the eighteenth century—according to which the sense of a lexical item is analyzed in terms of its basic conceptual bricks. However, the introduction of semantic markers as the technical device that performs the task of lexical decomposition and deals with concept combination does not help overcome the difficulties of the lexicographic definitions of senses pursued in *KF*. Besides, the usage of semantic markers was heavily criticized for its complete lack of real explanatory power as far as meaning is concerned, since, as Lewis (1972, 169–170) claims, "translation into Latin might serve as well."

Besides its own specific problems, *KF* has provoked one of the strongest disagreements between the new theory of grammar in generative linguistics and the logicophilosophical approach. For instance, Katz attacks the idea of logical form based on the distinction between form and content (Katz 1972, xvii), and claims that formal logic cannot provide a proper semantics for natural language because of its being exclusively concerned with the logical form of sentences, which in turn heavily depends on the contribution of logical words like connectives, quantifiers, and so on.[13] According to Katz, the truly semantic facts to be explained are those that philosophers called *analytic*, that is, truths and inferences based not on the structural properties of syncategorematic words (e.g., quantifiers and connectives) but rather on the lexical content of words, like (15), which should equally well deserve the status of logical truth:

(15) If x is a bachelor, then x is an unmarried adult man.

The importance and centrality assigned to the characterization of lexical meaning and lexical inference represents a genuine and positive contribution that *KF* brought to the semantic debate, independently of the problems of the formal representation of the semantic level offered by the theory. According to it, one of the main tasks of a semantic theory is to capture the contrast between, for instance, *The dog chases the cat* and *The cat chases the dog*, which differ semantically, despite having the same structural description. Thus *KF* shifts the focus from issues of purely compositional semantics—which on its view are basically solved by the proper formulation of a syntactic component paired with projection rules—to the issue of how to characterize lexical meaning. This stands in deep contrast to the conception of the logical grammar tradition, where semantic categories are carved out mainly to drive semantic composition, but fails to provide satisfactory insights into lexical content. The interest in the lexical aspects of meaning will constantly grow in the later stages of semantic inquiry in theoretical linguistics, and will largely influence the research in model-theoretic semantics, which will, in some cases, also incorporate and develop the formal treatment of lexical decomposition.[14]

Katz (1972) considers the failure of logical theory to deal with the core semantic aspects of language as another consequence of "the rise and eventual dominance of empiricism" (xxi) with its behavioristic perspective on meaning. Actually, *KF* should properly be regarded as another episode in the Chomskian program with its systematic opposition against empiricist approaches to language. *KF* attempts to develop a fully internalist, rationalist, and intensionalist analysis of meaning and semantic inference:

> Empiricists claim that concepts of the theory of meaning are unscientific, occult and useless, and should be banished from a scientific theory of language. … Thus, the constructive task for the rationalist approach to the study of language is to reply to these claims in the only way that can ultimately discredit them, that is by building a linguistic theory which demonstrates the scientific soundness of concepts such as sense, meaning, synonymy, analyticity, and so on. (Katz 1972, xxiii)

The heart of the polemics is the analytic-synthetic distinction and the notion of synonymy criticized by Quine, both related to his argument against any mentalist conception of meaning not reducible to purely behaviorist assumptions. When applied to nonlogical words, Quine finds the notion of analyticity based on that of synonymy as totally unreliable and inherently circular:

> But there is a second class of statements, typified by (2):
>
> (2) No bachelor is married
>
> The characteristic of such a statement is that it can be turned into a logical truth by putting synonyms for synonyms; thus (2) can be turned into [No unmarried man is married] by putting "unmarried man" for its synonym "bachelor." We still lack a proper characterization of this second class of analytic statements, and therewith of analyticity generally, inasmuch as we have had in the above description to lean on a notion of "synonymy" which is no less in need of clarification than analyticity itself. (Quine 1953a, 23)

In addition, Quine attacks the notion of synonymy as a pure relation between the senses of the words. Synonymy should instead be approached only "from the point of view of long segments of discourse" (Quine 1953b, 57; see also Quine 1960). Again, Quine's purpose is to substitute the notion of synonymy as meaning-sharing with the behavioristic notion of approximate likeness of the effects provoked by linguistic expressions on a hearer. With respect to these empiricist arguments, the theory of sense pursued in *KF* is, therefore, an attempt to develop an internalist theory of meaning, based on the conceptual analysis of word senses. The theoretical device of semantic markers is intended to provide a new foundation for the notion of word meaning and for the semantic relations of analyticity and synonymy *qua* relations between word meanings.

5. Deep Syntax and Generative Semantics

One of the most interesting aspects of the period from the late 1960s up to the mid-1970s is that philosophers and linguists found an unprecedented ground of agreement in carrying a strong attack against the model of grammar proposed by Chomsky in *Aspects* in 1965. This atmosphere is best illustrated by the volume *Semantics of Natural Language*, edited by D. Davidson and G. Harman in 1972, which contains contributions by generative linguists, logicians and philosophers, whose unifying leitmotif is the refusal of the purely syntactic nature assigned to deep structure representations in Chomsky's standard theory. The common claim is that deep structure is to be identified with logical form, that is to say, deep structure must be equated with the place in which the hidden logical structure of natural language is explicitly encoded. On the linguistic side, the convergence with logic was mainly carried out by J. McCawley, G. Lakoff, and other representatives of the *generative semantics* movement, the harsh opposition movement to the Chomskian theory of grammar, which quickly developed in 1968 to then rapidly decline around

1973–1974. The relevance of this heterodox movement stemming out of the body of Chomskian linguistics actually goes far beyond its short life and the particular linguistic solutions and analyses proposed by its representatives, most of which were soon to be dismissed. Crucially, starting from the assumption that "the linguists' and the logician's concerns are consistent with each other" (McCawley 1972, 540), generative semantics was able to raise a whole wealth of issues concerning the relation between grammatical structure and logical form, thus showing the inadequacies of the first transformational models to tackle these aspects of the theory of language. Many problems that attracted the attention of generative semanticists, like quantification, bound anaphora, and so on, continued then to occupy a key position in the later developments of the Chomskian framework, leading toward more elaborated hypotheses on the syntax-semantic interface.

As we saw in section 4, Chomsky's main claim is that syntactic representations must be designed strictly independently of semantics, which instead forms a separate module within the architecture of grammar. In the standard theory, deep structure occupies a particularly prominent role: It is, in fact, the level at which subcategorization and selectional restrictions are defined, grammatical relations are established, and lexical items are inserted from the lexicon. Moreover, deep structure also represents the main, actually the only interface with semantics. In fact, the relation between the syntactic component and the semantic module is regulated by the so-called *Katz–Postal hypothesis* (Katz and Postal 1964), according to which all the syntactic information necessary for the semantic interpretation is provided by deep structure. In other words, transformations are all meaning-preserving, since they do not affect the interpretation of syntactic structures. This hypothesis perfectly fits with the interpretive role assigned to the semantic component in *KF*, where the projection problem is intended to be solved at the syntactic level by the representations provided as input to the semantic rules, which have a strictly interpretive role. In fact, projection rules in *KF* are actually quite trivial, since they simply have to compose the semantic markers of the lexical items all up the syntactic tree.

One of the most crucial consequences of assuming the Katz–Postal hypothesis as the basis for the syntax-semantics pairing is that every nonlexical semantic ambiguity must be explained in terms of a difference at the level of deep structure. This is simply a corollary to the fact that all the structural information which determines semantic composition is already encoded in the deep structure, together with the fact that transformations cannot affect meaning. However, crucial problems arise with the analysis of sentences containing logical operators and quantifiers. For instance, (16a) is semantically ambiguous between a reading in which the negation has wide scope over the adverbial clause, and one in which it has narrow scope. Similarly, (16b) is ambiguous between a reading in which the universal quantifier has wide scope over the existential one, and a reading in which the universal quantifier has narrow scope:

(16) a. I don't steal from John because I like him.

b. Everyone loves someone.

The problem is that, given that deep structure is assumed to represent the only interface with semantics, such ambiguities can be accounted for only in terms of structural differences at this level of representation. To tackle this issue, Lakoff (1970) proposes to represent the ambiguity in (16a) at the level of deep structure by associating two readings with it:

(17) a. [$_S$ Neg [$_S$ I steal from John] [because I like John]].

b. [$_S$ [$_S$ Neg I steal from John] [because I like John]].

Similarly, Lakoff (1972) explains the scope ambiguity in (16b) by claiming that the reading in which the universal quantifier has wide scope derives from the deep structure (18a), while the reading in which the universal quantifier has narrow scope derives from the deep structure (18b):

(18) a. [$_S$ [Every x] [$_S$ [Some y] [$_S$ love x y]]].

b. [$_S$ [Some y] [$_S$ [Every x] [$_S$ love x y]]].

The scope of a quantifier, thus, includes whatever it *commands*, that is, every constituent dominated by the constituent dominating the quantifier.[15]

The crucial novelty in this line of analysis is that deep structure is now regarded as an abstract level with the same format as first-order logic representations. In other words, first-order quantificational representations are syntactically "wired" in deep structure syntactic representations. Accordingly, the latter contain not only lexical items but also abstract elements, such as variables, quantifiers, and other logical operators, like negation, modalities, and so on. These abstract constructs are then converted into surface phrase structures by the application of various types of transformations—such as, for instance, quantifier lowering—which replaces and inserts lexical items or deletes some of the abstract elements. To summarize, pushing to the extreme the assumption that deep structure provides all the compositionally relevant semantic information, generativist semantics was led to abandon the idea that deep structure is purely syntactic, thus breaking radically with the principle of the autonomy of syntax and the overall architecture of Chomsky's standard theory. Instead of postulating syntactic representations that serve as input to the interpretive semantic component, semantics is now conceived as a generative device that produces the deep layer—directly encoding the logical form of sentences—which is then converted by various transformations into surface structures. Therefore, generative semantics came to defend the view that in grammar "there is no dividing line between syntax and semantics" (McCawley 1972, 498), given that many logicosemantical phenomena—ranging from quantifier scope, to presuppositions, implicatures, and speech acts[16]—are represented directly as the level of deep structures. According to Lakoff (1972, 647), linguistics merges in this way with *natural logic* conceived as "the empirical study of human language and human reasoning." However, the price generative semanticists had to pay to achieve this progressive "logicization" of deep structures is the enormous complication of the transformation apparatus necessary to fill the wider and wider gap between deep and surface structure, due to the more and more abstract nature of the former.

Given the trend of regarding deep structures as abstract representations, they actually ceased to contain lexical items altogether, and progressively turned into predicate-argument structures familiar from first-order logic. Again, this shift stemmed from the need to solve important issues arising in the standard theory, like the representation of the so-called selectional restrictions. For instance, in (19), the verb *sink* has similar selectional restrictions with respect to its transitive and intransitive versions: The *NP* in the object position in (19a) has the same relation with the predicate as the *NP* in the subject position in (19b). There is, however, a difference between the two which consists in the presence of an extra *NP* in (19a) with the role of the agent causing the event described by the predicate:

(19) a. *John sank the boat.*
b. *The boat sank.*

On the hypothesis that selectional restrictions between a predicate and its arguments are determined at the level of deep structure (as claimed by the standard theory), one has now to analyze (19a) as derived from a deep structure representation containing as its proper part the structure associated with (19b). Lakoff analyzes (19a) as derived from the deep structure representation (20):

(20) $[_S$ John CAUSED $[_S$ the boat sink$]]$.

The predicate CAUSE in small caps marks the fact that it is actually an abstract item, which is then incorporated into the main predicate by a transformation which produces the transitive, agentive version of *sink*. This type of analysis has been extended to other causative verbs, such as *kill* that Lakoff took it to be equivalent to CAUSE-DIE, and so on.

The radical departure from the autonomy of syntax and the progressive logicization of deep structure pursued by generative semanticists found an incredibly high resonance among philosophers and logicians. For instance, Harman (1972) argues for the complete identification of deep structure with logical form regarded as the result of the paraphrase of a sentence into quantificational notation of the kind exemplified by (18). Moreover, following again generative semanticists, the subject-predicate asymmetry is considered as a surface feature of sentences, whose deep syntactic description is instead totally isomorphic with the standard predicate-arguments structure familiar from logic. Harman (1972, 30) arrives at the conclusion that "it is interesting to observe that what holds for logic holds for deep structure as well."

The idea that the deep structure of generative linguistics should be identified with logical form is also argued for by D. Davidson in *Semantics for Natural Languages* (1970). Adopting this stance allows for the possibility to view language inquiry as a common enterprise between linguists, logicians and philosophers:

> It is a question how much of a realignment we are talking about for linguistics. This depends largely on the extent to which the structure revealed by a theory of truth can be identified with the deep structure transformational grammarians seek. In one respect, logical structure (as we may call the structure developed by a theory of truth) and deep structure could be the same, for both are intended to be the foundation of semantics. (Davidson 1984, 63)

We find here two important points that make a difference with respect to the conception of language embodied in the standard theory, and which anticipate some of the crucial features of *Montague grammar*. First of all, logical form is to be identified with one of the levels of grammatical description, which in the case of transformational grammars is the deep structure level. Second, a proper semantics for natural language should take the form of a theory of truth, which assigns to sentences their truth-conditions in a recursive way. In particular, the claim that "a semantic theory for natural language cannot be considered adequate unless it provides an account for the concept of truth for that language along the general lines proposed by Tarski for formalized language" (Davidson 1984, 55) represents a major departure from Chomsky's radically internalist perspective on the study of language, which, as we saw with Katz and Fodor, rejected logical and truth-based approaches to the study of meaning. Although Davidson agrees with Chomsky that semantic differences in sentences sharing the same surface structure (e.g., *I persuaded John to leave* versus *I expected John to leave*) have to be accounted for in terms of differences at the level of deep structure, he claims, nevertheless, that these "intimations of structures" have to be derived ultimately from a suitable theory of truth which yields, for each sentence, its truth-conditions. The reader is referred to chapter 13 for the detailed description of Davidson's theory of truth. For him, it is such a theory that must serve as a ground for the notion of grammaticalness itself and must reveal the structure of sentences, which therefore is to be seen "through, the eyes of a theory of truth" (Davidson 1984, 61). One could not be further away from the principle of the autonomy of syntax. In a similar vein, Lewis (1972) argues for a referential, truth-conditional semantics for natural language, and at the same time claims that the ultimate criterion of adequacy for the grammar of a given language is its suitability to yield the truth-conditions of sentences in a recursive way.[17] To sum up, there is something common to both generative semanticists and certain logically minded philosophers of language: Both saw the convergence between logic and linguistics to be achievable only to the extent one abandons Chomsky's idea that the syntactic description of natural language is to be carried out independently of semantic constraints: "if we regard the structure revealed by a theory of truth as deep grammar, then grammar and logic must go hand in hand" (Davidson 1984, 59).

The idea that the logical form of sentences differs from their surface structure is a leitmotif in twentieth-century logic and analytic philosophy. It is therefore understandable that one of the crucial tenets of Chomskian linguistics, that of the surface structure being derived from a deep structure representation through various transformations, raised a huge amount of expectations concerning the possibility of finally identifying the level of linguistic description at which logical form is explicitly encoded. However, as we saw, these expectations could be really met only to the extent deep structures were conceived in a completely different way from that in which Chomsky himself conceived them. The deep structure in the *Aspects* was designed to be the interface with semantics and not encode logical form and other structural semantic properties. The move in this direction occurred with the reinterpretation of deep syntactic structures in generative semantics, a move that come to have an important influence in the first stages of the development of model-theoretic semantics. The reason of this influence is that it makes possible for the first time to see the relation between grammar and logic as internal, nonaccidental:

> Not all theories of linguistic structure guarantee that such a correspondence (*between grammatical structure and logical structure*) is not accidental. For example, the theory given in Chomsky's *Syntactic Structures* leaves open the question as to whether such correspondences are accidental. . . . Any rules relating English sentences to their logical forms would be independent of the rules assigning those sentences grammatical structures, though the rules assigning logical form might or might not depend on the grammatical structures assigned by rules of grammar. To the extent to which a theory of grammar assigns grammatical form independently of meaning, to that extent that theory will be making the claim that any correspondence between grammatical form and logical form is accidental. (Lakoff 1972, 546–547)

Chomsky has always claimed that such a correspondence exists, although its extension and form have to be established on empirical grounds. Apart from being a non sequitur, Lakoff's statement is a clear substantiation of the claim that to capture the evident correlation between syntax and semantics, syntactic rules should be couched in semantic terms. This idea is extremely close to the approach pursued by Montague in designing the formal architecture of language. According to it, grammatical forms are not determined independently of meaning, and the rules assigning grammatical structures to sentences run parallel to the rules that derive their logical form.

6. Montague Grammar and Model-Theoretic Semantics

In section 3.2 we mentioned Davidson's truth-functional program for the semantics of natural language, a challenge addressed to both logicians and linguists. At the beginning of the 1970s, Davidson's challenge was accepted by Richard Montague, who shared a similar view on the relation between artificial and natural language, as stated in the *incipit* of *English as a Formal Language* (1970) (Montague 1974):

> I reject the contention that an important theoretical difference exists between formal and natural languages. On the other hand, I do not regard as successful the formal treatments of natural languages attempted by certain contemporary linguists. Like Donald Davidson I regard the construction of a theory of truth ... as the basic goal of serious syntax and semantics; and the developments emanating from Massachusetts Institute of Technology offer little promise towards that end. (Montague 1974, 188)

In the short period of his activity, Montague pursued this goal by developing a rigorous formal system to describe the syntax and the semantics of natural language, as well as the relation between them, within the tradition of logical grammar. In particular, this was achieved by defining a fully compositional model-theoretic semantics in the spirit of Tarski (1936) and Carnap (1947), which also heavily relied on recent results in modal logic (Kripke 1963) and the foundations of intensional logic (Kaplan 1964).

What is usually known as Montague grammar (MG) corresponds roughly to the formal theory of natural language laid out by Montague in *English as a Formal Language* (EFL) (1970), *Universal Grammar* (UG) (1970), and *The Proper Treatment of Quantification in Ordinary English* (PTQ) (1973). Montague's work did not come out of the blue, but rather stood out in a research environment in which the possibility of exploiting the tools of logic for a formal description of natural language had come to a complete maturation and ramified into many directions. Thus, the term "Montague grammar" itself should be enlarged to include the important contributions made by Lewis (1972), Cresswell (1973), and many others, who together with Montague have been responsible for opening the field of model-theoretic semantics. Still, it is not possible to deny the central role occupied by Montague's own contribution, whose influence largely and rapidly outclassed other formal models for natural language semantics, particularly because of the extreme rigor with which the formalization of syntax and semantics was carried out in *MG*, as well as for the relevance and variety of linguistic phenomena to which Montague applied his system in the three papers.

Montague's work had a tremendous impact among both logicians and linguists, with the character of a true revolution. He managed to show that natural language, or some important fragments thereof, is amenable to formalization, thus bannishing the skepticism expressed by Carnap and Tarski before him. Moreover, Montague has also revealed the relevance and power of various logical tools—such as possible worlds semantics, intensional logic, higher-order logic, type theory, lambda calculus, and so on—for the purposes of providing a satisfactory formal description of natural language. However, this last claim has raised much concern. It is, in fact, quite controversial whether Montague's appeal to higher order logic and intensional logic is not only fully justified but also truly effective to tackle the problems it intends to address, thus making the departure from first-order logic unnecessary. Actually, this is one of the major points of disagreement between him and Davidson, who does not share Montague's appeal to possible world semantics. Still, it is undeniable that a large part of the logical machinery employed nowadays in formal semanticists derives form the logical tools underlying *MG*.

On the linguistic side, as Bach (1989) notices, the major advancement brought by Montague was to prove that natural language can be regarded as a formal system at the interpretive side, too. In fact, Chomsky's revolution had revealed that natural language can be satisfactorily described as a formal system at the syntactic level, but semantics was still regarded as lying largely beyond the possibility of such a treatment. Because of the deep mistrust in the application of logical techniques to semantic analysis, and of the suspicions toward truth-conditional semantics—particularly due to the Chomskian internalist and psychologist stand on language—semantic inquiry in the first years of generative linguistics was largely dominated by the *KF* paradigm, based on the decompositional analysis of meaning in terms of semantic markers. As we pointed out in section 4.3, *KF* has a strong lexicographically oriented approach to semantic analysis. The focus of the semantic analysis undertaken by Katz and Fodor is the representation of word senses, and of relations among them, such as analyticity, synonymy, semantic anomaly, polysemy, and so on. Although this new perspective on lexical meaning appears quite remarkable, still the inadequacies of the theoretical framework of *KF* have made the enterprise quite unsatisfactory. Among other things, unlike the truth-conditional approach, *KF* has been criticized for not being explanatory in a substantial way, given that semantic markers alone cannot provide any effective insight into interpretive processes. Despite heavily relying on lexical decomposition, generative semanticists should be given the credit for calling the linguists' attention to the centrality of problems such as quantification, operator scope, pronominal anaphora, and so on that pertain to the issue of logical form. Yet the quite protean and adventurous nature of generative semantics was not really able to lead to a solid framework within which to tackle these issues. The multilevel syntactic architecture typical of generative linguistics, notwithstanding its importance to overcome the shortcomings of traditional phrase structure grammar, had raised the important question of determining which representational layer is the input to semantic interpretation. The situation had become even more complex with the debate about the Katz–Postal hypothesis and the proposal advanced in generative semantics to regard deep structure as semantic in nature. In summary, few real advancements have been made on the issue of setting the relation between syntax and semantics on solid grounds, let alone of giving it a formal foundation. This situation partly justifies Montague's criticism (1974, 223) against transformational grammar:

> One could also object to existing syntactical efforts by Chomsky and his associates on grounds of adequacy, mathematical precision and elegance. ... In particular, I believe the transformational grammarians should be expected to produce a rigorous definition, complete in all details of the set of declarative sentences of some reasonably rich fragment of English ... before their work can be seriously evaluated.

Montague was actually able to provide a mathematically precise, logical analysis of a specific subfragment of English. But the revolutionary import of his contribution lies above all in the general framework he set up to formalize the relation between the logical semantics and the syntactic structure of natural language.

6.1. Compositionality and Universal Grammar

According to Frege (1984, 390),

> even if a thought has been grasped by an inhabitant of the Earth for the very first time, a form of words can be found in which it will be understood by someone else to whom it is entirely new. This would not be possible, if we could not distinguish parts in the thought corresponding to the parts of a sentence, so that the structure of the sentence can serve as a picture of the structure of the thought.

In this perspective, therefore, the *principle of compositionality*— stating that the interpretation of a complex expression is a function of the interpretations of its parts—is the key ingredient to explain linguistic creativity. Compositionality is usually satisfied by logical languages, in which the definition of semantics runs parallel to the recursive definition of syntax, like in the case of Tarski's definition of the satisfaction predicate. Actually, compositionality provides a finite method for the semantic interpretation of an infinite number of expressions.

On the other hand, Chomsky has claimed that the explanation of linguistic creativity cannot be based on the assumption of a systematic pairing between syntax and semantics: Unlike formal languages, he did not find this correspondence warranted for natural languages. Thus, the capacity of understanding and producing a potentially infinite number of sentences would rather be grounded in the generative capacity of the syntactic component, which can and must be identified independently of any semantic considerations. Syntactic rules generate structures that in turn drive semantic composition belonging to an external interpretive module. This kind of architecture of the grammar, having at its center the principle of the autonomy of syntax, is strongly criticized by Montague, who rejects the possibility of considering a syntactic theory of language, independently of semantic considerations. Actually, the main objection that Montague addresses to transformational grammar in *UG* is exactly its lack of relevance for the enterprise of developing a semantics for natural language.[18] He thinks that

i. A proper semantic theory must be grounded on a theory of truth, and

ii. The core function of syntax is to provide the necessary structural backbone for semantic interpretation.

Thus, while for Chomsky (1957) the purpose of syntax is to generate the grammatical sentences of a language, for Montague syntax is mainly subservient to the goal of defining how the interpretation of a sentences depends on the interpretations of its components. In other terms, in the case of Montague the problem of finding the right syntactic structure becomes part of the problem of how to implement the requirement of compositionality.

Consistent with his tenet that no actual difference exists between formal and natural language, Montague solves the problem of the interpretability of a potentially infinite number of sentence in the same way as Frege (Montague 1974, 217). Interpreting means for him determining the truth values of sentences, something to be achieved

> by assigning extra-linguistic entities to all expressions involved in the generation of sentences (including among these, sentences themselves) in such a way that (a) the assignment of a compound will be a function of the entities assigned to its components, and (b) the truth value of a sentence can be determined from the entity assigned to it.

UG represents the most general formulation of Montague's formal framework, where the principle of compositionality is given an algebraic formulation in terms of an homomorphism between a syntactic algebra and a semantic algebra. The algebraic perspective allows Montague to specify the structure of syntax, the structure of semantics and the relation between them by abstracting away from specific ontological and epistemological commitments, as well as from the particular format of the syntactic rules. The aim of the paper is to provide the universal architecture of syntax, semantics, and of their relation. However, it is crucial to keep in mind that the term *universal* in Montague has a radically different content and import than in generative linguistics. In the latter, universal grammar means the rules and principles that define the class of human learnable languages and that form the innate component of the faculty of language, while in the former universal grammar intends to capture the constraints on the structure of whatever possible language, artificial or natural.

In *UG*, the *syntax* is defined as the system $\langle A, F, X_\delta \rangle_{\delta \in \Delta}$, such that:

(21) i. $\langle A, F\rangle$ is an algebra with A a nonempty set of expressions, and F a set of operations on A;

ii. Δ is the set of syntactic categories;

iii. for all $\delta \in \Delta$, X_δ is a subset of A, that is, the set of basic expressions of category δ (e.g., intransitive verbs, common nouns, etc.).

The operations in F apply to tuples of basic expressions to generate other expressions like A. For instance, F may include a simple operation like concatenation, or any other operation of arbitrary complexity. This algebra generates a *disambiguated language*, the set of all expressions which can be formed starting from some basic expressions and applying operations on them a finite number of times. An *interpretation* for the disambiguated language is a system $\langle B, G, f\rangle$, such that:

(22) i. $\langle B, G\rangle$ is the semantic algebra similar to $\langle A, F\rangle$, such that B is the set of meanings prescribed by the interpretation, G is the set of semantic operations corresponding to the syntactic operations F and which apply to tuples of elements in B;

ii. f is a function from $\mathrm{U}_{\delta\in\Delta}X_\delta$ into B, that is, it assigns meanings to the basic expressions of the generated language.

G may contain operations like function-argument application, function composition, and so on. Crucially, given the system $\langle A, F, X_\delta\rangle_{\delta\in\Delta}$ and the interpretation $\langle B, G, f\rangle$, the meaning assignment to the generated language is defined by Montague as the unique *homomorphism* g from $\langle A, F\rangle$ into $\langle B, G\rangle$ such that:

(23) i. F and G are sets of operations with the same number of places;

ii. g is a function with domain A and range included in B;

iii. for every n-ary operation F and G and every sequence $a_1, \ldots, a_n$ in A, we have $g(F(a_1, \ldots, a_n)) = G(g(a_1), \ldots, g(a_n))$;

iv. $f \subseteq g$.

The principle of compositionality is implemented as the homomorphism requirement, and not an isomorphism requirement, to allow for the fact that two distinct syntactic expressions may have the same meaning, but each syntactic expression must have at most one meaning. Defining the compositionality as a homomorphism between two algebras requires a disambiguated level of representation in syntax. This is not, however, what happens in natural language, where the same linear sequence of elements can be structurally ambiguous, for example, in the case of *Every man love some woman* or *John saw a man in the park with a telescope*, etc. To account for this fact, in addition to defining a disambiguated language DL generated by an algebra, Montague also defines a *language* L as the pair $\langle DL, R\rangle$, where R is a relation with domain in A. R maps expressions of A onto expressions of A, which so to speak represent their surface representation. This relation is often referred to as an "ambiguating" relation,[19] because it maps expressions of a disambiguated language onto expressions to which more than one syntactic description may correspond. So, an expression ζ of the language L is ambiguous if and only if there are at least two expressions ζ' and ζ'' generated by the relevant algebra such that $\zeta' R \zeta$ and $\zeta'' R \zeta$. This solution amounts to saying that a language may contain expressions to which there actually correspond two different syntactic representations generated by the syntax. The interpretation is defined on the disambiguated algebra: If ζ is an expression of the language L and g the homomorphism of the interpretation, then g means b if and only if there is a $\zeta' \in DL$, such that $\zeta' R \zeta$ and $g(\zeta') = b$. This implies that an ambiguous expression will also have two or several interpretations, each corresponding to a particular syntactic representation.

In *MG* the principle of compositionality is implemented in terms of the so-called *rule-by-rule interpretation* (Bach 1976). According to this procedure, the syntax is given by a recursive definition starting from a set of basic expressions of given categories with rules that operate on them to produce new expressions. Here is an example with F_I an arbitrary syntactic operation:

(24) *Syntactic Rule* S_I
If α is a well-formed expression of category A and β is a well-formed expression of category B, then γ is a well-formed expression of category G, such that $\gamma = F_i(\alpha, \beta)$.

Semantics is then given by a parallel recursive definition, in which basic expressions are assigned basic semantic values, and for each syntactic rule S_I there is a semantic rule of the following form:

(25) *Semantic Rule* S_I
If α is interpreted as α' and β is interpreted as β', then γ is interpreted as γ', with $\gamma' = G_k(\alpha', \beta')$.

G_k is a semantic operation (e.g., function-argument application) that combines the semantic values of expressions to produce the semantic value of the complex expression. The rule-by-rule interpretation is actually the method that is normally employed to define the interpretation of formal languages, and is employed by Montague in *PTQ* to provide a compositional formal semantics of English. When the systems of rules that make up the syntax and the semantics are recast as algebras, the rule-by-rule correspondence becomes the requirement of homomorphism. So again, the framework defined in *UG* is intended to provide the most general method to satisfy the constraint of compositionality.[20]

Because Montague's goal is to define a theory of truth for a language, the notion of interpretation just given is not per se sufficient, given that it is simply defined as a particular type of mapping between algebras without further constraints on the format of semantics, consistently with the full generality of the approach pursued in *UG*. This is the reason why Montague introduces the notion of *Fregean interpretation*, a semantic algebra consisting of a model-theoretic structure containing domains with a typed structure. The extensive use of type-theory and intensional logic to define the formal semantics of natural language is one the most important innovations brought by Montague. Actually, in the years immediately preceding the three papers devoted to the formalization of English, Montague did important work in intensional logic, leading to the unification of temporal logic and modal logic and more generally to the unification of intensional logic and formal pragmatics, defined by Bar-Hillel (1954) as the study of indexical expressions, that is, words and sentences whose reference cannot be determined without knowledge of the context of their use. In addition, Montague integrated the work of Carnap (1947), Church (1951), and Kaplan (1964) into a fully typed intensional logic, in which the function-argument structure typical of type theory (Russell) merges with the functional treatment of intensions. The latter are in fact regarded by Montague as functions from possible-world and time moments to extensions. The results of this more foundational work are contained in *Pragmatics* (1968), *On the Nature of Certain Philosophical Entities* (1969), and *Pragmatics and Intensional Logic* (1970).

To guarantee that the mapping from the syntactic to the semantic algebra is a homomorphism, it is necessary that the model-theoretic structure contains a domain of interpretation for every syntactic category. In *UG* and in *PTQ*, Montague defines recursively an infinite system of domains via an intensional type theory, and then establishes a relation between syntactic categories and a relevant sets of defined types.[21] Montague first defines the set of *types* T in the following way:

(26) i. e is a type;

ii. t is a type;

iii. if a and b are types then $\langle a, b\rangle$ is also a type;

iv. if a is a type, then so is $\langle s, a\rangle$.

Each type individuates a certain domain, which will provide the interpretation of the expressions of the language having this type. Thus in (i)–(ii) the two basic types, e and t are introduced. Their interpretation varies: In *UG* and *PTQ* e is the type of entities, and t is the type of truth values, while in the *EFL* system, t is the type of propositions defined as functions from possible worlds to truth values. The clause in (iii) defines the functional types, that is, the types of functions from objects of type a to objects of type b. Finally, the clause in (iv) defines the intensional types, that is, the types of functions from indices (usually possible worlds or world-time pairs) to objects of type a. Notice that the type s has no independent existence, that is, it does not belong to the domain of objects of the structure itself, and does not represent the interpretation of any category of expressions.

Given a nonempty set A (to be regarded as the set of entities or individuals), a set I of possible worlds, and a set J of moments of time, for every type $\tau \in T$ Montague recursively defines the domain associated to τ, $D_{\tau,A,I,J}$ (*the domain of possible denotations of type τ relative to A, I, and J*) in the following way:[22]

(27) i. $D_{e,A,I.J} = A$;

ii. $D_{t,A,I,J} = \{0,1\}$;

iii. $D_{\langle a,b\rangle,A,I,J} = D_{b,A,I,J}^{D_{a,A,I,J}}$;

iv. $D_{\langle s,a\rangle,A,I,J} = D_{a,A,I,J}^{IxJ}$.

In *PTQ*, the domain $D_{\langle s,a\rangle,A,I,J}$ is defined as the set of *senses* (*meanings* in the *UG* terminology) of type a,[23] regarded as intensional entities, that is, functions from pairs of indices to objects of type a.

The relation between the syntactic categories of a language L and the semantic types is determined by a function of *type assignment*, defined as the function σ from Δ (the set of syntactic categories) into T, such as $\sigma(\delta_0) = \tau$. Finally, a *Fregean interpretation* for L is defined as an interpretation $\langle B, G, f\rangle$ such that for some non-empty sets A, I, J and type assignment σ:

(28) i. For every type τ, B includes at least the domain of possible denotations for τ, that is, $B \subseteq \mathrm{U}_{\tau\in T} D_{\tau,A,I,J}$;

ii. For every syntactic category δ, such that $\sigma(\delta) = \tau$, and every basic expressions $\zeta \in X_\delta$, $f(\zeta) \in D_{\tau,A,I,J}$;

iii. For every syntactic operation F_I there is a corresponding semantic operation G_I, such that if F_I applies to expressions of category δ' to produce expressions of category δ'', then G_I applies to entities of type $\sigma(\delta')$ to produce entities of type $\sigma(\delta'')$.

The system defined by these rules is then applied to two specific examples, the language of *intensional logic*, and a fragment of English, with the purpose of showing that the same procedure allows both formal and natural languages to be treated alike. The fragment of English formalized by Montague is very complex, including intensional verbs, relative clauses, quantifiers, and so on. Before giving some of the details of Montague's analysis, it is important to spend a few words to describe two notions that play a crucial role in *MG*, namely, the *method of fragments* and the *method of indirect interpretation*.

The former, one of the novelties of Montague's approach, made its first appearance in *EFL*. It consists in writing a complete syntax and truth-conditional semantics for a specific fragment of a given language to make fully explicit assumptions employed in the formalization.

The method of indirect interpretation consists of interpreting a fragment of a given language via its translation into a formal language, which is in turn interpreted in a Fregean structure. It contrasts with the method of direct interpretation where the syntactic algebra, the semantic algebra (corresponding to the Fregean interpretation), and the homomorphism between them are given explicitly. The direct method is employed by Montague in *EFL* to formalize a fragment of English, while in *UG* and *PTQ* the indirect method is adopted, with intensional logic serving as an intermediate language into which the fragment of English is translated. Montague provides a general theory of *compositional translation*, in which a homomorphism g is built up between the syntactic algebra Syn_1 defining the source language L_1 and the syntactic algebra Syn_2 defining the target language L_2, for which, in turn, there is a homomorphism h with a semantic algebra *Sem* which provides an interpretation for L_2. Because one can define an operation of composition of g with h and show it is a homomorphism k from Syn_1 to *Sem*, then it follows that *Sem* may serve directly as an interpretation for the source language L_1. In other words, the compositionality of translation makes the intermediate level totally dispensable. Nevertheless, the compositionality of translation provides, according to Montague, a more perspicuous representation of the logical form of expressions, thus making the indirect method of interpretation preferable as a way to define a formal semantics for a given fragment of English.

6.2. PTQ: The Standard Model of Montague Grammar

The formal analysis of the fragment of English presented in *The Proper Treatment of Quantification in Ordinary English* (*PTQ*) represents an illustration of the general algebraic method for a compositional analysis of language exposed in *UG*. This paper is the best vantage point to see at work Montague's approach to natural language, not only because the fragment discussed there is the largest of the three that Montague formalized, but also because it is the paper that had the strongest impact on the linguistic community and on the subsequent development of model-theoretic semantics. Thus, *PTQ* represents a sort of standard model of *MG* up to the point of being almost identifiable with it.

In *PTQ*, the *syntax* of the fragment of English makes use of a categorial grammar reminiscent of Ajdukiewicz's system, whose set of categories *Cat* is defined as the smallest set X such that:

(29) i. $t \in X$, with t the category that corresponds to sentences (the letter t marks the fact that sentences are the expressions that can have a truth value);

ii. $e \in T$, with e the category of entities;

iii. If A, $B \in X$, $A/B \in X$ and $A//B \in X$.

The "double-slash" category is the only actual innovation brought to Ajdukiewicz's categorial grammar, and it is used only to mark the syntactic difference from A/B. That is to say, A/B and $A//B$ are semantically alike, although they have a different syntactic role. The set *Cat* contains an infinite number of categories, out of which only a restricted number is actually used in *PTQ*, which is listed in table 16.1, together with the abbreviations given

Table 16.1 SYNTACTIC CATEGORIES IN *PTQ*

Category	*Abbreviation*	*Linguistic description*	*Example*
t/e	IV	Intransitive verb phrases	run, walk
t/IV	T	Terms	John, Mary, he_0, he_1, ...
IV/T	TV	Transitive verb phrases	find, eat, seek
IV/IV	IAV	IV-modifying adverbs	slowly, allegedly
t//e	CN	Common nouns	man, unicorn, temperature
t/t		Sentence-modifying adverbs	necessarily
IAV/T		IAV-making prepositions	in, about
IV/t		Sentence-taking verb phrases	believe that, assert that
IV//IV		IV-taking verb phrases	try to, wish to

to some of them by Montague and their standard linguistic equivalent. The "double-slash" category is, for instance, used to distinguish IV from CN: In fact, while semantically they are both interpreted on the domains of functions from individuals to truth values, at the syntactic level IV combine with terms to produce sentences, while CN are used to build up terms. For each of the categories listed in table 16.1, Montague introduces a set of basic expressions, some of which are exemplified in the fourth column. Basic expressions form what we might call the lexicon of the selected fragment of English, although it is important to notice that in many respects it greatly differs from the linguists' conception of lexicon. For instance, the categories IV/t and IV//IV contain basic expressions like *believe that* or *try to*, which are not lexical under a strictly linguistic point of view. Similarly, the basic expressions belonging to category T, include, besides proper nouns, an infinite set of variables, $he_0, he_1, he_2, \ldots,$ which play a crucial role in Montague's analysis of relative clauses, quantification and anaphora. Moreover, one of the characteristics of *MG* is that there is no expression in the language, neither basic nor derived by syntactic rules, belonging to the category *e*. Thus, this category is only used to create other categories, but, in Montagovian terms, it has no linguistic exemplification.

The grammar of *PTQ* includes the set of syntactic rules described in (24), which generate the set of expressions of various categories (sentences included). The categories determine which expressions are to be combined with which, as well as the category of the resulting expressions. In contrast to the categorial grammars of Ajdukiewicz and Bar-Hillel (see section 2.5), the syntactic operations include a rule of concatenation. Montague introduces 17 syntactic rules, grouped in five clusters. Some of the *rules of functional application* (S4–S10) coincide with mere concatenation, other rules of the same group include operations that inflect the verb to the third singular person of the simple present, to satisfy the agreement with the subject, or inflect a pronominal variable to the accusative form, when it combines with a transitive verb. In the cluster of *basic rules* (S1–S3), S2 introduces the determiners *every*, *the*, and *a* syncategorematically,[24] while S3 is a rule for forming relative clauses. The *rules of conjunction and disjunction* (S11–S14) introduce *and* and *or* syncategorematically, and the *rules of tense and sign* (S17) inflect a verb for tenses other than present (future and present perfect) and adds negation. Finally, the *rules of quantification* (S14–S16) replace a variable inside an expression with a term: These rules have a crucial role in Montague's treatment of scope and quantification, as we will see in 6.2.3. The *meaningful expressions* of the given fragment of English are those generated by the finite application of the syntactic rules to the basic expressions. The way a sentence is constructed through a finite application of the set of syntactic rules is called by Montague the *analysis tree* of the sentence.

The Montagovian conception of grammar greatly differ from the generative-transformational grammars, both from a formal and a substantial point of view. First of all, the syntactic operations operate on strings and not on trees or labeled bracketing. The analysis tree marks the history of the derivation leading to a meaningful expression, but it is not in itself a symbolic object which can be manipulated and transformed by syntactic rules. Second, there is no notion of grammaticalness other than that of meaningfulness, in agreement with Montague's rejection of Chomsky's autonomy of syntax. Therefore, the syntactic rules in *MG* resemble more Husserl's meaning connection rules and the rules of traditional logical grammar, rather than the rules of the syntactic component in generative grammar. The structure of the analyses tree in *MG* is intended to reflect meaning constitution and semantic structural ambiguities, and not so much purely syntactic criteria of constituency.

PTQ implements the algebraic framework set up in *UG* in terms of a rule-by-rule interpretation procedure described in section 6.1. Since the interpretation is performed according to the indirect method, syntactic rules are actually paired with translation rules into the language of intensional logic (*IL*), which is then interpreted in a Fregean structure through a homomorphic mapping. Notice that Montague's rule-by-rule method bears some similarities to the projection rules which in *KF* operate on tree structures (see 4.3). However, the similarity should not hide a deeper difference between the two: In *KF* the projection rules provide an interpretation to an autonomous syntactic component, while in *MG* the rules of syntax are designed in such a way as to display the semantic structures.

Given the set of types T defined in (26), the set *ME* of meaningful expressions of *IL* includes an infinite number of constants and variables for each type $\tau \in T$, and a set of expressions generated by a list of recursive rules. *IL* is then interpreted in a (*Fregean*) *interpretation* or *intensional model*, which is a quintuple $M = \langle A, I, J, \leq, F \rangle$, such that:

(30) i. A, I, and J are nonempty sets, the domain of individual entities, the set of possible worlds, and the set of moments of times, respectively;

ii. for every type $\tau \in T$, $D_{\tau,A,I,J}$ is the set of possible denotations of τ, as defined in (27);

iii. $\leq$ is a linear order over J;

iv. F is a function taking as arguments constants of *IL*, such that for every type $\tau \in T$ and every constant α of type τ, $F(\alpha) \in D_{\langle s,\tau\rangle A,I,J}$.

As we pointed out in 6.1, for every type τ, the domain $D_{\langle s,\tau\rangle A,I,J}$ corresponding to $D^{IxJ}_{a,A,I,J}$, is called by Montague the set of *intensions* or *senses* of expressions of type τ. Intensions—which Montague regards to be the equivalent to Fregean senses—are defined as functions from world-time pairs $\langle i,j\rangle$, to entities of appropriate type. The *extension* (or *denotation*) of a certain expression with respect to the pair $\langle i,j\rangle$ is obtained in the standard way by application of the intension function to the argument $\langle i,j\rangle$. Montague interprets the constants of *IL* as intensions (30iv), and then, given an assignment g, he recursively defines the notion of *extension with respect to M and g* for all the *ME* of *IL*.

The lambda calculus is an important part of *IL*, the intermediate language employed in the interpretation of English. Here are some central definitions:

(31) i. If $\alpha \in ME_a$ and u is a variable of type b, then $\lambda u\alpha \in ME_{\langle b,a\rangle}$;

ii. If $\alpha \in ME_{\langle a,b\rangle}$ and $\beta \in ME_b$, then $\alpha(\beta) \in ME_b$;

iii. If $\alpha \in ME_a$ then $[\hat{}\alpha] \in ME_{\langle s,a\rangle}$;

iv. If $\alpha \in ME_{\langle s,a\rangle}$ then $[\check{}\alpha] \in ME_a$.

Given an expression of type a and a variable of type b, the extension of an expression like $\lambda u\alpha$ (31i) is a function belonging to the domain $D_{\langle b,a\rangle,A,I,J}$, which associates with every argument x of type b the value that α has when the variable u denotes x. Montague must be given credit for introducing lambda calculus into the linguistic community, to whom it was virtually unknown before him. This calculus has rapidly become one of the most powerful tools for the formal description of natural language semantics. Montague himself used lambda expressions for the analysis of relative clauses, conjunction, and quantification.

In (31ii) we find another crucial ingredient of *MG*, *functional application*: The expression $\alpha(\beta)$ denotes the result of applying the function denoted by α to the argument denoted by b. That is to say, the extension of $\alpha(\beta)$ is the value of the extension of α, when applied to the extension of b. Moreover, every expression γ of type $\langle a,t\rangle$ denotes a set B of entities of type a, or equivalently the *characteristic function* of B, that is, the function from the domain of entities of type a to $\{1,0\}$, such that it assigns 1 to all the entities that are elements B, and 0 otherwise. Then, if γ has type $\langle a,t\rangle$ and α has type a, "we may regard the formula $\gamma(\alpha)$ ... as asserting that the object denoted by α is a member of the set denoted by γ" (Montague 1974, 259).[25]

Finally, (31iii, iv) introduce operators that move back and forth between an expression and its intensional type. The "$^\wedge$" (up or cap) operator takes an expression α and forms a new expression denoting the intension of α: The extension of $^\wedge\alpha$ is then the intension of α. The "$^\vee$" (down or cup) operator performs the reverse operation when applied to intensional types. Thus, the extension of $^\vee\alpha$ with respect to a certain pair $\langle i, j\rangle$, is the result of applying the intension of α to $\langle i, j\rangle$, that is to say, it is the extension of α at $\langle i, j\rangle$.[26]

Given the indirect method adopted in *PTQ*, the bulk of the interpretation procedure consists in providing a compositional translation from English into expressions of *IL*, which takes place in three steps.

First, a mapping f is introduced, defined on the categories of English with arguments in the types T, so that every English expression of category A is translated into an expression of *IL* of type $f(A)$. The mapping is defined as follows:

(32) i. $f(e) = e$,
ii. $f(t) = t$,
iii. $f(A/B) = f(A//B) = \langle\langle s, f(B)\rangle, f(A)\rangle$, for every category A and B.

The particularity of this definition lies in the third clause. In fact, it states that functional categories (such as IV or T, etc.) correspond to a function from intensions of objects of type $f(B)$ to objects of type $f(A)$. That is to say, expressions of functional types are always translated into expressions of *IL* which denote functions operating on the intensions of their argument. For instance, in *PTQ* expressions of category CN or IV are assigned the type $\langle\langle s, e\rangle, t\rangle$, where $\langle s, e\rangle$ is the type of what Montague refers to as *individual concepts,* that is, functions from world-time pairs to individual entities. As it will be seen in 6.2.2, the reason for this choice lies in the analysis of the expressions creating intensional contexts. In particular, the use of individual concepts and the characterization of IV and CN expressions as sets of individual concepts are motivated by Montague by the failure of the following argument:[27]

(33) a. *The temperature is ninety.*
b. *The temperature is rising.*
c. *Ninety is rising.*

The fact that the truth of (33c) does not follow from the truth of the two premises can be explained in the following way: The value of a number word like *ninety* is always equal to itself in every point of reference in which it is evaluated, but this is not true for noun phrases like *the temperature* or *the price*, whose denotations can change from context to context (temperatures and prices can rise and fall). Montague (1974, 267–268) accounts for this difference by assuming that a noun phrase like *the temperature* does not denote an individual entity, but a function from world-time pairs to individual entities (i.e., an individual concept), and that the IV *rise* is inherently intensional, that is, "unlike most verbs, depends on its applicability on the full behavior of individual concepts, not just on their extension to the actual world and ... moment of time." In this way the failure of (33) is seen to follow from the fact that in (33b) the verb *rise* applies to an individual concept denoted by a subject noun phrase, while the identity in (33a) holds between the individual entity ninety and the individual entity that is the extension of the individual concept denoted by *the temperature* at the actual world and moment of time. In his attempt to grant the maximum level of generality to the formal framework for the interpretation of the English fragment, Montague generalizes the interpretation of *price* and *rise* as sets of individual concepts to every IV and CN, while the fully extensional behavior of other elements of these categories is captured through meaning postulates (see 6.2.1.). However, Montague's explanation of (33) as well as his interpretation of IV and CN as functions of individual concepts has been widely criticized. Actually, the idea of individual concepts soon became quite controversial, and Bennett (1974) proposed an amendment to Montague's type-assignment to syntactic categories which assumes no individual concepts at all and in which IV and CN are assigned the type $\langle e, t\rangle$. As a result, the function f in (32) can be redefined as follows:

(34) i. $f(e) = e$,
ii. $f(t) = t$,
iii. $f(\mathrm{IV}) = f(\mathrm{CN}) = \langle e, t\rangle$,
iv. $f(A/B) = f(A//B) = \langle\langle s, f(B)\rangle, f(A)\rangle$, for every category A and B.

This amendment has led to a major simplification of the translation procedure, which has quickly found its way in standard expositions of *MG* (e.g., Dowty et al. 1981) and will also be assumed in the rest of this chapter to describe the formalization of the English fragment in *PTQ*.[28]

The second step of the translation procedures takes care of the translation of the lexical items, that is, of the basic expressions, into *IL*. To this purpose, Montague defines a function g defined on the set of basic expressions, except for the verb *be*, sentence modifying adverbs (e.g., *necessarily*), and basic expressions of type T, that is, proper nouns and variables (see table 16.1), all of which are translated into complex logical expressions of *IL*:

- the verb *be* is translated as $\lambda P \lambda x \,{}^{\vee}P({}^{\wedge}ly(x = y))$, which has type $\langle\langle s, \langle\langle s, \langle e, t\rangle\rangle, t\rangle\rangle, \langle e, t\rangle\rangle$;
- *necessarily* is translated as $\lambda p \Box {}^{\vee}p$, which has type $\langle\langle s, t\rangle, t\rangle$;
- the translations of proper nouns and variables will be discussed in section 6.2.1.

All other lexical expressions are translated as constants of *IL* of appropriate type. For instance, the intransitive verb *walk* of category IV is translated into the constant $walk'$ of type $\langle\langle s, e\rangle, t\rangle$, which denotes a function from individual concepts to truth values.

In the third step of the procedure for compositional translation, Montague provides a translation rule for each syntactic rule that generates a meaningful expression of the English fragment. A sample of the 17 translation rules proposed in *PTQ* will be closely inspected in the following sections, discussing some of the most influential and controversial solutions offered in *MG* for the characterization of the logical form of natural language, that is, the analysis of noun phrases (6.2.1), the treatment of intensional constructions (6.2.2), and the representation of scope ambiguities (6.2.3). However, as a general remark, it is worth emphasizing that for Montague the function-argument application, which appears in the interpretation of all the basic grammatical relations, is the fundamental semantic glue. In fact, in *PTQ* all the nonbasic semantic types are constructed as functional types. This way of building the semantic interpretation of complex expressions represented an absolute novelty for the linguistic community and had an enormous impact. This procedure stands in deep contrast to the procedure of semantic composition in *KF*, which is performed through a process of unification of feature clusters in which functional application had no role to play. Montague's extensive use of function-argument structures in semantics brought to a new life the machinery of categorial grammar, neutralizing some of the criticisms levelled against it by Chomsky on account of its limited explanatory power. Given a function-argument based semantics, categorial grammar seemed to offer a very good syntactic layer to build a fully compositional model-theoretic semantics for natural language, especially if it is enriched in such a way that it can handle the structural complexity of natural language. As we showed, one such emendation was performed by Montague himself in *PTQ*, when he used categorial grammar to define the mapping between syntactic categories and semantic types as the basis for the homomorphic translation: He did not limit the operation of syntactic composition to concatenation. As Partee and Hendriks (1997, 30) remark, while in classical categorial grammar the derivation tree that displays the application of the syntactic rules is isomorphic to the surface structure of the relevant string, in *PTQ* this is no longer true, and "it is the analysis tree which displays the semantically relevant syntactic structure."

6.2.1. The Interpretation of Noun Phrases

In *PTQ*, proper names, pronouns, and noun phrases prefixed by determiners like *every* or *the* belong to the same syntactic category of terms, T, despite the fact that the first two are basic expressions, while determiners are introduced syncategorematically via syntactic rules. Because T = t/IV, then given (34), every term is assigned the type $\langle\langle s, \langle e, t\rangle\rangle, t\rangle$. The type $\langle s, \langle e, t\rangle\rangle$ is the type of functions from world-time pairs to sets of individuals, and is called by Montague the type of *properties*. Thus terms are regarded as denoting sets of properties, or equivalently functions from properties to truth values. This approach resembles Frege's analysis of quantifiers as second-order concepts: In *PTQ* a noun phrase like *every man* is a second-order predicate, true of a property of individuals if every individual that is a man has that property. For instance, the sentence *Every man dreams* is interpreted as stating that the property of dreaming has the property of being true of every man. *PTQ* provides the following translation rule (T2) for phrases containing the determiners *every*, *a/an* and *the*, where P is a variable of type $\langle s, \langle e, t\rangle\rangle$, and x and y are variables of type e:

(35) If ζ is of category CN and translates as ζ', then:

a. *every* ζ translates into $\lambda P \forall x[\zeta'(x) \rightarrow {}^{\vee}P(x)]$;

b. *the* ζ translates into $\lambda P \exists x[\forall y[\zeta'(y) \rightarrow x = y] \wedge {}^{\vee}P(x)]$;

c. *a/an* ζ' translates into $\lambda P \exists x[\zeta'(x) \wedge {}^{\vee}P(x)]$.

A major difference between the *PTQ* analysis of noun phrases and Frege's analysis is the view of proper names, which in the former are interpreted as sets of properties.[29] While in Frege proper names denote objects, that is, individual entities, in Montague their type is the same as that of quantificational expressions, $\langle\langle s, \langle e, t\rangle\rangle, t\rangle$. Thus the translation of a proper name like *John* is $\lambda P {}^{\vee}P(j)$, with j a constant of type e which denotes the individual John, and the denotation of the name *John* is the set of properties that John has. Montague's analysis is careful to obey the principle of compositionality as a strict architectural constraint on the semantics for natural language: Proper names belong to the same category as quantified phrases, that is, they are terms, and therefore they must have the same type.[30] In fact, if two expressions of the same syntactic category would be assigned two different types, the principle of compositionality would be violated. Accordingly, no expression in natural language is assigned the type e in *PTQ*: Even those expressions that would most naturally seem to denote individual entities (i.e., pronouns and proper names) are actually assigned a higher type.

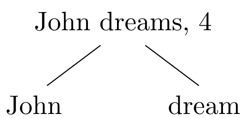
One of the most interesting consequences of the analysis of terms in *PTQ* is that it is possible to provide a uniform formalization of the logical form of the sentences in which they occur, irrespective of whether they contain quantified noun phrases or truly referential terms. As an example, let us consider the case of simple subject-predicate sentences:

(36) a. John dreams.

b. Every man dreams.

Traditionally, these sentences are taken to provide a clear example of the mismatch between linguistic surface form and semantic structure. Although they have the same structure, they are attributed distinct logical forms, $\text{dream}'(j)$ and $\forall x[\text{man}'(x) \rightarrow \text{dream}'(x)]$, respectively. This is not any longer so in Montague's *PTQ* where their syntactic analysis is as follows:

(37) a.



b.

Every man dreams, 4

Every man   dream

The syntactic rule that combines a subject term and an intransitive verb (S4) is associated to the following translation rule (T4):

(38) If δ is an expression of category T and β is an expression of category IV, and δ and β translate into an δ' and β' respectively, than $F_4(\delta, \beta)$ translates into $\delta'({}^{\wedge}\beta')$.

Therefore, the two syntactic trees in (37) give rise to the following parallel translations into expressions of *IL*.

(39) a.

$\text{dream}'(j)\ t\ \ {}^{\vee}{}^{\wedge}\text{-elimination}$

$\Updownarrow$

${}^{\vee}{}^{\wedge}\text{dream}'(j)\ t\ \ \lambda\text{-conversion}$

$\Updownarrow$

$\lambda P\,{}^{\vee}P(j)({}^{\wedge}\text{dream}')t$

$\lambda P\,{}^{\vee}P(j)\langle\langle s, \langle e,t\rangle\rangle, t\rangle$ $\text{dream}'\ \langle e,t\rangle$

b.

$\forall x[\text{man}'(x) \to \text{dream}'(x)]\ t\ \ {}^{\vee}{}^{\wedge}\text{-elimination}$

$\Updownarrow$

$\forall x[\text{man}'(x) \to {}^{\vee}{}^{\wedge}\text{dream}'(x)]\ t\ \ \lambda\text{-conversion}$

$\Updownarrow$

$\lambda P \forall x[\text{man}'(x) \to {}^{\vee}P(x)]\ ({}^{\wedge}\text{dream}')\ t$

$\lambda P \forall x[\text{man}'(x) \to {}^{\vee}P(x)]$ $\langle\langle s, \langle e,t\rangle\rangle, t\rangle$ $\text{dream}'\ \langle e,t\rangle$

In both cases, the top node of the tree receives a structurally similar translation, consisting of the functional application of a *IL* expression denoting a set of properties to the property denoting the expression associated to the predicate. The double arrows in (39) show that this formula can be further simplified through the "meaning-preserving" operations of *IL* logic (such as λ-conversion and the ${}^{\vee}{}^{\wedge}$-theorem), which produce logically equivalent formulas. The expressions that are usually regarded as exhibiting the logical form of the sentences in (36) then correspond to the result of such simplification. However, notwithstanding the fact that (36a) and (36b) come to be associated with very different expressions of *IL*, they share the same pattern of composition, exhibited by the analysis tree and the step-by-step translation procedure. This also opens the issue of the real status of the notion of logical form in Montague, since prima facie it seems that expressions of *IL* and analysis trees are both plausible candidates to play this role, and as in the case of (39) it can happen that two natural language expressions have the same analysis tree, but they ultimately correspond to different expressions of *IL*. As Partee and Hendriks (1997, 43) argue,[31] actually "the analysis trees are . . . the best candidates for a level of 'logical form,' if by 'logical form' one means a structural representation of an expression from which its meaning is transparently derivable." In support of this claim we have to remember that for Montague the level of *IL* is totally dispensable, its role being solely to increase perspicuity (see 6.1).

6.2.2. Intensionality and Meaning Postulates

The existence of nonextensional contexts has attracted the efforts of logicians and philosophers of language from Frege to Carnap, from Quine to Hintikka among many others, who have tried to provide an explanation for this widespread phenomenon of natural language. Nonextensional constructions are typically identified by the failure of the substitutability *salva veritate* of expressions having the same extensions. Among the core examples of nonextensional contexts we find the constructions containing epistemic verbs like *believe* and *know*, modal expressions like the adverb *necessarily*, verbs expressing intention like *look for*, *seek*, and *want*, temporal expressions, and so on. Since Frege, these constructions seem to challenge the universal validity of the principle of extensionality according to which the extension of every expression would depend only on the extension of its components. As we have said, Montague uses intensional logic and structures containing possible worlds and time indexes to provide a compositional semantics of natural language with the goal of accounting for the contexts in which extensionality failures occur. In *MG*, nonextensional contexts are treated as *intensional constructions*, that is, expressions in which the determination of the extension of the whole depends not simply on the extensions of the parts but on the intension of at least one of the parts. Interestingly, in *PTQ* intensionality is taken by Montague to represent the general case, in the sense that, rather than providing a special representation for the expressions giving rise to nonextensional contexts, he treats all basic grammatical relations as intensional, while the subset of extensional expressions is then singled out through meaning postulates. This choice, which is motivated by Montague's overall plan to formulate a translation procedure for an English fragment that has maximum level of generality, is visible in the type-assignment function (34): Every expression belonging to the functional category A/B or $A//B$[32] is assigned as type a function which takes as arguments the intensions of expressions belonging to category A. For instance, the basic expression *believe that* of category IV/S is assigned the type $\langle\langle s,t\rangle,\langle e,t\rangle\rangle$, where $\langle s,t\rangle$ is the type of *propositions*, that is, functions from world-time pairs to truth value. So in *PTQ believe that* is interpreted as a relation between an individual and a proposition,[33] and a sentence like (40a) is translated as (40b):

(40) a. John believes that Mary dreams.

b. $\text{believe}'(j, {}^\wedge\text{dream}'(m))$.

In *PTQ* Montague departs from the tradition and does not regard all intensional constructions as having the form of a propositional operator acting on some implicitly embedded propositional structure. Quine (1960) is a typical example of the received view: He explains the nonextensional character of *John seeks a unicorn* by rephrasing it as *John endeavours that he finds a unicorn.* Although Montague considers *seek* as being equivalent to *try to find*, he nevertheless claims that the intensionality of the former should not be explained by reducing it to the latter. Rather, intensionality is an inherent character of *seek* as a basic transitive expression. In fact, in *PTQ* all transitive verbs, being of category IV/T are assigned the type $f(\text{IV/T}) = \langle\langle s, f(\text{T})\rangle, f(\text{IV})\rangle = \langle\langle s, f(\text{S/IV})\rangle, f(\text{IV})\rangle = \langle\langle s,\langle\langle s,\langle e,t\rangle\rangle,t\rangle\rangle,\langle e,t\rangle\rangle$. Because terms are translated in *PTQ* as expressions of type $\langle\langle s,\langle e,t\rangle\rangle,t\rangle$, that is, sets of properties, then semantically a transitive verb denotes a relation between the intension of a set of properties (i.e., a property of properties or a second-order property) and an individual. This results in the following translation for the sentence *John seeks a unicorn* (in its *de dicto* reading):[34]

$$\text{seek}'(j, {}^\wedge\lambda P\exists x[\text{unicorn}'(x) \wedge {}^\vee P(x)]) \tag{41}$$

Similarly, Montague's translation regards as inherently intensional prepositions like *about* (see *John is talking about a unicorn*), intransitive verbs like *raise* or *change* and adverbs like *necessarily* or *allegedly.*

As we said, extensional expressions are captured by Montague by letting a set of formulas of *IL* play the role of *meaning postulates.* The terminology is from Carnap (1952) and their role is to restrict the class of possible models of *IL*. Carnap introduced them to explain analytical relations between lexical items and to overcome the shortcomings of the model of intensions and L-truth formulated in his *Meaning and Necessity* (1947). More generally, meaning postulates have come to be widely used in model-theoretic semantics as a powerful tool to represent relations about words meanings. In fact, one of the main features of Montague's model-theoretic semantics is its *lexical underspecification.* As we saw in 6.1, for Montague interpreting a language amounts to determine the type of reference of the different categories of its expressions. The interpretation procedure consists essentially in assigning, for instance, to the category of intransitive verbs the type of sets of entities, while nothing is said of the way in which specific members of this category, say, *eat* and *run*, differ semantically. Actually, making further distinctions about the semantic content of elements within the same categories is far beyond

Montague's aim, because as Marconi (1997, 10) remarks, he did not need to go any further, "for that was enough to make his point, namely, the availability of a formal method for the construction of a definition of truth for a language that met his own formal and material constraints." In *MG*, meaning postulates are the formal instrument with which finer-grained semantic distinctions within a given category of expressions can be expressed, and lexical properties of words are captured in terms of implications between propositions containing them.

Montague applies the method of meaning postulates to characterize the extensional character of certain expressions of English. For instance, (42) illustrates the postulate introduced in *PTQ* to capture extensional verbs, where S is a variable of type $\langle s, \langle e, \langle e, t\rangle\rangle\rangle$, and X is a variable for the intension of a set of properties:

(42) $\exists S \forall x \forall X (\delta(x, X) \leftrightarrow {}^{\vee}X({}^{\wedge}\lambda y {}^{\vee}S(x, y)))$, where $\delta = \text{love}'$, find', kiss', etc.

This postulate says that although the object of extensional verbs like *kiss*, *love*, and so on is semantically a second-order property, for each of these verbs there is an expression denoting a relation between two entities, to which it is equivalent. For instance, unlike *John seeks a unicorn*, the sentence *John finds a unicorn* implies the existences of unicorns, because the verb *find* is fully extensional:

(43) John finds a unicorn.

To preserve the compositional mapping between syntactic categories and semantic types, *find* is assigned the same type as *seek*, that is, $\langle\langle s, \langle\langle s, \langle e, t\rangle\rangle, t\rangle\rangle, \langle e, t\rangle\rangle$, which gives the flowing translation for (43):

(44) $$\text{find}'(j, {}^{\wedge}\lambda P \exists x[\text{unicorn}'(x) \wedge {}^{\vee}P(x)]).$$

However, since the interpretations of *IL* are restricted to those in which the meaning postulate (42) holds, then there is a relation between individuals, say find* of type $\langle s, \langle e, \langle e, t\rangle\rangle\rangle$, such that (44) is equivalent to (45):

(45) $${}^{\vee}{}^{\wedge}\lambda P \exists x[\text{unicorn}'(x) \wedge {}^{\vee}P(x)]({}^{\wedge}\lambda y {}^{\vee}\text{find*}(j, y)).$$

Applying the ${}^{\vee}{}^{\wedge}$-theorem and the λ-conversion, we obtain the logically equivalent (46), which actually means that there is an individual entity such that it is a unicorn and John finds it:

(46) $$\exists x[\text{unicorn}'(x) \wedge {}^{\vee}\text{find*}(j, x)].$$

Other meaning postulates in *PTQ* capture the extensional nature of the preposition *in*; the extensionality of intransitive verbs other than *rise* and *change*; the fact that proper names are rigid designators in the sense of Kripke (1972), that is, they denote the same individual in every possible world; the fact that verbs like *seek*, and *believes that*, which are intensional in their direct object, are nevertheless extensional in their subject position; the truth-conditional equivalence of *seek* and *try to find*, and so on.

Prima facie, meaning postulates provide an alternative to lexical decomposition to capture linguistically relevant aspects of word meaning. In fact, rather than assuming a set of noninterpreted primitive elements, such as the semantic markers in *KF*, meaning postulates allow for expressing inferences between lexical items in a fully model-theoretic fashion. However, introducing a meaning postulate is by itself not less ad hoc than introducing a certain conceptual primitive, and therefore meaning postulates are unable to achieve a real breakthrough with respect to word meaning analysis. In fact, the real challenge, for both meaning postulates and semantic decomposition, lies in the empirical issue of determining which aspects of lexical meaning are systematically relevant in the lexicon and active in affecting the linguistic behavior of lexical items. The problem is thus to motivate in a principled way the adoption of a given postulate or semantic primitive.

6.2.3. The Treatment of Scope

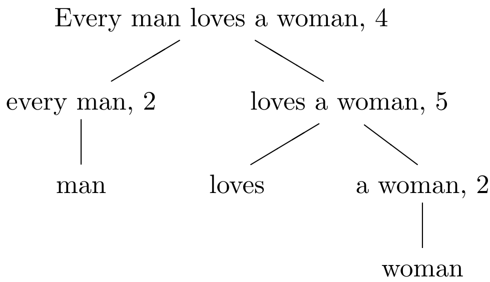
The claim that nonlexical ambiguities are syntactic ambiguities is one of the most important features of *MG*, and is a direct consequences of the principle of compositionality as defined by Montague. Since the syntax-semantics mapping is defined in terms of a homomorphism between algebras, every aspect of semantics that is not related to the interpretation of basic expressions must be traced back to a syntactic opposition. In other terms, every nonlexical ambiguity of a natural language expression must be explained by assigning to it more than one truth-conditionally distinct analysis tree. In fact, the input to semantic interpretation must be provided by a fully disambiguated syntax.

One of the novelties of *PTQ* with respect to the linguistic theory of its time is the way disambiguation is resolved in terms of the order of application of the syntactic rules as encoded in the derivational trees. A typical example is the treatment of *scope ambiguities*, as in the following sentence:

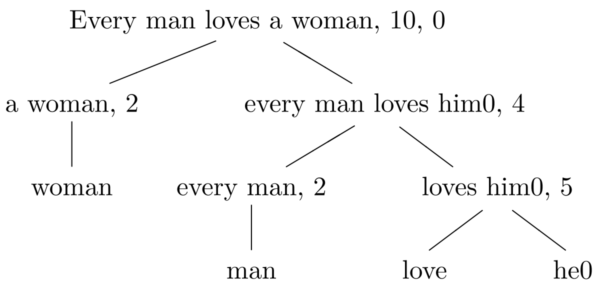
(47) Every man loves a woman.

Montague's ingenious solution to this problem is to see terms as entering syntactic composition "indirectly," through the process of replacement of a free variable in a sentence. As mentioned in 6.2, the category of terms include an infinite set of syntactic variables, he_0, $he_1, \ldots$, which are the only lexical "abstract" (i.e., not corresponding to actual English expressions) elements in *PTQ*. (S14), the most important of the three rules of quantification (also known as the quantifying in rules), combines a term T with a formula,[35] which, unless the rule applies vacuously, must be "open," that is, it contains one or more free variables. In this case, the first occurrence of the variable is replaced with T, and all the other occurrences of the same variable are replaced with an appropriate pronoun. As a consequence, the ambiguity of (47) is traced back to the alternative ways in which the sentence can be derived from its basic expressions, as shown here.

(48) a.



b.



In the translation rule (T14) associated to (S14), first the interpretation of the "open" sentence is lambda-abstracted over the variable, and then the interpretation of the term is applied to the intension of the lambda abstraction:

(49) If α is an expression of category T and φ is an expression of category t, and α and φ translate into an α' and φ' respectively, then $F_{10,n}(\alpha, \varphi)$ translates into $\alpha'({}^\wedge\lambda x_n \varphi')$.

Details aside, it can be proved that the top node in the analysis tree (48a) translates into

$$\lambda P \forall x[\text{man}'(x) \rightarrow {}^\vee P(x)]({}^\wedge \text{love}'(\lambda Q \exists z[\text{woman}(z) \wedge {}^\vee Q(z))]),$$

which after several conversions comes to be logically equivalent to

$$\forall x[\text{man}'(x) \rightarrow \exists z[\text{woman}'(z) \wedge \text{love*}(x, z)]]$$

(with love* the extensional variant of love$'$).

On the other hand, the top node of (48b) translates into

$$\lambda P \exists z[\text{woman}'(z) \wedge {}^\vee P(z)]({}^\wedge \lambda y \forall x[\text{man}'(x) \rightarrow \text{love*}(x, y)]),$$

which is logically equivalent to

$$\forall z[\text{woman}'(z) \wedge \forall x[\text{man}'(x) \rightarrow \text{love*}(x, z)]],$$

thereby giving the wide scope reading of the existential quantifier.

It is interesting to notice that the function F_{10} in (49) has an index n, ranging over natural numbers. In fact, the rules of quantification are rule schemas such that for every n there is a different rule instantiating them. Consequently, distancing himself from the principle of the autonomy of syntax, Montague claims that each declarative sentence of the fragment he is interested in, has infinitely many analysis trees, which are nevertheless "inessential," exactly because they do not amount to semantic differences. To see this, we have to remember that every term can be either interpreted "in situ" or introduced into a sentence via the rules of quantification and the use of the syntactic variables; in addition, basic expressions contain an infinite number of variables. Therefore, there are an infinite number of analysis trees which are syntactic variants of (48b), that is, they differ only with respect to the pronominal variable he_0. Moreover, there is another possible derivation for (47) on top of (48a) and (48b), one in which the object term *a woman* can also be composed by applying (S14). Similarly, for the sentences *John dreams* and *Every man dreams* the analyses reported in (37) are not the only possible ones. In fact in both cases the subject term can be introduced "indirectly" via the rule (S14). In both cases, the alternative analyses yield interpretations logically equivalent to the given ones.

Montague's analysis of quantification bears a strong resemblance to the one proposed in generative semantics based on the operation of quantifier lowering (section 5). However, it is important to stress that in contrast to generativist semantics, in *MG* quantifier scope is determined *in a purely derivational fashion*. In fact, the analysis tree is not properly to be regarded as a syntactic representational level in the sense of generative grammar, but rather as a way of keeping trace of the process of syntactic composition. Accordingly, the relative scope of quantifiers is defined in terms of the order of their introduction into the analysis tree, and not in terms of the geometry of the tree itself. Nevertheless, the two models gave quickly rise to convergent studies, as witnessed by a number of efforts to stress the synergies between them. For instance, Cooper and Parsons (1976) define a transformational grammar for English equivalent to *PTQ* which is very close to Lakoff's and McCawley's analyses of quantification. Similarly, Dowty (1979) combines *PTQ* with the lexical decomposition approach widely adopted in generative semantics.

Montague's original approach to the logical syntax of natural language has motivated a whole stream of research in model-theoretic semantics, which has widely enlarged Montague's original fragment and has addressed some of the open issues in *PTQ*. Some of the most active areas of application of Montague's method have been the interpretation of pronouns, adverbial quantification, verb aspect, and so on. These developments have often led to major changes in Montague's original solutions. A typical example is the analysis of "donkey sentences" in Kamp (1981), which has been the starting point of one of the most important and influential model-theoretic frameworks of logical semantics, *discourse representation theory*, which radically departs from *PTQ* under many respects. The work of Partee, Heim, Kratzer, and Chierchia among many others also represent important contributions to the logical investigation of natural language stemming from the Montagovian tradition. These have led to important developments of the original framework, while sticking to its spirit.

7. The Problem of Logical Form in Generative Linguistics

As stated in Chomsky (1981, 17), "At the most general level of description, the goal of a grammar is to express the association between representations of form and representations of meaning." In *MG* this association is resolved by defining the algebra of syntax as homomorphic to the algebra of meaning, thereby implementing the general constraint of compositionality to which Montague's system adheres. In the early period of generative linguistics, the form-meaning relation is determined through the Katz–Postal hypothesis, which assigns to deep syntactic structures the whole burden of providing the input to semantic interpretation. As we saw in section 5, generative semantics brings this principle up to the extreme consequence of overthrowing the assumption of the autonomy of syntax itself. The progressive implosion of the generative semantics movement has then run parallel to the exploration of new solutions for the form-meaning relation in grammar, with a twofold goal:

i. Reasserting Chomsky's principle of the autonomy of syntax, and

ii. Improving the grammar so that it can accommodate the phenomena (quantification) that had led generative semanticists to depart from this principle.

In the 1970s, the so-called *extended standard theory* revises the *Aspects* model by abandoning the Katz–Postal hypothesis and by proposing that both deep and surface syntactic structures contribute to the semantic interpretation of sentences (Jackendoff 1972). In particular, the deep structure would be responsible for those aspects of meaning concerning *thematic relations*, while structural aspects like quantification and anaphora would be established at the level of surface structure. Thematic relations include notions like agent, patient, goal, source, and so on and define the semantic roles of predicate arguments. Transformations like the passive do not actually alter these relations. For instance, John is the patient of the killing event in (50a), and does not change this role after the passive transformation:

(50) a. The car killed John.

b. John was killed by the bomb.

On the other side, transformations seem to affect pronominal coreference, and constitute in this way positive evidence for this relation to be marked at the level of the surface structure:

(51) a. John saw himself.
b. *Himself was seen by John.

Under the pressure to find solutions to this kind of phenomena, the generative paradigm underwent its most critical changes since its rising in the 1950s, which led to a huge reorganization of the architecture of grammar. In its early stages, the generative theory of grammar included a set of base phrase structure rules that generated deep structure representations. Then, transformational rules derived surface structure representations by moving some of the constituents, inserting lexical material or deleting some of the elements. Some of the main shortcomings of this model had their roots in the fact that the machinery of transformations was too powerful, the rules too loosely constrained, and they lacked generality. In its new developments, the generative paradigm tries to overcome these shortcomings by adopting a much more general and constrained description of the architecture of language. A new grammatical architecture is now proposed organized around the following modules:

1. A basic module that generates the constituent structures. It includes the principles of *X-bar theory* (Jackendoff 1972), which represents a major generalization and abstraction of phrase structure rules;
2. One single transformation or operation, *Move-*α, which moves elements from one position to another within the phrase markers system generated by the *X-bar* principles;
3. Principles and filters constraining the structures produced by the generative component together with *Move-*α. These principles are organized in several subsystems: bounding theory, government theory, theta-theory, binding theory, Case theory, and control theory.

The new architecture corresponds to the *Government and Binding* (*GB*) approach to the formal study of grammar (Chomsky 1981, 1982, 1986), and represents the mainstream version of the generative paradigm up to the minimalist turn that occurred in the 1990s. In the *GB* model, various syntactic constructions, like passive or relative clauses, are not projected into specific rules of the grammar, but are regarded rather as epiphenomenal distinctions to be analyzed and explained in terms of the interaction of the different principles of the modules of grammar. As a consequence, “The notions ‘passive’, ‘relativization’, etc., can be reconstructed as processes of a more general nature, with a functional role in grammar, but they are not ‘rules of grammar’.” (Chomsky 1981, 7).

One of the most interesting aspects of the *GB* model concerns the relation between syntax and meaning, and the way it has opened new important connections with model-theoretic semantics and with the tradition of logical grammar. The locus of meaning in *GB* depends on two main innovations that characterize this stage of the generative paradigm. The first one is the notion of *trace*, that is, an "empty" syntactic category that appears within syntactic representations as an effect of the application of the rule *Move-α*. In the previous versions of generative grammar, an element which at surface structure had to appear in a different position from the one occupied at deep structure was simply displaced by a specific transformation rule, leaving behind a gap in the original position, as in the case of the interrogative pronoun in (52):

(52) a. you see + PAST what.
b. what did you see.

Here *Move-α* moves a syntactic constituent to a new position in the syntactic representation, but the moved element leaves behind it a trace t coindexed with it:

(53) a. you see + PAST who.
b. who$_i$ did you see t_i.

Traces are syntactic elements voided of any phonological content, but carrying important information: the index of the moved element. According to Chomsky (1976), traces are like variables bound by the syntactic constituent with which they are coindexed.

The appearance of traces in syntactic theory has a number of important consequences. First of all, traces preserve the syntactic and semantic relations that obtained prior to movement. Thus, if in (53a) *who* is the theme argument of the verb, it keeps this role also after it has moved, via the coindexed trace t_i. This fact makes it substantially unnecessary to have both deep and surface structure as input to the semantic component (as assumed by the extended standard theory), since thematic relations are preserved also at surface structure. Second, with traces, syntactic variables appear on the scene of generative linguistics, thereby adding and important element of similarity with Montague's system. Third, the original notion of surface structure now disappears, at least in the way it was understood in the standard theory. In fact, the syntactic representations resulting from applications of *Move-α* contain traces that do not have any phonological content. This is the reason why in *GB* the syntactic representations derived from deep structures via *Move-α* are called *S-structures* (SS), a notational way to indicate they resemble and yet at the same time are distinct from surface structures.

The second major innovation of the *GB* model, closely interrelated with the former, is the appearance of *logical form* (*LF*) as a new and independent layer of syntactic representation, which replaces S-structure as the only interface with the semantic component. One of the main motivations for the introduction of *LF* in the architecture of generative grammar was the need to account for the behavior of interrogative pronouns and quantifiers in natural language which semantically behave as operators binding a variable occurring within their scope. For instance, identifying a question with its possible answers

(Karttunen 1977; Higginbotham 1983), allows for the possibility to take the logical form of (54a) to be (54b):

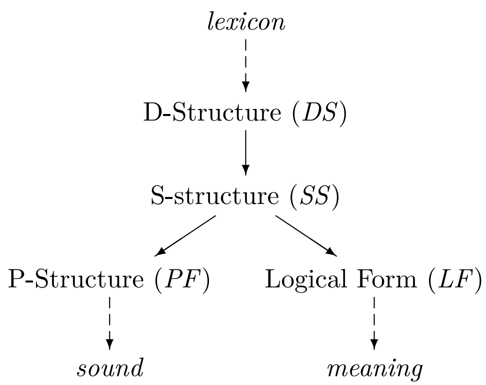
(54) a. Who did you see?
b. For which person x: you saw x.

Once again, this is a move made possible by associating (54a) to the S-structure (53b), making the manipulation mentioned, and then interpreted traces as bound variables. However, since the late 1970s various types of evidence have been brought to show that this could not be the whole story. For instance, in (55a) the pronoun *what* appears in the S-structure in its original position at the end of the clause (in English only one interrogative pronoun is allowed to appear at the left end of the clause). Nevertheless, semantically *what* behaves exactly like *who*, that is, as an operator with its scope over the whole sentence, as showed by (55b), which expresses the *LF* of the sentence:

(55) a. Who saw what?
b. For which person x and thing y: x saw y.

If syntactic representation must display the proper scopes of operators, then it follows that S-structures cannot perform this role. (See Williams 1977 and Fiengo 1977, for similar conclusions.) The conclusion that emerged was that if syntactic representations must provide the proper input to the sound and to the meaning systems, then this input cannot come from the same level of description, because there are elements that are interpreted as if they occurred at a different place from the one in which they are pronounced. This hypothesis has been incorporated into the theory starting from Chomsky and Lasnik (1977) and has a stable place in Chomsky (1981), who adopts the following architecture of grammar (i.e., the "T-model").

(56)



According to this proposal, each sentence is associated with four levels of *linguistic representation*. Lexical items enter into syntactic representations at *DS* (roughly equivalent to previous deep structures), prior to any transformation. *Move-α* links *DS* to *SS* by moving constituents within phrase markers, and leaving traces in the original place coindexed with the moved element. In addition, *SS* is related, on one hand to *PF* (phonetic form), which represents the interface with the phonological and phonetic component by providing the grammatical information relevant to sound assignment. On the other hand, further applications of *Move-α* generate *LF*, which May (1985, 2) defines as follows:

> [*LF*] represents whatever properties of syntactic form are relevant to semantic interpretation—those aspects of semantic structure which are expressed syntactically. Succinctly, the contribution of grammar to meaning.

LF is the level at which proper scope relations are assigned to operator-like elements (e.g., interrogative pronouns). For example, (55a) would be rendered at *LF* as:

(57) $[_S \text{ what}_j\ [_S \text{ who}_I\ [_S\ t_I \text{ saw } t_j]]]$.

Similarly, May (1977, 1985) proposes that quantificational *NP*s are assigned scope at *LF* through *Quantifier Raising* (*QR*), a particular instance of Move-α that "adjoins" the *NP* to a proper scope position in the phrase marker, typically the S node.[36] For instance, the *SS* of the sentence in (47) (repeated here as (58a)) is presented in (58b). Then two successive applications of *QR* produce two possible *LF* representations (58c) and (58d), depending on their order of application:

(58) a. Every man loves a woman.
b. $[_S \text{ every man } [_{VP} \text{ loves a woman}]]$.
c. $[_S \text{ every man}_I\ [_S \text{ a woman}_j\ [_S t_I\ [_{VP} \text{ loves } t_j]]]]$.
d. $[_S \text{ a woman}_I\ [_S \text{ every man}_j\ [_S t_I\ [_{VP} \text{ loves } t_j]]]]$.

The *scope* of a quantifier (and in general of any expression behaving as a variable-binding operator) is defined as a relation over syntactic representations:

(59) The *scope* of α is the set of nodes that α c-commands at *LF*.

In turn, c-command is defined as follows:[37]

(60) α *c-commands* β if and only if (i) the first branching node dominating α dominates β, and (ii) α does not dominate β.

A trace is bound by α if and only if it is within the scope of α and is coindexed with it. Therefore, in (58c)–(58d), both *NP*s have wider scope over the whole clause, since they c-command it, and the traces left by *QR* are the variables bound by the quantifiers.

In May (1977), this definition of scope also accounts for the structural ambiguity of (58a). Actually, the existence of two interpretations (wide and narrow scope of the existential quantifier) is explained by the fact that (58a) corresponds to two possible *LF*s, the first one in which the universal quantifier has wider scope than the existential one (since it c-commands it) and the second one in which the scope order is reversed. In a later version of his theory of quantification, May revises the definition of scope, by assuming that, given a *LF* of the form $[_S Q_1 [_S Q_2 [_S \ldots]]]$, the operators Q_1 and Q_2 belong to the same Σ-*sequence*. Then, May defines a general *scope principle*, such that the members of a Σ-sequence are free to take any ordering relation (May 1985, 34) Therefore, (58c) and (58d) are both compatible with the narrow and the wide scope interpretation of the existential quantifier, and also with a further reading in which the quantifiers are interpreted independently of one another, that is, as "branching quantifiers."[38]

It is interesting to compare the foregoing treatment of quantification in *GB* with Montague's analysis.

First of all, in *MG* the relative scope of quantifiers is the result of the order of their introduction in the compositional process, that is, the result of the derivational history as shown in the analysis tree (6.2.3). On the other side, according to May, the scope of quantifiers and quantifier-like elements is determined by the particular structure of the representations at the level of *LF*.

Second, Montague defines a rule of *Quantifying in* over terms, a general category including every type of *NP*: proper nouns, syntactic variables, quantificational expressions, and so on. Terms, as we have seen, have the type $\langle\langle s, \langle e, t\rangle\rangle, t\rangle$, and denote sets of properties. Consequently, Montague assigns the same derivation (and thus the same logical form) to *John dreams* and to *Every man dreams* (see (37a,b)), which has the effect that at the interpretive level, both are analyzed as the functional application of the denotation of the subject *NP* to the denotation of the *VP*. On the other hand, the analysis of *NP*s in generative grammar radically departs form such a model. Chomsky (1976, 198) explicitly criticizes Montague for blurring crucial syntactic and semantic differences within the set of *NP*s. Similarly, in May's theory, the two sentences would come to have two distinct *LF* representations:

(61) a. $[_S$ John $[_{VP}$ dreams$]]$.

b. $[_S$ Every man$_I$ $[_S t_I$ $[_{VP} t_I$ dreams$]]]$.

In fact, "at *LF*, quantified and nonquantified phrases are distinguished not only in their interpretation, but in those aspects of their syntax to which the rules of interpretation are sensitive" (May 1985, 25). While for Montague all the *NP*s are alike in their type (and consequently in their syntactic and semantic behavior), *GB* grammarians make a distinction, in the spirit of Frege, between *referential* and *quantificational expressions*. Referential expressions include proper names, pronouns, and in general the traces left by *Move-α*. Quantificational expressions include all the expressions that behave like operators binding variables (i.e., quantifiers, interrogative pronouns, etc.), and whose correct interpretation requires the assignment of scope. Since scope is now defined as a relation over *LF* syntactic representations, one of the consequences of this distinction is that the movement of quantificational expressions at *LF* becomes obligatory in *GB*. This also represents another crucial difference between May's theory of quantification and *MG*: In fact, for Montague the application of the rule *Quantifying in* is optional, and every term can roughly be interpreted "in situ." Conversely, in the *GB* analysis, *NP*s can be interpreted "in situ" (i.e., in the position in which they appear at SS) only if they are referential, while quantificational *NP*s must necessarily move to a position at which they are assigned a scope.

The appearance of *LF* in the generative theory of grammar has brought with it new ways to understand the relationship between formal linguistics, on one side, and model-theoretic semantics and logical grammar on the other:

> LF-Theory and model-theoretic semantics are very similar in some respects. Both are concerned with "structural meaning," abstracting away from word meaning and pragmatics, and both postulate a logical language as the representation of the "structural meaning" of sentences. (Riemsdijk and Williams 1986, 183)

In the early stages of the theory, the dominant semantic paradigm in generative linguistic, *KF* semantics (4.3), was essentially concerned with the explanation of phenomena like analyticity, entailment, synonymy, and so on, in terms of an "internalist" approach to word meanings based on their decomposition into conceptual primitives. *KF* rejected categorically talk of truth-conditions and reference as the ground for any semantic theory aiming to model linguistic competence. The introduction and the growing role of *LF* in the architecture of grammar radically changes this perspective, with the effect of leaving outside the domain of the theory of grammar all the notions that are beyond structural semantics. In particular, one of the goals of the generative enterprise is to study those aspects of meaning—such as scope ambiguities, anaphora—which depend on structural conditions. Moreover, as Riemsdijk and Williams (1986, 188) remark, *LF* basically models the same aspects of meaning represented in a "predicate calculus": "the scope of operators and quantifiers, sameness and distinctness of variables, and predicate-argument structure."

Nevertheless, *LF* differs from Montagovian logical syntax in different respects. First of all, the algebra of syntax in *MG* provides a disambiguated input to interpretation. In contrast, *LF* representations do not need to be semantically unambiguous. For instance, May's scope principle establishes that a *LF* can be totally underspecified with respect to the actual interpretation of the relation between the quantifiers. Second, although *LF* provides the proper structural information for the semantic interpretive component, it is still a layer of syntax, with the same status as *DS* or *SS*. *QR* is actually an instance of *Move-α*, the same movement operation that accounts for the displacement of constituents in passive sentences. As such, *QR* is subject to the same type of constraints that regulate the "overt" movement of constituents. In *MG*, a sentence is assigned those syntactic derivations that are necessary to explain its possible interpretations. In *GB*, a sentence is assigned those *LF* representations that are licensed by the independently motivated principles of grammar. It is therefore an empirical issue whether the principles of universal grammar are able to assign to a given sentence all and only all the *LF* representations that will correspond to its possible structural interpretations. *LF* is in fact a way to reassert the validity and centrality of the principle of the autonomy of syntax.

As a consequence, while Montague pursues his program within the general tradition of logical syntax, the introduction of *LF* in the architecture of grammar opens the way to a wide program that aims to convert logical semantics into syntax: As long as the distribution of linguistic phenomena like quantification, or bound anaphora can be explained on the ground of general, independently motivated principles of syntax, these phenomena can be regarded as part of a theory of syntax. A particularly radical version of this type of approach is given by Hornstein in *Logic as Grammar* (1984, 1):

> semantic theories of meaning for natural language—theories that exploit semantic notions such as "truth," "reference," "object," "property"—are not particularly attractive in explaining what speakers know when they know the meaning of a sentence in their native language. In short a theory of meaning that is viewed as responsive to the same set of concerns characteristic of work in generative grammar to a large extent will not be a semantic theory, ... but instead a syntactic one.

He concludes that "many of the phenomena earlier believed to be semantic ... are better characterized in syntactic terms" (ibid.), where syntactic refers now to the autonomous theory of universal grammar.

Actually, as a layer of syntactic representation *LF* is not directly committed to a particular interpretation. A theory of syntax for Chomsky is supposed to take one up to the point of specifying the structural information relevant for the interpretation of sentences. He does not, strictly speaking, take any position with respect to the nature and form of this interpretation. Although it is the only interface level with the interpretive module, *LF* is in fact an "uninterpreted" level of representation. This leaves the door open to new synergies between the generativist grammar and the truth-conditional semantics. One step in this direction is May (1985) who analyzes quantifiers truth-conditionally as *generalized quantifiers* in the sense of Barwise and Cooper (1981).

Similarly, Higginbotham (1985, 1986, 1989) argues for the possibility of pursuing Davidson's program in semantics by giving a recursive definition of truth for natural language using as input *LF* syntactic representations. We recall that Montague also intended his logical grammar to be a realization of Davidson's program. It then seems that this program is compatible with the principles of the generative enterprise. Chierchia (1995a, 1995b) represents interesting attempts to provide a model-theoretic analysis that pairs *GB* syntactic representations with type-theoretical semantic interpretations in the style of Montague. Both Higginbotham's and Chierchia's works, among many others, are indicative of the new dialectic that characterizes the most recent developments in generative linguistics and in the logical semantics tradition. A few decades ago the principles and methods arising from formal linguistics seemed to be radically orthogonal to the logical investigation of natural language, whereas the intense work and changes on both sides have allowed them to reach important and unprecedented convergences in the inquiry into the universal principles of language.

Notes

1. De Saussure introduced the notion of *phoneme*. As for syntax, he considered it as mostly belonging to *parole*, that is, not to the language as a system, but to language usage.

2. Bloomfield wrote on linguistics in *the International Encyclopaedia of Unified Science* (1939).

3. See our comments at the end of section 2.1.

4. A third level of adequacy is the *descriptive* one, which is intermediate between observative and explanatory adequacy. Descriptive adequacy is defined as follows: "the grammar gives a correct account of the linguistic intuition of the native speaker, and specifies the observed data (in particular) in terms of significant generalizations that express underlying regularities in the language" (Chomsky 1964, 63).

5. For instance, "In fact, the realization that this creative aspect of language is its essential characteristic can be traced back at least to the seventeenth century. Thus we find the Cartesian view that man alone is more than mere automatism, and that its is the possession of true language that is the primary indicator of this" (Chomsky 1964, 51).

6. -*en* is the past participle affix.

7. Some of these are context-sensitive rules, meant to account for the subcategorization properties of lexical items (e.g., *V* cannot be intransitive if it is followed by a *NP*, etc.)

8. Chomsky (1957), 15.

9. Particularly through the contributions in Fillmore (1968), Gruber (1976), and Jackendoff (1972).

10. The strongest and most definitive attack against the behaviorist view of language is in Chomsky (1959), which critically reviews B. F. Skinner's *Verbal Behavior*.

11. The output of the semantic component is actually formed by the set of readings that the projection rules can derive by all the possible senses that form the dictionary entry of the lexical items (by excluding, at the same time, those combinations that violate semantic selectional constraints).

12. For the chemical theory of concepts, see Coffa (1991).

13. The explicit target of Katz's critique is the definition of the domain of logic in Quine (1955).

14. See for instance Dowty (1979).

15. This characterization of the scope of operators in terms of command will be incorporated in the later stages of the Chomskian framework (see the definition of c-command in section 7).

16. The fuzziness of the notion of meaning adopted by generativist semanticists has also lead them to try to incorporate more and more phenomena within grammar, including presuppositions, speech acts, different sorts of pragmatic phenomena, and so on. Each step was also associated with the posit of more and more abstract deep structures and with the necessity of complex mechanism to derive the surface ones. For a history of this stage, see Newmeyer (1986).

17. "If it were necessary to choose between a categorial base that was convenient for semantics and a non-categorial base that was convenient for transformational syntax, I might still choose the former" (Lewis 1972, 22).

18. "It appears to me that the syntactic analyses of particular fragmentary languages that have been suggested by transformational grammarians, even if successful in correctly characterizing the declarative sentences of those languages, will prove to lack semantic relevance; and I fail to see any great interest in syntax except as a preliminary to semantics" (Montague 1974, 223).

19. See for instance Partee and Hendriks (1997).

20. As Partee and Hendriks (1997, 22) remark, in the rule-by-rule interpretation the homomorphism applies at the level of rules or derivation trees, not at the level of syntactic or semantic operations employed in the rules: "This is frequently a point of confusion.... But it is clear that while there my be a uniform compositional interpretation of the Subject-Predicate combining rule ..., there could not be expected to be a uniform semantic interpretation of a syntactic operation such as concatenation, a syntactic operation which may be common to many rules."

21. This actually amounts to a generalization of the technique employed in *EFL*, where eight semantic domains are individually defined.

22. See Montague (1974, 228, 258).

23. In *UG*, senses are instead the members of $D_{\langle s,a\rangle A,I}$, that is, functions of only one argument, regarded as a possible world.

24. In *PTQ*, quantified terms are introduced syncategorematically, in the sense that there is no syntactic category to which quantifiers and determiners are assigned, and they are rather introduced directly by the syntactic rule forming the term. The same holds true for conjunction and disjunction (see the following).

25. Notice that in *MG* there are only unary functions. Expressions denoting binary relations are of type $\langle e, \langle e, t\rangle\rangle$. The expression $\gamma(\alpha)(\beta)$ is then taken to assert that the objects denoted by β and α stand in the relation denoted by γ. In fact, as is well known, every binary function f from A into $\{1, 0\}$ is equivalent to the function g of type $\langle e, \langle e, t\rangle\rangle$ such that for every $x \in A$, $g(x)$ is the function of type $\langle e, t\rangle$ such that for every $y \in A$, $g(x)(y) = f(y, x)$.

26. It can be shown that for every expression α, $\check{}\hat{}\alpha$ is equivalent to α. However, it is not always the case that α is equivalent to $\hat{}\check{}\alpha$.

27. The argument is attributed by Montague to Barbara Hall Partee.

28. The usefulness of individual concepts has been again advocated from time to time, as for instance in Janssen (1984).

29. Notice also the Russellian treatment of definite descriptions in (35).

30. Montague's analysis is also claimed to have some empirical advantages, since it is then possible to give a straightforward representation of proper nouns when

they appear as elements in a conjunction with some quantified term, as in *John and a student*.

31. For a similar position see Gamut (1991).

32. Others than IV and CN, which are assigned the type $\langle e, t\rangle$, see 6.1.

33. Hintikka (1962, 1968) also analyzes *believe* as a relation between an individual and a proposition, the latter intended as set of possible worlds. Notice that, while for Hintikka (40a) is true if and only if the proposition expressed by the embedded sentence includes the set of the possible worlds *compatible* with John's beliefs, Montague does not set any specific constraint on the type of the relation denoted by *believe that*.

34. Montague's analysis of transitive intensional verbs has been deeply revised, and alternative solutions have been proposed by many scholars. Notice however that the *PTQ* translation of *seek* is actually able to account for interesting semantic properties of this verb. For instance, (41) is able to explain why the fact that John seeks a unicorn does not entail the existence of these animals. In fact, (41) is true even if the set of unicorns is empty in the real world. Moreover, we can also explain why from the fact that neither unicorns nor chimeras exist and that John seeks a unicorn we can not infer that John seeks a chimera. For further details, see Gamut (1991).

35. The other two rules of quantification defined in *PTQ*, (S15) and (S16), combine terms with expressions of category CN and IV, respectively, so that it is possible to quantify also over these types of expressions, besides sentences.

36. Given a constituent β and a node α in a phrase marker, adjoining β to α means to yield either a structure of the form $[_\alpha \beta [_\alpha \ldots]]$ (left adjunction) or a structure of the form $[_\alpha [_\alpha \ldots] \beta]$ (right adjunction).

37. See Reinhart (1976).

38. See Hintikka (1974), Barwise (1979).

References

Ajdukiewicz, K. 1935. Die syntaktische Konnexität. *Studia Philosophica* 1 (1935): 1–27. The references here are to the English translation (H. Weber), Syntactic connection. In *Polish Logic*, ed. Storrs McCall, 207–231. Oxford: Clarendon Press, 1967.

Bach, E. 1976. An extension of classical transformational grammar. *Problems of Linguistic Metatheory*. Proceedings of the 1976 Conference. Michigan State University.

Bach, E. 1989. *Informal Lectures on Formal Semantics*. Albany: State University of New York Press.

Bar-Hillel, Y. 1954. Logical syntax and semantics. *Language* 30: 230–237.

Bar-Hillel, Y. 1964. *Language and Information*. Palo Alto: Addison-Wesley.

Bar-Hillel, Y. 1970. *Aspects of Language*. Jerusalem: Magnes Press.

Bar-Hillel, Y., C. Gaifman, and E. Shamir. 1960. On categorial and phrase structure grammars. *Bulletin of the Research Council of Israel*, 1–16.

Barwise, J. 1979. On branching quantifiers in English. *Journal of Philosophical Logic* 8: 47–80.

Barwise, J., and R. Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4: 159–219.

Bennett, M. 1974. *Some Extensions of a Montague Fragment of English*. Bloomington: Indiana University Linguistics Club.

Bloomfield, L. 1933. *Language*. New York: Holt.

Bloomfield, L. 1939. *Linguistic Aspects of Science*. International Encyclopaedia of Unified Science, vol. 1.4. Chicago: University of Chicago Press.

Carnap, R. 1937. *Logical Syntax of Language*. New York and London.

Carnap, R. 1947. *Meaning and Necessity. A Study in Semantics and Modal Logic*. Chicago: University of Chicago Press.

Carnap, R. 1952. Meaning postulates. *Philosophical Studies* 3: 65–73.

Casadio, C. 1988. Semantic categories and the development of categorial grammars. In *Categorial Grammars and Natural Language Structures*, eds. Richard T. Oehrle et al., 95–123. Dordrecht: Reidel.

Chierchia, G. 1995a. *Dynamics of Meaning. Anaphora, Presupposition, and the Theory of Grammar*. Chicago: University of Chicago Press.

Chierchia, G. 1995b. Individual level predicates as inherent generics. In *The Generic Book*, eds. G. N. Carlson and F. J. Pelletier, 176–223. Chicago: University of Chicago Press.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. 1959. Review of Skinner, *Verbal Behavior*. *Language* 35: 26–58.

Chomsky, N. 1964. Current issues in linguistic theory. In *The Structure of Language: Readings in the Philosophy of Language*, eds. J. A. Fodor and J. J. Katz, 50–118. Englewood Cliffs, N.J.: Prentice Hall.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1976. Conditions on rules of grammar. *Linguistic Analysis* 2: 303–351. Reprinted in Chomsky, N. 1977. *Essays on Form and Interpretation*. New York: North-Holland.

Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1986. *Knowledge of Language, its Nature, Origin, and Use*. New York: Praeger.

Chomsky, N., and H. Lasnik. 1977. Filters and control. *Linguistic Inquiry* 8: 425–504.

Church, A. 1951. A formulation of the logic of sense and denotation. In *Essays in Honor of H. M. Sheffer*, eds. P. Henle, H. Kallen, and S. Langer. New York.

Coffa, J. A. 1991. *The Semantic Tradition from Kant to Carnap: To the Vienna Station*. Cambridge: Cambridge University Press.

Cooper, R., and T. Parsons. 1976. Montague grammar, generative semantics and interpretive semantics. In *Montague Grammar*, ed. B. H. Partee, 311–362. New York: Academic Press.

Cresswell, M. J. 1973. *Logics and Languages*. London: Methuen.

Davidson, D. 1970. Semantics for natural languages. In *Linguaggi nella Società e nella Tecnica*, ed. di Comunità. Milan. Reprinted in Davidson (1984), 55–64.

Davidson, D. 1984. *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press.

Davidson, D., and G. Harman, eds. 1972. *Semantics of Natural Language*. Dordrecht: Reidel.

Dowty, D. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.

Dowty, D., R. Wall, and S. Peters. 1981. *Introduction to Montague Semantics*. Dordrecht: Reidel.

Fiengo, R. W. 1977. On trace theory. *Linguistic Inquiry* 8: 35–61.

Fillmore, C. J. 1968. The case for case. In *Universals in Linguistic Theory*, eds. E. Bach and R. Harms, 1–90. New York: Holt, Rinehart & Winston.

Frege, G. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* Eng. *Concept Notation, a formal language of pure thought, modelled upon that of arithmetic.* Translated in *Conceptual Notation and Related Articles*, trans. and ed. T. W. Bynum. Oxford: Clarendon Press, 1972.

Frege, G. 1891. Function and concept. In Gottlob Frege: *Collected Papers on Mathematics, Logic, and Philosophy*, ed. Brian McGuinness, 137–156. Oxford: Basil Blackwell, 1984.

Frege, G. 1892. On sense and meaning. In Brian McGuinness, ed., 157–177.

Frege, G. 1892. On concept and object. In Brian McGuinness, ed., 182–194.

Frege, G. 1923. *Compound Thoughts.* In Brian McGuinness, ed., 390–406.

Gamut, L. T. F. 1991. *Logic, Language, and Meaning.* Vol 2: *Intensional Logic and Logical Grammar.* Chicago: University of Chicago Press.

Gruber, J. S. 1976. *Lexical Structures in Syntax and Semantics.* Amsterdam: North-Holland.

Harman, G. 1972. Deep structure as logical form. In Davidson and Harman (1972), 25–47.

Harris, Zellig. 1951. *Methods in Structural Linguistics.* Chicago.

Harris, Zelling. 1957. Co-occurrences and transformations in linguistic structure. *Language* 33: 283–340.

Higginbotham, J. 1983. Logical form, binding and nominals. *Linguistic Inquiry* 14: 395–420.

Higginbotham, J. 1985. On semantics. *Linguistic Inquiry* 16: 547–593.

Higginbotham, J. 1986. Linguistic theory and Davidson's program in semantics. In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, ed. E. LePore, 29–48. Oxford: Blackwell.

Higginbotham, J. 1989. Elucidations of meaning. *Linguistics and Philosophy* 12: 465–517.

Hintikka, J. 1962. *Knowledge and Belief.* Ithaca, N.Y.: Cornell University Press.

Hintikka, J. 1968. Semantics for propositional attitudes. In *Philosophical Logic*, eds. W. Davis et al., 21–45. Dordrecht: Reidel.

Hintikka, J. 1974. Quantifiers vs. quantification theory. *Linguistic Inquiry* 5: 154–177.

Hornstein, N. 1984. *Logic as Grammar.* Cambridge, Mass.: MIT Press.

Husserl, E. 1913. *Logische Unterscuchungen*, 2nd ed. Halle: Max Niemeyer.

Jackendoff, R. 1972. *Semantic Interpretation in Generative Grammar.* Cambridge, Mass.: MIT Press.

Janssen, T. M. V. 1984. Individual concepts are useful. In *Varieties of Formal Semantics*, eds. F. Landman, and F. Veltman, 171–192. Dordrecht: Foris.

Kamp, H. 1981. A theory of truth and semantic representation. In *Formal Methods in the Study of Language*, eds. J. Groenendjik, T. M. V. Janssen, and M. Stokhof, part 1, 277–322. Amsterdam: Mathematisch Centrum Tracts.

Kaplan, D. 1964. Foundations of Intensional Logic. PhD Dissertation. University of California at Los Angeles.

Karttunen, L. 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1: 3–44.

Katz, J. J. 1972. *Semantic Theory.* New York: Harper and Row.

Katz, J. J., and J. A. Fodor. 1963. The structure of a semantic theory. *Language* 39. Reprinted in *The Structure of Language: Readings in the Philosophy of Language*, eds. J. A. Fodor and J. J. Katz, 479–518. Englewood Cliffs, N.J.: Prentice Hall.

Katz, J. J., and P. M. Postal. 1964. *An Integrated Theory of Linguistic Descriptions*. Cambridge, Mass.: MIT Press.

Kripke, S. 1963. Semantical considerations on modal logic. *Acta Philosophica Fennica* 16.

Kripke, S. 1972. Naming and necessity. In Davidson and Harman (1972), 253–355.

Lakoff, G. 1970. *Irregularity in Syntax*. New York: Holt, Rinehart & Winston.

Lakoff, G. 1972. Linguistics and natural logic. In Davidson and Harman (1972), 545–665.

Lesńiewski, S. 1929. Grundzuge eines neuen Systems der Grundlagen der Mathematik. *Fundamenta Mathematica* 14: 1–81.

Lewis, D. 1972. General semantics. In Davidson and Harman (1972), 169–218.

Marconi, D. 1997. *Lexical Competence*. Cambridge, Mass.: MIT Press.

May, R. 1977. *The Grammar of Quantification*. Bloomington: Indiana University Linguistics Club.

May, R. 1985. *Logical Form: Its Structure and Derivation*. Cambridge, Mass.: MIT Press.

McCawley, J. D. 1972. A program for logic. In Davidson and Harman (1972), 498–544.

Montague, R. 1974. *Formal Philosophy. Selected Papers of Richard Montague*, edited and with an introduction by Richard Thomason. New Haven, Conn.: Yale University Press.

Newmeyer, F. J. 1986. *Linguistic Theory in America*. San Diego: Academic Press.

Partee, B. H., and H. L. W. Hendriks. 1997. Montague grammar. In *Handbook of Logic and Language*, eds. J. Benthem, and A. ter Meulen, 5–91. New York: Elsevier Science.

Quine, W. van O. 1953a. Two dogmas of empiricism. In Quine, W. van O., *From a Logical Point of View*, 20–46. Cambridge, Mass.: Harvard University Press.

Quine, W. van O. 1953b. The problem of meaning in linguistics. In Quine, W. van O., *From a Logical Point of View*, 47–64. Cambridge, Mass.: Harvard University Press.

Quine, W. van O. 1955. *Mathematical Logic*. Cambridge, Mass.: Harvard University Press.

Quine, W. van O. 1960. *Word and Object*. Cambridge, Mass.: MIT Press.

Quine, W. van O. 1961. Logic as a source of syntactical insights. *Proceedings of Symposia in Applied Mathematics* 12: 1–5.

Reinhart, T. 1976. The Syntactic Domain of Anaphora. PhD dissertation. Cambridge, Mass.: MIT.

Riemsdijk, H. van, and E. Williams. 1986. *Introduction to the Theory of Grammar*. Cambridge, Mass.: MIT Press.

Saussure, F. de. 1916. *Cours de linguistique générale*. Paris: Payot.

Tarski, A. 1936. Grundlegung der wissenschaftlichen Semantik. *Actes du Congrès International de Philosophie Scientifique*, vol. 3, 1–18. Paris.

Whitehead, A. N., and B. Russell. 1910, 1912, 1913. *Principia Mathematica*, 3 vols. Cambridge: Cambridge University Press.

Williams, E. 1977. Discourse and logical form. *Linguistic Inquiry* 8: 101–139.

17

Logic and Artificial Intelligence

Richmond H. Thomason

1. Logic and Artificial Intelligence

Artificial intelligence (which I'll refer to hereafter by its nickname, "AI") is the subfield of computer science devoted to developing programs that enable computers to display behavior that can (broadly) be characterized as intelligent. Most research in AI is devoted to fairly narrow applications, such as planning or speech-to-speech translation in limited, well-defined task domains. But substantial interest remains in the long-range goal of building generally intelligent, autonomous agents.[1]

Throughout its relatively short history, AI has been heavily influenced by logical ideas. AI exhibits a rather eclectic assortment of theories and research methodologies; the value and relative importance of logical formalisms is questioned by some leading practitioners and has been debated in the literature from time to time.[2] But most members of the AI community would agree that logic has an important role to play in at least some central areas of AI research, and an influential minority considers logic to be the most important factor in developing strategic, fundamental advances.

1.1. Guide to This Chapter

I imagine that the audience for this chapter will consist primarily of logicians and historians of logic who have little or no familiarity with AI. In writing this chapter, I have tried to provide an overview of the issues that arise when logic is used in helping to understand problems in intelligent reasoning and to guide the design of mechanized reasoning systems. Logic in AI is a large and rapidly growing field—I could not hope to achieve anything like complete coverage. In sections 3 and 4 I have tried to provide an overview with some historical and technical details concerning nonmonotonic logic and reasoning about action and change, a topic that is not only central in AI but that should be of considerable interest to philosophers. The remaining sections provide brief and inadequate sketches of selected topics, with references to the primary literature.

While this chapter was being written, Minker (2000b) appeared. This book is a comprehensive, up-to-date collection of survey papers and original contributions to the field of logic-based AI, with extensive references to the literature and with an introduction (to the book and to the field), Minker (2000a). I highly recommend this volume as a beginning point for any readers who wish to pursue this topic further.

1.2. The Role of Logic in Artificial Intelligence

Theoretical computer science developed out of logic, the theory of computation (if this is to be considered a different subject from logic), and some related areas of mathematics.[3] So theoretically minded computer scientists are well informed about logic even when they aren't logicians. Computer scientists in general are familiar with the idea that logic provides techniques for analyzing the inferential properties of languages, and with the distinction between a high-level logical analysis of a reasoning problem and its implementations. Logic, for instance, can provide a specification for a programming language by characterizing a mapping from programs to the computations that they license. A compiler that implements the language can be incomplete, or even unsound, as long as in some sense it approximates the logical specification. This makes it possible for the involvement of logic in AI applications to vary from relatively weak uses in which the logic informs the implementation process with analytic insights, to strong uses in which the implementation algorithm can be shown to be sound and complete. In some cases, a working system is inspired by ideas from logic, and acquires features that at first seem logically problematic but can later be explained by developing new ideas in logical theory. This sort of thing has happened, for instance, in logic programming.

In particular, logical theories in AI are independent from implementations. They can be used to provide insights into the reasoning problem without directly informing the implementation. Direct implementations of ideas from logic—theorem-proving and model-construction techniques—are used in AI, but the AI theorists who rely on logic to model their problem areas are free to use other implementation techniques as well. Thus, in R. C. Moore (1995, chapter 1), Robert C. Moore distinguishes three uses of logic in AI: as a tool of analysis, as a basis for knowledge representation, and as a programming language.

A large part of the effort of developing limited-objective reasoning systems goes into the management of large, complex bodies of declarative information.

It is generally recognized in AI that it is important to treat the representation of this information, and the reasoning that goes along with it, as a separate task, with its own research problems.

The evolution of expert systems illustrates the point. The earliest expert systems, such as MYCIN (a program that reasons about bacterial infections, see Buchanan and Shortliffe 1984), were based entirely on large systems of procedural rules, with no separate representation of the background knowledge—for instance, the taxonomy of the infectious organisms about which the system reasoned was not represented.

Later generation expert systems show a greater modularity in their design. A separate knowledge representation component is useful for software engineering purposes—it is much better to have a single representation of a general fact that can have many different uses, since this makes the system easier to develop and to modify. And this design turns out to be essential in enabling these systems to deliver explanations as well as mere conclusions.[4]

1.3. Knowledge Representation

In response to the need to design this declarative component, a subfield of AI known as *knowledge representation* (KR) emerged during the 1980s. Knowledge representation deals primarily with the representational and reasoning challenges of this separate component. The best place to get a feel for this subject is the proceedings of the meetings that are now held every other year: see Brachman et al. (1989), Allen et al. (1991), Nebel et al. (1992), Doyle et al. (1994), Aiello et al. (1996), Cohn et al. (1998), Cohn et al. (2000), and Fensel et al. (2002).

Typical articles in the proceedings of the KR and Reasoning conferences deal with the following topics.

1. Topics in logical theory and the theory of computation, including
   a. Nonmonotonic logic
   b. Complexity theory
2. Studies in application areas, including
   a. Temporal reasoning
   b. Formalisms for reasoning about planning, action, and change
   c. Metareasoning
   d. Reasoning about context
   e. Reasoning about values and desires
   f. Reasoning about the mental states of other agents, and especially about knowledge and belief
   g. Spatial reasoning
   h. Reasoning about vagueness

3. Studies in application techniques, including
   a. Logic programming
   b. Description logics
   c. Theorem proving
   d. Model construction
4. Studies of large-scale applications, including
   a. Cognitive robotics
   b. Merging, updating, and correcting knowledge bases

These topics hardly overlap at all with the contents of the *Journal of Symbolic Logic* (*JSL*), the principal research archive for mathematical logic. But there is substantial overlap in theoretical emphasis with the *Journal of Philosophical Logic* (*JPL*), where topics such as tense logic, epistemic logic, logical approaches to practical reasoning, belief change, and vagueness account for a large percentage of the contributions. Very few *JPL* publications, however, deal with complexity theory or with potential applications to automated reasoning.

1.4. Philosophical Logic

I do not know of a good history of philosophical logic. In fact, the distinction between mathematical and philosophical logic may well be incidental in relation to the overall goals of the subject, since technical rigor and the use of mathematical methods seem to be essential in all areas of logical research. However, the distinction between the two subfields has been magnified by differences in the sorts of professional training that are available to logicians, and by the views of individuals on what is important for the field. The statement of policy presented in volume 1, no. 1 of the *Journal of Symbolic Logic* (1936) lists bringing together the mathematicians and philosophers working in logic among the goals of the new journal. Probably at this time both the mathematicians and the philosophers shared a sense that their subject was considered to be somewhat marginal by their colleagues, and may have felt a primary loyalty to logic as a subject rather than to any academic discipline. Articles in the first volume of the *JSL* were divided about equally between professional mathematicians and philosophers, and the early volumes of the *JSL* do not show any strong differences between the two groups as to topic.

This situation changed in the 1960s. The 1969 volume of the *JSL* contained 39 articles by mathematicians, and only 9 by philosophers. By the early 1970s, many philosophers felt that philosophical papers on logic were unlikely to be accepted by the *JSL,* and that if they were accepted they were unlikely to be read by philosophers. At this point, the goals of the two groups had diverged considerably. Mathematicians were pursuing the development of an increasingly technical and complex body of methods and theorems. Many philosophers felt that this pursuit was increasingly irrelevant to the goal of illuminating philosophical issues. These divisions led to the founding of the *Journal of Philosophical Logic* in 1972. The list of sample topics in the first issue included:

1. Contributions to branches of logical theory directly related to philosophical concerns, such as inductive logic, modal logic, deontic logic, quantum logic, tense logic, free logic, logic of questions, logic of commands, logic of preference, logic of conditionals, many-valued logic, relevance logics;
2. Contributions to philosophical discussions that utilize the machinery of formal logic;
3. Discussions of philosophical issues relating to logic and the logical structure of language;
4. Philosophical work relating to the special sciences.

Most of the articles over the subsequent 28 years of the *JPL* belong to the first of these four categories. But the description with which this list begins is not particularly illuminating: Why should these particular topics be of interest to philosophers? I believe that the most important feature they share is a sense that despite successes in formalizing areas of mathematical logic, the scope of logic remained severely limited. There are unsolved problems in formalizing the nonmathematical sciences that seem to require thinking through new and different logical issues (quantum logic and the logic of induction, for instance). The remaining topics cover a part, at least, of the even more pressing problems involved in extending logical theory to nonscientific reasoning. The dominant goal, then, of philosophical logic is the extension of logical methods to nonmathematical reasoning domains. This goal has a theoretical dimension if (as many philosophical logicians seem to feel) it requires reworking and extending logical formalisms.

The development and testing of applications (applications such as the problem of formalizing the reasoning involved in getting to the airport, that was posed as a challenge in McCarthy 1959—see section 2.2) doesn't even appear as a category in the list of *JPL* topics, and in fact most of the philosophical logic literature is theoretical.

1.5. Logic in AI and Philosophical Logic

The rough comparison in section 1.3 of the contents of the main publications for research in logical AI and philosophical logic suggests the following picture. Theoretical work in logical AI and in philosophical logic overlap to a large extent. Both are interested in developing nonmetamathematical applications of logic, and the core topics are very similar. This overlap is due not only to commonality of interest but to direct influence of philosophical logic on logical AI; there is ample evidence, as we will see, that the first generation at least of AI logicists read and were influenced by the literature in philosophical logic.

Since that point, the specialties have diverged. New logical theories have emerged in logical AI (nonmonotonic logic is the most important example) which are familiar only to a subcommunity of the philosophical logicians. Other differences are due to the AI community's interest in the theoretical analysis of algorithms and, of course, to their sense of the importance of implementations. Some have to do with the emerging development in computer science of ambitious applications using unprecedentedly large bodies of logical axioms. The sheer size of these applications produces new problems and new methodologies. And other differences originate in the interest of philosophical logicians in some topics (metaphysical topics, for instance) that are primarily inspired by purely philosophical considerations.

Concern for applications can be a great influence on how research is carried out and presented. The tradition in philosophical logic predates applications in automated reasoning, and to this day remains relatively uninterested in such applications. The methodology depends on intuitions, but operates without any generally accepted methodology for articulating and deploying these intuitions. And the ideas are illustrated and informed by artificial, small-scale examples.[5] In general, the philosophical literature does not deal with implementability or efficiency of the reasoning, or indeed with any features of the reasoning process. It is hard to find cases in which the philosophical theories are illustrated or tested with realistic, large-scale reasoning problems.

These differences, however, are much more a matter of style than of substance or of strategic research goals. It is difficult to think through the details of the reasoning process without the computational tools to make the process concrete, and difficult to develop large-scale formalizations of reasoning problems without computational tools for entering, testing, and maintaining the formalizations. Because the core theoretical topics (modal, conditional, and temporal logic, belief revision, and the logic of context) are so similar, and because the ultimate goal (the formalization of nonmathematical reasoning) is the same, I think of logic in AI as a continuous extension of the philosophical logic tradition.

The early influence of philosophical logic on logic in AI was profound. The bibliography of McCarthy and Hayes (1969), one of the most influential early papers in logical AI, illustrates the point well. There are 58 citations in the bibliography. Of these, 35 refer to the philosophical logic literature. (There are 17 computer science citations, 1 mathematical logic citation, 1 economics citation, and 1 psychology citation.) This paper was written at a time when there were hardly any references to logical AI in the computer science literature. Naturally, as logical AI has matured and developed as a branch of computer science, the proportion of cross-disciplinary citations has decreased. A sampling of articles from the first Knowledge Representation conference, Brachman et al. (1989), held in 1989, shows only 12 philosophical logic citations out of a total of 522 sampled citations; a sampling of articles from Cohn et al. (1998) shows 23 philosophical logic citations out of a total of 468 sampled.[6]

Despite the dramatic decrease in quantity of explicit citations, the contemporary literature in logical AI reflects an indirect acquaintance with the earlier literature in philosophical logic, since many of the computational papers that are explicitly cited in the modern works were influenced by this literature. Of course, the influence becomes increasingly distant as time passes, and this trend is accelerated by the fact that new theoretical topics have been invented in logical AI that were at best only dimly prefigured in the philosophical literature.

Although philosophical logic is now a relatively small field in comparison to logical AI, it remains a viable area of research, with new work appearing regularly. But references to contemporary research in philosophical logic are rare in the AI literature. Similarly, the papers currently published in the *Journal of Philosophical Logic*, at least, do not show much influence from AI.[7] In Europe, the lines are harder to draw between professional divisions among logicians: some European journals, especially the *Journal of Logic, Language, and Information*, are successful in maintaining a focus in logic while attracting authors from all the disciplines in which logic is represented.

1.6. The Role of Artificial Intelligence in Logic

The importance of applications in logical AI, and the scale of these applications, represents a new methodology for logic—one that would have been impossible without mechanized reasoning. This methodology forces theoreticians to think through problems on a new scale and at a new level of detail, and this in turn has a profound effect on the resulting theories. The effects of this methodology will be illustrated in the following sections, dealing with various topics in logical AI. But the point is illustrated well by reasoning about action and change. This topic was investigated in the philosophical literature. Reasoning about change, at least, is part of tense logic, and the consequences of actions are investigated in the literature on "seeing to it that"; see, for instance, Belnap (1996). The latter theory has no very robust account of action. The central construct is a variation on a branching-time modality of the sort that has been familiar since Prior (1967). Although it represents an interesting development in philosophical logic, the scale of the accomplishment is very different from the research tradition in logical AI reported in section 4. The formalisms in this tradition not only support the formalization of complex, realistic planning problems, but provide entirely new insights into reasoning about the causal effects of actions, the persistence of states, and the interactions between actions and continuous physical processes. Developments such as this would have been impossible without the interactions between the logical theories and large-scale, practical applications in automated planning.

Rudolf Carnap (1955) attempted to clarify intensional analyses of linguistic meaning, and to justify them from a methodological point of view, by imagining how the analysis could be applied to the linguistic usage of a hypothetical robot. Carnap hoped that the fact that we could imagine ourselves to know the internal structure of the robot would help make the case for an empirical science of semantics more plausible. This hope proved to be unjustified; the philosophical issue that concerned Carnap remains controversial to this day, and thought experiments with robots have not proved to be particularly rewarding in addressing it. Real robots, though, with real applications,[8] are a very different matter. Though it is hard to tell whether they will prove to be helpful in clarifying fundamental philosophical problems, they provide a laboratory for logic that is revolutionary in its potential impact on the subject. They motivate the development of entirely new logical theories that I believe will prove to be as important for philosophy as the fundamental developments of the late nineteenth century proved to be.

The emergence of separate mathematical and philosophical subspecialties within logic was not an entirely healthy thing for the field. The process of making mathematical logic rigorous and demonstrating the usefulness of the techniques in achieving mathematical ends that was pursued so successfully in the first half of the twentieth century represents a coherent refinement of logical methodology. All logicians should be pleased and proud that logic is now an area with a body of results and problems that is as substantial and challenging as those associated with most areas of mathematics.

But these methodological advances were gained at the expense of coverage. In the final analysis, logic deals with reasoning—and relatively little of the reasoning we do is mathematical, while almost all of the mathematical reasoning that nonmathematicians do is mere calculation. To have both rigor and scope, logic needs to keep its mathematical and its philosophical side united in a single discipline. In recent years, neither the mathematical nor the philosophical professions—and this is especially true in the United States—have done a great deal to promote this unity. But the needs of computer science provide strong unifying motives. The professional standards for logical research in computer science certainly require rigor, but the field also puts its practitioners into contact with reasoning domains that are not strictly mathematical, and creates needs for innovative logical theorizing.

The most innovative and ambitious area of computer science, in terms of its coverage of reasoning, and the one that is closest in spirit to philosophical logic, is AI. This chapter provides an introduction, for outsiders who are familiar with logic, to the aspects of AI that are closest to the philosophical logic tradition. This area of logic deserves, and urgently needs, to be studied by historians. But I am not a historian, and this document does not pretend to be a history.

2. John McCarthy and Commonsense Logicism

2.1. Logical AI

The most influential figure in logical AI is John McCarthy. McCarthy is one of the founders of AI, and has consistently advocated a research methodology that uses logical techniques to formalize the reasoning problems that AI needs to solve.[9] All but the most recent work in McCarthy's research program can be found in Lifschitz (1990a), which also contains an introduction to McCarthy's work, and Lifschitz (1990b); for additional historical background, see Israel (1991).

McCarthy's methodological position has not changed substantially since it was first articulated in 1959 and elaborated and amended in McCarthy and Hayes (1969). The motivation for using logic is that—even if the eventual implementations do not directly and simply use logical reasoning techniques like theorem proving—a logical formalization helps us understand the reasoning problem itself. The claim is that without an understanding of what the reasoning problems are, it will not be possible to implement their solutions. Plausible as this Platonic argument may seem, it is in fact controversial in the context of AI; an alternative methodology would seek to learn or evolve the desired behaviors. The representations and reasoning that this methodology would produce might well be too complex to characterize or to understand at a conceptual level.

From McCarthy and Hayes (1969), it is clear that McCarthy thinks of his methodology for AI as overlapping to a large extent with traditional philosophy, but adding to it the need to inform the design of programs capable of manifesting general intelligence. This idea is not uncongenial to some philosophers (see, for instance, Carnap 1956, 244–247, and Pollock 1995), and I personally believe that logical AI is potentially of great value for philosophy. In practice, however, the actual theories that have emerged from McCarthy's methodology are influenced most strongly by work in philosophical logic, and the research tradition in logical AI represents a more or less direct development of this work, with some changes in emphasis. This review concentrates on logical AI in relation to philosophical logic, without further comment on relations to philosophy in general or to the feasibility of developing human-level intelligent systems.

2.2. The Formalization of Common Sense

McCarthy's long-term objective is to formalize *commonsense reasoning*, the prescientific reasoning that is used in dealing with everyday problems. An early example of such a problem, mentioned in McCarthy (1959), is getting from home to the airport. Other examples include:

1. *Narrative understanding.* The reasoning involved in reconstructing implicit information from narratives, such as sequencing of eventualities, and inferred causal connections.

2. *Diagnosis.* For instance, detecting faults in physical devices.

3. *Spatial reasoning.* For instance, reasoning about the parts of rigid bodies and their shapes, and their relation to the shape of the whole.

4. *Reasoning about the attitudes of other agents.* For instance, making informed guesses about the beliefs and desires of other agents, not from "keyhole observation" but from conversational clues of the sort that could be obtained in a brief, interactive interview.

Stated baldly, the goal of formalizing common sense would probably seem outrageous to most philosophers, who are trained to think of common sense as rather elusive. But whether or not the ultimate goal is appropriate and achievable, the specific formalization projects that have emerged from this program have been successful in several ways. They have succeeded in breaking new territory for logic by extending the scope of the reasoning problems to which logical techniques can be successfully applied. They have demonstrated that logical techniques can be an important part of the solutions to specific AI problems—planning is the most successful of these, but some success has been achieved in other areas as well.[10] They form the basis of one approach to developing complete, autonomous agents.[11] And they have illuminated many specific forms of nonscientific reasoning—for instance, qualitative reasoning about the behavior of physical devices.[12]

3. Nonmonotonic Reasoning and Nonmonotonic Logics

3.1. Nonmonotonicity

Aristotle believed that most reasoning, including reasoning about what to do and about sublunary natural phenomena, dealt with things that hold "always or for the most part." But Aristotelian logic deals only with patterns of inference that hold without exception. We find at the very beginning of logic a discrepancy between the scope of logical theory and commonsense reasoning. Nonmonotonic logic is the first sustained attempt within logical theory to remedy this discrepancy. As such, it represents a potential for a sweeping expansion of the scope of logic, as well as a significant body of technical results.

The consequence relations of classical logics are monotonic. That is, if a set Γ of formulas implies a consequence C then a larger set $\Gamma \cup \{A\}$ will also imply C. A logic is *nonmonotonic* if its consequence relation lacks this property. *Preferred models* provide a general way to induce a nonmonotonic consequence relation. Invoke a function that for each Γ produces a subset $\mathcal{M}_\Gamma$ of the models of Γ; in general, we will expect $\mathcal{M}_\Gamma$ to be a proper subset of the models of Γ. We then say that Γ implies C if C is satisfied by every model in $\mathcal{M}_\Gamma$. As long as we do not suppose that $\mathcal{M}_{\Gamma \cup \{A\}} \subseteq \mathcal{M}_\Gamma$, we can easily have an implication relation between Γ and C without imposing this relation on supersets of Γ.[13]

This model theoretic behavior corresponds to expectation-guided reasoning, where the expectations allow certain cases to be neglected. Here is an important difference between common sense and mathematics. Mathematicians are trained to reject a proof by cases unless the cases exhaust all the possibilities; but typical instances of commonsense reasoning neglect some alternatives. In fact, it is reasonable to routinely ignore outlandish possibilities. Standing in my kitchen in California, wondering if I have time to wash my dishes before leaving for work, I do not take the possibility of an earthquake into account.

There seem to be many legitimate reasons for neglecting certain cases in commonsense reasoning. A qualitative judgment that the probability of a case is negligible is one reason. But, for instance, in a planning context it may be reasonable to ignore even nonnegligible probabilities, as long as there is no practical point in planning on these cases.

The motivations for nonmonotonicity seem to involve a number of complex factors; probability (perhaps in some qualitative sense), normality, expectations that are reasonable in the sense that one can't be reasonably blamed for having them, mutual acceptance, and factors having to do with limited rationality. As far as I know, no one has succeeded in disentangling and clarifying these motivating considerations. In the early stages of its emergence in logical AI, many researchers seem to have thought of nonmonotonic reasoning as a general method for reasoning about uncertainty; but by the end of the 1980s, implementations of fully quantitative probabilistic reasoning were not only possible in principle, but were clearly preferable in many sorts of applications to methods involving nonmonotonic logic. A plausible and realistic rationale for nonmonotonic logic has to fit it into a broader picture of reasoning about uncertainty that also includes probabilistic reasoning.[14]

3.2. Historical Motivations

Three influential papers on nonmonotonic logic appeared in 1980: McCarthy (1980), McDermott and Doyle (1980), and Reiter (1980). In each case, the formalisms presented in these papers were the result of a gestation period of several years or more. To set out the historical influences accurately, it would be necessary to interview the authors, and this I have not done. However, there seem to have been two motivating factors: strategic considerations having to do with the long-range goals of AI, and much more specific, tactical considerations arising from the analysis of the reasoning systems that were being deployed in the 1970s.

Section 2.2 drew attention to McCarthy's proposed goal of formalizing common sense reasoning. The brief discussion in section 3.1 suggests that monotonicity may be an obstacle in pursuing this goal. An additional motive was found in Minsky (1974), which was widely read at the time. This paper presents an assortment of challenges for AI, focusing at the outset on the problem of natural language understanding.[15] Minsky advocates frame-based knowledge representation techniques,[16] and (conceiving of the use of these representations as an alternative to logic) he throws out a number of loosely connected challenges for the logical approach, including the problem of building large-scale representations, of reasoning efficiently, of representing control knowledge, and of providing for the flexible revision of defeasible beliefs. In retrospect, I think most AI researchers would agree that these problems are general challenges to any research program in AI (including the one Minsky himself advocated at the time) and that logical techniques are an important element in addressing some, perhaps all, of the issues. (For instance, a well-structured logical design can be a great help in scaling up knowledge representation.)

Minsky apparently intended to provide a general argument against logical methods in AI, but McCarthy (1980) and McDermott and Doyle (1980) interpret it as a challenge that can be met by developing logics that lack the monotonicity property. Perhaps unintentionally, Minsky's paper seems to have provided some incentive to the nonmonotonic logicians by stressing monotonicity as a source of the alleged shortcomings of logic. In fact, the term "monotonicity" apparently makes its first appearance in print in this paper.

The development of nonmonotonic logic also owes a great deal to the applied side of AI. In fact, the need for a nonmonotonic analysis of a number of AI applications was as persuasive as the strategic considerations urged by McCarthy, and in many ways more influential on the shape of the formalisms that emerged. Here, I will mention three such applications that appear to have been important for some of the early nonmonotonic logicians: belief revision, closed-world reasoning, and planning.

Belief Revision Doyle (1979) provides an analysis and algorithm for a "truth maintenance system." The TMS answers a general need, providing a mechanism for updating the "beliefs" of knowledge bases. The idea of the TMS is to keep track of the support of beliefs, and to use the record of these support dependencies when it is necessary to revise beliefs. In a TMS, part of the support for a belief can consist in the *absence* of some other belief. This introduces nonmonotonicity. For instance, it provides for defaults; that Wednesday is the default day for scheduling a meeting means the belief that the meeting will be on Wednesday depends on the *absence* of the belief that it will not be on Wednesday.

The TMS algorithm and its refinements had a significant impact on AI applications, and this created the need for a logical analysis. (In even fairly simple cases, it can be hard in the absence of analytic tools to see what consequences a TMS should deliver.) This provided a natural and highly specific challenge for those seeking to develop a nonmonotonic logic. The TMS also provided specific intuitions: The idea that the key to nonmonotonicity has to do with inferences based on unprovability was important for the modal approaches to nonmonotonic logic and for default logic. And the TMS's emphasis on interactions between arguments began a theme in nonmonotonic logic that remains important to this day. (See the discussion of argument-based approaches, in section 3.4.)

Closed-World Reasoning The study of databases belongs to computer science rather than to AI. But one of the research paradigms in the scientific analysis of databases uses logical models of the representations and reasoning (see Minker 1997 for a recent survey of the field), and this area has interacted often with logical AI. The deductive database paradigm was taking shape at about the same time that many AI researchers were thinking through the problems of nonmonotonic logic, and provided several specific examples of nonmonotonic reasoning that called for analyses. Of these, perhaps the most important, is the *closed-world assumption,* according to which—at least as far as simple facts are concerned, represented in the database as positive or negative literals—the system assumes that it knows all that there is to be known. It is the closed-world assumption that justifies a negative answer to a query "Is there a direct flight from Detroit to Bologna?" when the system finds no such flight in its data. This is another case of inference from the absence of a proof; a negative is proved, in effect, by the failure of a systematic attempt to prove the positive. This idea, which was investigated in papers such as Clark (1978) and Reiter (1978), also provided a challenge for nonmonotonic logics, as well as specific intuitions—note that again, the idea of inference rules depending on the absence of a proof is present here.

Planning The need for inertial defaults in temporal reasoning—defaults to the effect that things will stay put in the absence of a reason for them to change—arises in attempting to formalize the reasoning needed in planning. This application (the apparent need for a nonmonotonic logic in developing an economical formal solution to the frame problem) provided another specific formal need. One of the earliest attempts to formalize nonmonotonic reasoning, Sandewall (1972), addresses this problem. Inertial defaults are an especially important and instructive case study; I will say no more about them here, because they are discussed in detail in section 4.4.

3.3. The Earliest Formalisms

The three 1980 papers mentioned at the beginning of section 3.2 represent three approaches to nonmonotonic logic that remain important subfields to this day: *circumscription* (McCarthy), *modal approaches* (Doyle and McDermott), and *default logic* (Reiter).

In (1993a), McCarthy urges us, when considering the early history of circumscription, to take into account a group of three papers (McCarthy 1987, 1980, 1986). The first paper connects the strategic ideas of McCarthy and Hayes (1969) with the need for a nonmonotonic logic, and sketches the logical ideas of *domain circumscription*, which is now classified as the simplest case of circumscription. The second paper provides more thorough logical foundations, and introduces the more general and powerful *predicate circumscription* approach. The third paper concentrates on developing techniques for formalizing challenging commonsense examples.

All forms of circumscription involve restricting attention to models in which certain sets are minimized; for this reason, circumscription can be grouped with the preferred models approaches to nonmonotonicity; see section 3.4. McCarthy's formalism is fairly conservative; though it raises interesting logical issues in higher order logic and complexity, it uses familiar logical frameworks. And much of the focus is on the development of formalization techniques. The other varieties of nonmonotonic logic, including default logic and the modal nonmonotonic logics, raise issues of the sort that are familiar to philosophical logicians, having to do with the design of new logics, the systematic investigation of questions concerning validity, and managing the proliferation of logics.

As the discussion of truth maintenance indicated, it is very natural to think of nonmonotonic inferences as being *hedged*. That is, a nonmonotonic inference may require not merely the presence of a set of proved conclusions, but the *absence* of certain other conclusions. The general form of such a rule is:

DR In the presence of $\{A_1, \ldots, A_n\}$
and in the absence of $\{B_1, \ldots, B_n\}$,
conclude C.

An important special case of **DR** is a *normal default*, a simple rule to the effect that C holds by default, conditionally on A. This can be formalized by taking the condition that must be absent to simply be the negation of the conclusion.

NDR In the presence of $\{A_1, \ldots, A_n\}$
and in the absence of $\neg C$,
conclude C.

At first sight, it is somewhat perplexing how to formalize this notion of nonmonotonic inference, since it seems to require a circular definition of provability that can't be replaced with an inductive definition, as in the nonmonotonic case. The difficulty with the early theory of Sandewall (1972) is that it does not address this difficulty successfully. McDermott and Doyle (1980) and Reiter (1980) use fixpoint definitions to solve the problem. In both cases, the logical task is (1) to develop a formalism in which rules like **DR** can be expressed, and (2) to define the relation between a theory DT (which may incorporate such rules) and the theories E which could count as reasonable consequences of DT. In the terminology that later became standard, we need to define the relation between a theory DT and its *extensions*.

In retrospect, we can identify two sorts of approaches to nonmonotonic logic: those based on *preference* and those based on *conflict*. Theories of the first sort (like circumscription) involve a relatively straightforward modification of the ordinary model-theoretic definition of logical consequence that takes into account a preference relation over models. Theories of the second sort (like default logic) involve a more radical rethinking of logical consequence. The possibility of multiple extensions—different possible coherent, inferentially complete conclusion sets that can be drawn from a single set of premises—means that we have to think of logical consequence not as a function taking a set of axioms into its logical closure but as a *relation* between a set of axioms and alternative logical closures. Because logical consequence is so fundamental, this represents a major theoretical departure. With multiple extensions, we can still retrieve a consequence relation between a theory and a formula in various ways, the simplest being to say that DT nonmonotonically implies C if C is a member of every extension of DT. Still, the conflict-based account of consequence provides a much richer underlying structure than the preferential one.

Reiter approaches the formalization problem conservatively. Nonmonotonicity is not expressed in the language of default logic, which is the same as the language of first-order logic. But a theory may involve a set of *default rules*—rules of the form **DR**. Reiter (1980) provides a fixpoint definition of the extensions of such a theory, and develops the theoretical groundwork for the approach, proving a number of the basic theorems.

Of these theorems, I mention one in particular, which will be used in section 4.5, in connection with the Yale shooting anomaly. The idea is to take a conjectured extension (which will be a set T^*) and use this set for consistency checks in a proof-like process that applies default rules in $\langle W, D\rangle$ successively to stages that begin with W.

We define a default proof process $T_0, T_1, \ldots$ for W, D, relative to T^*, as follows.

> Let $T_0 = W$.
>
> If no default rule in D is nonvacuously applicable to T_i relative to T^*, then $T_{i+1} = \mathrm{Th}_{\mathrm{FOL}}(T_i)$.
>
> Otherwise, choose some default rule
>
> $$\frac{A : B_1, \ldots, B_n}{C}$$
>
> that is nonvacuously applicable to T_i relative to T^*, and let
>
> $$T_{i+1} = \mathrm{Th}_{\mathrm{FOL}}(T_i \cup \{C\}).$$

In other words, as long as we can nonvacuously close the stage we are working on under an applicable default, we do so; otherwise, we do nothing.

A theorem of Reiter's says that under these circumstances:

> T is an extension of $\langle W, D\rangle$ if and only if there is a proof process $T_0, T_1, \ldots$ for W, D, relative to T, such that
>
> $$T = \bigcup_{i=0}^{\infty} T_i.$$

Thus, we can show that T is an extension by (1) using T for consistency checks in a default reasoning process from $\langle W, D\rangle$, (2) taking the limit T' of this process, and (3) verifying that in fact $T' = T$.

The modal approach represents a "higher level of nonmonotonic involvement" than default logic. The unprovability construct is represented explicitly in the language, by means of a modal operator L informally interpreted as "provable" (or, as in McDermott and Doyle 1980, by the dual of this operator).[17] Although McDermott and Doyle's terminology is different from Reiter's, the logical ideas are very similar—the essence of their approach, like Reiter's, is a fixpoint definition of the extensions of a nonmonotonic logic. Incorporating nonmonoticity in the object language creates some additional complexities, which in the early modal approach show up mainly in proliferation of the logics and difficulties in evaluating the merits of the alternatives. As better foundations for the modal approach emerged, it became possible to prove the expected theorems concerning equivalence of modal formalisms with default logic.[18]

Reiter's paper (1980) appears to have developed primarily out of tactical considerations. The earlier paper (1978) is largely concerned with providing an account of database queries. Unlike the other seminal papers in nonmonotonic logic, Reiter's shows specific influence from the earlier and independent work on nonmonotonicity in logic programming—the work seems to have been largely inspired by the need to provide logical foundations for the nonmonotonic reasoning found in deductive databases. Doyle and McDermott's paper shows both strategic and tactical motivation—citing the earlier literature in logicist AI, it motivates nonmonotonic logic as part of a program of modeling commonsense rationality. But the theory is also clearly influenced by the need to provide a formal account of truth maintenance.

3.4. Approaches to Nonmonotonic Logic

Nonmonotonic logic is a complex, robust research field. Providing a survey of the subject is made difficult by the fact that there are many different foundational paradigms for formalizing nonmonotonic reasoning, and the relations between these paradigms is not simple. An adequate account of even a significant part of the field requires something like a book-length treatment. A number of books are available, including Antoniou (1997), Besnard (1992), Brewka (1991), Brewka et al. (1997), Lukaszewicz (1990), Marek and Truszczyński (1994), and Schlechta (1997). Two collections are especially useful: Ginsberg (1987) and Gabbay et al. (1994). The former is a useful source for readers interested in the early history of the subject and has an excellent introduction. The handbook chapters in Gabbay et al. (1994) provide overviews of important topics and approaches. My current recommendation for readers interested in a quick, readable introduction to the topic would be Brewka et al. (1997) and self-selected chapters of Gabbay et al. (1994). I will rely on these references for technical background and concentrate on intellectual motivation, basic ideas, and potential long-term significance for logic.

Preference Semantics At the outset in section 3.1, I mentioned how preferred models could be used to characterize a nonmonotonic consequence relation. This general model theory of nonmonotonicity emerged in Shoham (1988) five years after the work discussed in section 3.2, and represents a much more general and abstract approach.

Preferential semantics rely on a function S taking a set K of models into a subset $S(K)$ of K. The crucial definition of *preferential entailment* stipulates that A is a (nonmonotonic) consequence of Γ if every model M of $S([\![\Gamma]\!])$ implies A. Shoham's theory is based on a partial order $\preceq$ over models: $S(K)$ can then be characterized as the set of models in K that are $\preceq$-minimal in K. To ensure that no set can preferentially entail a contradiction unless it classically entails a contradiction, infinite descending $\preceq$ chains need to be disallowed.

This treatment of nonmonotonicity is similar to the earlier modal semantic theories of conditionals—the similarities are particularly evident using the more general theories of conditional semantics, such as the one presented in Chellas (1975). Of course, the consequence relation of the classical conditional logics is monotonic, and conditional semantics uses possible worlds, not models. But the *left-nonmonotonicity* of conditionals (the fact that $A \mathrel{\Box\!\!\rightarrow} C$ does not imply $[A \wedge B] \mathrel{\Box\!\!\rightarrow} C$) creates issues that parallel those in nonmonotonic logics. Early work in nonmonotonic logic does not seem to be aware of the analogy with conditional logic. But the interrelations between the two have become an important theme more recently; see, for instance, Alcourrón (1995), Arlo-Costa and Shapiro (1992), Asher (1995), Benferat et al. (1997), Boutilier (1992), Delgrande (1998), Gabbay (1995), Gärdenfors and Makinson (1994), and Pearl (1994).

Preference semantics raises an opportunity for formulating and proving representation theorems relating conditions over preference relations to properties of the abstract consequence relation. This line of investigation began with Lehmann and Magidor (1992).

Modal and Epistemic Theories Neither Doyle or McDermott pursued the modal approach much beyond the initial stages of McDermott and Doyle (1980) and McDermott (1982). With a helpful suggestion from Robert Stalnaker (see Stalnaker 1993), however, Robert C. Moore produced a modal theory that improves in many ways on the earlier ideas. Moore gives the modal operator of his system an epistemic interpretation, stressing the interpretation of a default rule that licenses a conclusion for a reasoning agent unless something that the agent knows blocks the conclusion. In Moore's *autoepistemic logic*, an extension E of a theory T is a superset of T that is *stable,* that is, deductively closed, and that satisfies the following two rules:

(1) if $A \in E$ then $\Box A \in E$;

(2) if $A \notin E$ then $\neg\Box A \in E$;

It is also usual to impose a *groundedness* condition on autoepistemic extensions of T, ensuring that every member of an extension has some reason tracing back to T. Various such conditions have been considered; the simplest one restricts extensions to those satisfying

(3) E is the set of nonmodal consequences of
$T \cup \{ A : \Box A \in E \} \cup \{ \neg\Box A : A \notin E \}$.

Autoepistemic logic remains a popular approach to nonmonotonic logic, in part because of its usefulness in providing theoretical foundations for logic programming. For more recent references, see Antoniou (1997), Konolige (1994), Marek and Truszczyński (1989, 1991), and R. C. Moore (1993, 1995).

Epistemic logic has inspired other approaches to nonmonotonic logic. Like other modal theories of nonmonotonicity, these use modality to reflect consistency in the object language, and so allow default rules along the lines of **DR** to be expressed. But instead of consistency, these use *ignorance*. See Halpern and Moses (1985) and Levesque (1987) for variations on this idea. These theories are explained, and compared to other nonmonotonic logics, in Meyer and van der Hoek (1995). In more recent work, Levesque's ideas are systematically presented and applied to the theory of knowledge bases in Levesque and Lakemeyer (2000).

3.5. Further Topics

This brief historical introduction to nonmonotonic logic leaves untouched a number of general topics that might well be of interest to even a nonspecialist. These include graph-based and proof-theoretic approaches to nonmonotonic logic, results that interrelate the various formalisms, complexity results, tractable special cases of nonmonotonic reasoning, relations between nonmonotonic and abductive reasoning, relations to probability logics, the logical intuitions and apparent patterns of validity underlying nonmonotonic logics, and the techniques used to formalize domains using nonmonotonic logics. For these and other topics I have to refer the reader to the literature. As a start, I highly recommend the chapters in Gabbay et al. (1994).

4. Reasoning about Action and Change

4.1. Priorian Tense Logic

Time and temporal reasoning have been associated with logic since the origins of scientific logic with Aristotle.[19] The idea of a logic of tense in the modern sense has been familiar since at least the work of Jan Łukasiewicz (see, for instance, Łukasiewicz 1970), but the shape of what is commonly known as tense logic was standardized by Arthur Prior's work in the 1950s and 1960s (see Prior 1956, 1967, 1968).[20] As the topic was developed in philosophical logic, tense logic proved to be a species of modal logic; Prior's work was heavily influenced by both Hintikka and Kripke, and by the idea that the truth of tense-logical formulas is relative to *world-states* or temporal stages of the world; these are the tense-theoretic analogs of the timeless possible worlds of ordinary modal logic. Thus, the central logical problems and techniques of tense logic were borrowed from modal logic. For instance, it became a research theme to work out the relations between axiomatic systems and the corresponding model theoretic constraints on temporal orderings. See, for instance, Burgess (1984) and van Benthem (1983).

Priorian tense logic shares with modal logic a technical concentration on issues that arise from using the first-order theory of relations to explain the logical phenomena, an expectation that the important temporal operators will be quantifiers over world-states, and a rather distant and foundational approach to actual specimens of temporal reasoning. Of course, these temporal logics do yield validities, such as

$$A \rightarrow PFA$$

(if A, then it was the case that A was going to be the case), which certainly are intuitively valid. But at most, these can only play a broadly foundational role in accounting for realistic reasoning about time. It is hard to think of realistic examples in which they play a leading part.

This characteristic, of course, is one that modal logic shares with most traditional and modern logical theories; the connection with everyday reasoning is rather weak. Although modern logical techniques do account with some success for the reasoning involved in verifying mathematical proofs and logic puzzles, they do not explain other cases of technical or commonsense reasoning with much detail or plausibility. Even in cases like legal reasoning, where logicians and logically minded legal theorists have put much effort into formalizing the reasoning, the utility of the results is controversial.

4.2. Planning Problems and the Situation Calculus

Planning problems provide one of the most fruitful showcases for combining logical analysis with AI applications. On the one hand, there are many practically important applications of automated planning, and on the other hand, logical formalizations of planning are genuinely helpful in understanding the problems and in designing algorithms.

The classical representation of an AI planning problem, as described in Amarel (1968), evidently originates in early work of Herbert Simon, published in a 1966 CMU technical report (Simon 1966). In such a problem, an agent in an initial world-state is equipped with a set of *actions,* which are thought of as partial functions transforming world-states into world-states. Actions are feasible only in world-states that meet certain constraints (these constraints are now called the "preconditions" of the action). A planning problem then becomes a search for a series of feasible actions that successively transform the initial world-state into a desired world-state.

The *situation calculus*, developed by John McCarthy, is the origin of most of the later work in formalizing reasoning about action and change. It was first described in 1969, in McCarthy (1983); the earliest generally accessible publication on the topic is McCarthy and Hayes (1969).

Apparently, Priorian tense logic had no influence on Amarel (1968). But there is no important difference between Amarel's world-states and those of Priorian tense logic. The "situations" of the situation calculus are these same world-states, under a new name.[21] They resemble possible worlds in modal logic in providing abstract locations that support a consistent and complete collection of truths. As in tense logic, these locations are ordered, and change is represented by the variation in truths from one location to another. The crucial difference between the situation calculus and tense logic is that change in the situation is *dynamic*—changes do not merely occur, but occur for a reason.

This difference, of course, is inspired by the intended use of the situation calculus: It is meant to formalize Simon's representation of the planning problem, in which a single agent reasons about the scenarios in which a series of actions is performed.[22] In this model, what drives change is the performance of actions, so the fundamental model theoretic relation is the relation

$$\textsc{Result}(\mathrm{a}, \mathrm{s}, \mathrm{s}')$$

between an action a, an initial situation s in which a is performed, and a resulting situation s' immediately subsequent to the performance of the action. Usually (though this is not absolutely necessary) the deterministic assumption is made that s' is unique. In general, actions can be successfully performed only under certain limited circumstances. This could be modeled by allowing for cases in which there is no s' such that $\textsc{Result}(\mathrm{a}, \mathrm{s}, \mathrm{s}')$. But usually, it is assumed that Result is in fact a total function, but that in cases in which s does not meet the "preconditions" of a, there are no restrictions on the s' satisfying $\textsc{Result}(\mathrm{a}, \mathrm{s}, \mathrm{s}')$, so that the causal effects of a will be entirely unconstrained in such cases.

A planning problem starts with a limited repertoire of actions (where sets of preconditions and effects are associated with each action), an initial situation, and a goal (which can be treated as a formula). Solving such a problem is a matter of finding a sequence of actions that will achieve the goal, given the initial situation. That is, given a goal G and initial situation s, the problem will consist of finding a sequence $\mathrm{a}_1, \ldots, \mathrm{a}_n$ of actions that will transform s into a final situation that satisfies G. This means (assuming that Result is a function) that G will be satisfied by the situation s_n, where $\mathrm{s}_0 = \mathrm{s}$ and s_{i+1} is the s' such that $\textsc{Result}(\mathrm{a}_{i+1}, \mathrm{s}_i, \mathrm{s}')$. The planning problem is in effect a search for a sequence of actions meeting these conditions. The success conditions for the search can be characterized in a formalism like the situation calculus, which allows information about the results of actions to be expressed.

Nothing has been said up to now about the actual language of the situation calculus. The crucial thing is how change is to be expressed. With tense logic in mind, it would be natural to invoke a modality like $[a]A$, with the truth condition

$$\models_{\mathrm{s}} [a]A \text{ iff } \models_{\mathrm{s}'} A, \text{ where } \textsc{Result}(\mathrm{a}, \mathrm{s}, \mathrm{s}').$$

This formalization, in the style of dynamic logic, is in fact a leading candidate; see section 4.7.

But McCarthy and Hayes (1969) deploy a language that is much closer to first-order logic. (This formalization style is characteristic of McCarthy's work; see McCarthy 1979.) Actions are treated as individuals. And certain propositions whose truth values can change over time (propositional *fluents*) are also treated as individuals. Where s is a situation and f is a fluent, $\mathit{Holds}(\mathrm{f}, \mathrm{s})$ says that f is true in s.

4.3. Formalizing Microworlds

Since the pioneering work of the nineteenth- and early twentieth-century logicians, the process of formalizing mathematical domains has largely become a matter of routine. Although (as with set theory) there may be controversies about what axioms and logical infrastructure best serve to formalize an area of mathematics, the methods of formalization and the criteria for evaluating them are relatively unproblematic. This methodological clarity has not been successfully extended to other domains; even the formalization of the empirical sciences presents difficult problems that have not yet been resolved.[23]

The formalization of commonsense reasoning presents an extreme with respect to such methodological difficulties. The work in logical AI has not converged successfully on a solution to this problem. But it has provided the idea of formalizing *microworlds* that represent limited domains of knowledge and reasoning, and work on formalizing these domains has provided some instructive case studies. In addition, there are a few projects that strive for more extensive coverage, as well as some useful methodological ideas. An adequate study of this work would take up a great deal of space. Here, I only mention some topics and provide some references to the literature.

Temporal reasoning, and in particular reasoning about actions and plans, is the best-developed domain. At least one important methodology will emerge in section 4.5: the development of a library of *scenarios* for testing the adequacy of various formalisms, as the creation of specialized domains like the blocks-world domain (mentioned in section 4.2) that serve a laboratories for testing ideas. For more on the blocks world, see Davis (1991) and Genesereth and Nilsson (1987). McCarthy's ideas about *elaboration tolerance* (McCarthy 1999) provide one interesting attempt to provide a criterion for the adequacy of formalizations. Still other important ideas have emerged in the course of formalizing commonsense domains. One is the importance of an explicit *ontology*; see, for instance, Lenat and Guha (1989) and Fikes (1996). Another is the potential usefulness of explicit representations of context; see Guha (1991). Finally, Davis (1991) provides many extended examples of formalizations of commonsense domains.

4.4. Prediction and the Frame Problem

To tell whether a plan achieves its goal, you need to see whether the goal holds in the plan's final state. Doing this requires *predictive* reasoning, a type of reasoning that was, as far as I know, entirely neglected in the tense-logical literature. As in mechanics, prediction involves the inference of later states from earlier ones. But (in the case of simple planning problems at least) the dynamics are determined by actions rather than by differential equations. The investigation of this qualitative form of temporal reasoning, and of related sorts of reasoning (e.g., plan recognition, which seeks to infer goals from observed actions, and narrative explanation, which seeks to fill in implicit information in a temporal narrative) is one of the most impressive chapters in the brief history of commonsense logicism.

The essence of prediction is the problem of inferring what holds in the situation that ensues from performing an action, given information about the initial situation. I will assume that the agent has complete knowledge about the initial situation—this assumption is usual in classical formalizations of planning.[24]

A large part of the qualitative dynamics that is needed for planning consists in inferring what does *not* change. Take a simple plan to type the word "cat" using word processing software: My plan is to first enter "c," then enter "a," then enter "t." Part of my confidence in this plan is that the actions are independent—for instance, entering "a" does not also erase the "c." The required inference can be thought of as a form of *inertia*. The *frame problem* is the problem of how to formalize the required inertial reasoning.

The frame problem was named and introduced in McCarthy and Hayes (1969). Unlike most of the philosophically interesting technical problems to emerge in AI, it has attracted the interest of philosophers; most of the relevant papers, and background information, can be found in Pylyshyn (1987) and Ford and Pylyshyn (1996). Both of these volumes document interactions between AI and philosophy.

The quality of these interactions is discouraging; as a philosopher, I even find it somewhat embarrassing. Like any realistic commonsense reasoning problem, the frame problem is open-ended, and can depend on a wide variety of circumstances. If I put $20 in my wallet and go to the store with the wallet in my pocket, I can safely assume that the $20 is still in my wallet. If I leave the $20 on the counter at the store while shopping, I can't safely assume it will be there when I get back. This may account for the temptation that makes some philosophers[25] want to construe the frame problem very broadly, so that very soon it becomes indiscernible from the problem of formalizing general common sense in arbitrary domains. Such a broad construal may serve to introduce speculative discussions concerning the nature of AI, but it loses all contact with the genuine, new logical problems in temporal reasoning that have been discovered by the AI community. It provides a forum for repeating some familiar philosophical themes, but it brings nothing new to philosophy.

I find this approach disappointing because I believe that philosophy can use all the help it can get, and that the AI community has succeeded in extending and enriching the application of logic to commonsense reasoning in dramatic ways that are highly relevant to philosophy. The clearest account of these developments to be found in the volumes edited by Pylyshyn is Morgenstern (1996). A recent extended treatment can be found in Shanahan (1997); also see Sandewall (1994).

The purely logical frame problem can be solved using monotonic logic, by simply writing explicit axioms stating what does *not* change when an action is performed. This technique can be successfully applied to quite complex formalization problems.[26] But *nonmonotonic* solutions to the framework have been extensively investigated and deployed; these lead to new and interesting lines of logical development.

Some philosophers (Fodor 1987; Lormand 1996) have felt that contrived propositions will pose special difficulties in connection with the frame problem. As Shanahan points out (Shanahan 1997, 24), Fodor's "fridgeon" example is readily formalized in the situation calculus and poses no special problems. However, as Lormand suggests, Goodman's examples (Goodman 1946) do create problems if they are admitted as fluents; there will be anomalous extensions in which objects change from green to blue to preserve their grueness.

This is one of the few points that I can find in the philosophical literature on the frame problem that raises a genuine difficulty for the formal solutions. But the difficulty is peripheral, since the example is not realistic. Recall that fluents are represented as first-order individuals. Although fluents are situation-dependent functions, an axiom of comprehension is certainly not assumed for fluents. In fact, it is generally supposed that the domain of fluents will be a very limited set of the totality of situation-dependent functions; typically, it will be a relatively small finite set of important variables, and will be chosen in particular cases much as a set of variables is chosen in statistical modeling.

I know of no systematic account in the AI literature of how to choose an appropriate set of fluents, but it would certainly be part of such an account that all fluents should correspond to projectable predicates, in Goodman's sense.

4.5. Nonmonotonic Treatments of Inertia and a Package of Problems

The idea behind nonmonotonic solutions to the frame problem is to treat inertia as a default; changes are assumed to occur only if there is some special reason for them to occur. In an action-centered account of change, this means that absence of change is inferred when an action is performed unless a reason for the change can be found in axioms for the action.

For explicitness, I use Reiter's default logic to illustrate the formalization. Recall that in Reiter's theory, defaults are represented as rules, not formulas, so that they are not subject to quantification. To formalize inertia, then, we need to use default rule schemata. For each fluent f, action a, and situation s, the set of these schemata will include an instance of the following schema:

$$\mathrm{IR}(\mathrm{f},\mathrm{a},\mathrm{s}) \quad \frac{\mathsf{T} : \mathit{Holds}(\mathrm{f},\mathrm{s}) \leftrightarrow \mathit{Holds}(\mathrm{f},\textsc{Result}(\mathrm{a},\mathrm{s}))}{\mathit{Holds}(\mathrm{f},\mathrm{s}) \leftrightarrow \mathit{Holds}(\mathrm{f},\textsc{Result}(\mathrm{a},\mathrm{s}))}.$$

This way of doing things makes any case in which a fluent changes truth value a prima facie anomaly. But it follows from Reiter's account of extensions that such defaults are overridden when they conflict with the monotonic theory of situation dynamics. So if, for instance, there is a monotonic causal axiom for the action `blacken` ensuring that blackening a block will make it black in the resulting situation, then the appropriate instance of **IR** will be inefficacious, and there will be no extension in which a white block remains white when it is blackened.

The frame problem somehow managed to capture the attention of a wide community—but if one is interested in understanding the complex problems that arise in generalizing formalisms like the situation calculus, while at the same time ensuring that they deliver plausible solutions to a wide variety of scenarios, it is more useful to consider a larger range of problems. For the AI community, the larger problems include the frame problem itself, the qualification problem, the ramification problem, generalizability along a number of important dimensions including incomplete information, concurrency (multiple agents), and continuous change, and finally a large assortment of specific challenges such as the scenarios mentioned later in this section.

The *qualification problem* arises generally in connection with the formalization of commonsense generalizations. Typically, these involve exceptions, and these exceptions—especially if one is willing to entertain far-fetched circumstances—can iterate endlessly. The same phenomenon, under labels like "the problem of ceteris paribus generalizations," is familiar from analytic philosophy. It also comes up in the semantics of *generic* constructions found in natural languages.[27] In a sense, this problem is addressed at a general level by nonmonotonic logics, which—though they do not provide a way to enumerate exceptions—do allow commonsense generalizations to be formulated as defaults, as well as enabling further qualifications to be added nondestructively. Ideally, then, the initial generalization can be stated as an axiom and qualifications can be added incrementally in the form of further axioms.

The qualification problem was raised in McCarthy (1986), where it was motivated chiefly by generalizations concerning the consequences of actions; McCarthy considers in some detail the generalization that turning the ignition key in an automobile will start the car. Much the same point, in fact, can be made about virtually any action, including stacking one block on another—the standard action that is used to illustrate the situation calculus. A circumscriptive approach to the qualification problem is presented in Lifschitz (1987); this explicitly introduces the precondition relation between an action and its preconditions into the formalism, and circumscriptively minimizes preconditions, eliminating from preferred models any "unknown preconditions" that might render an action inefficacious.

Several dimensions of the qualification problem remain as broad, challenging research problems. For one thing, not every nonmonotonic logic provides graceful mechanisms for qualification. Default logic, for instance, does not deliver the intuitively desired conclusions. Suppose one formalizes the common sense generalization that if I press the "a" key on my computer it will type "a" as a normal default:

$$\frac{\mathsf{T} : \textsc{Value}(\texttt{text}, \textsc{Result}(\texttt{Press-a}, \mathrm{s})) = \textsc{Value}(\texttt{text}, \mathrm{s}) + \text{“a”}}{\textsc{Value}(\texttt{text}, \textsc{Result}(\texttt{Press-a}, \mathrm{s})) = \textsc{Value}(\texttt{text}, \mathrm{s}) + \text{“a”}}.$$

If I then formalize the exception to this generalization that if I press the "a" key while the Alt key is depressed the cursor moves to the beginning of the current sentence as a similar normal default, I get *two* extensions: one in which pressing "a" while the Alt key is depressed adds "a" to the text and another in which it moves the cursor. The problem is that default logic does not provide for more specific defaults to override ones that are more general. This principle of specificity has been discussed at length in the literature. Incorporating it in a nonmonotonic logic can complicate the theory considerably; see, for instance, Asher and Morreau (1991) and Horty (1994). As Elkan (1995) points out, the qualification problem raises computational issues.

Relatively little attention has been given to the qualification problem for characterizing actions, in comparison with other problems in temporal reasoning. In particular, the standard accounts of *unsuccessful* actions are somewhat unintuitive. In the formalization of Lifschitz (1987), for instance, actions with some unsatisfied preconditions are distinguished from actions whose preconditions all succeed only by the fact that the conventional effects of the action are ensured when the preconditions are met. It is as if an action of spending $1,000,000 can be performed at any moment—although if you don't have the money, no effects in particular will be guaranteed.[28] And there is no distinction between actions that cannot even be attempted (like boarding a plane in London when you are in Sydney), actions that can be attempted but in which the attempt can be expected to go wrong (like making a withdrawal when you have insufficient funds), actions that can be attempted with reasonable hope of success, and actions that can be attempted with guaranteed success.

As J. L. Austin made clear (1961), the ways in which actions can be attempted, and in which attempted actions can fail, are a well-developed part of commonsense reasoning. Obviously, in contemplating a plan containing actions that may fail, one may need to reason about the consequences of failure. Formalizing the pathology of actions, providing a systematic theory of ways in which actions and the plans that contain them can go wrong, would be a useful addition to planning formalisms, and one that would illuminate important themes in philosophy.

The challenge posed by the *ramification problem* (characterized first in Finger 1987) is to formalize the indirect consequences of actions, where "indirect" effects are not delayed,[29] but are temporally immediate and causally derivative. If I walk into a room, the direct effect is that I am now in the room. There are also many indirect effects: For instance, my shirt also is now in the room. You can see from this that the formulation of the problem presupposes a distinction between direct consequences of actions (ones that attach directly to an action, and that are ensured by the successful performance of the action) and other consequences. This assumption is generally accepted without question in the AI literature on action formalisms. You can make a good case for its commonsense plausibility—for instance, many of our words for actions ("to warm," to "lengthen," "to ensure") are derived from the effects that are conventionally associated with them. And in these cases, success is entailed: If someone has warmed something, this entails that it became warm.[30] A typical example is discussed in Lin (1995): a certain suitcase has two locks, and is open if and only if both locks are open. Then (assuming that actions are not performed concurrently) opening one lock will open the suitcase if and only if the other lock is open. Here, opening a lock is an action, with direct consequences; opening a suitcase is not an action, it is an indirect effect.

Obviously, the ramification problem is intimately connected with the frame problem. In approaches that adopt a nonmonotonic inertial axiom to solve the frame problem, inertial defaults will need to be overridden by conclusions about ramifications to obtain correct results. In case the left lock of the suitcase is open, for instance, and an action of opening the right lock is performed, then the default conclusion that the suitcase remains closed, simply because it is closed initially, needs somehow to be overridden. The most detailed and promising approaches to the ramification problem depend on the development of theories of commonsense causation, and therefore are closely related to the causal approaches to reasoning about time and action discussed in section 4.6. See, for instance, Lin (1995), Thielscher (1989), and Giunchiglia et al. (1997).

Philosophical logicians have been content to illustrate their ideas with relatively small-scale examples. The formalization of even large-scale mathematical theories is relatively unproblematic. Logicist AI is the first branch of logic to undertake the task of formalizing large examples involving nontrivial commonsense reasoning. In doing so, the field has had to invent new methods. An important part of the methodology that has emerged in formalizing action and change is the prominence that is given to challenges, posed in the form of *scenarios*. These scenarios represent formalization problems that usually involve relatively simple, realistic examples designed to challenge the logical theories in specific ways. Typically, there will be clear commonsense intuitions about the inferences that should be drawn in these cases. The challenge is to design a logical formalism that will provide general, well-motivated solutions to these benchmark problems.

Among the many scenarios that have been discussed in the literature are the baby scenario, the bus ride scenario, the chess board scenario, the ferryboat connection scenario, the furniture assembly scenario, the hiding turkey scenario, the kitchen sink scenario, the Russian turkey scenario, the Stanford murder mystery, the Stockholm delivery scenario, the stolen car scenario, the stuffy room scenario, the ticketed car scenario, the walking turkey scenario, and the Yale shooting anomaly. Accounts of these can be found in Shanahan (1997) and Sandewall (1994, esp. chapters 2 and 7).

Many of these scenarios are designed to test advanced problems that I do not discuss here—for instance, challenges dealing with multiple agents, or with continuous changes. Here, I concentrate on one of the earliest, and probably the most subtle of these scenarios: the Yale shooting anomaly, first reported in Hanks and McDermott (1985) and published in Hanks and McDermott (1986, 1987).

The Yale shooting anomaly involves three actions: `load`, `shoot`, and `wait`. A propositional fluent `Loaded` tracks whether a certain pistol is loaded; another fluent, `Alive`, tracks whether a certain person, Fred, is alive. `load` has no preconditions; its only effect is `Loaded`. `shoot` has `Loaded` as its only precondition and `Alive` as a negative effect; `wait` has no preconditions and no effects.

Causal information regarding the axioms is formalized as follows.

Load $\forall s\, \mathit{Holds}(\texttt{load}, \textsc{Result}(\texttt{load}, s))$

Shoot 1 $\forall s[\mathit{Holds}(\texttt{Loaded}, s) \rightarrow \mathit{Holds}(\neg\texttt{Alive}, \textsc{Result}(\texttt{shoot}, s))]$

Shoot 2 $\forall s[\mathit{Holds}(\texttt{Loaded}, s) \rightarrow \mathit{Holds}(\neg\texttt{Loaded}, \textsc{Result}(\texttt{shoot}, s))]$

There is no wait axiom—that is, `wait` has no preconditions and no effects.

We will formalize the inertial reasoning in this scenario using a nonmonotonic logic—to be specific, we use Reiter's default logic. The set D of defaults for this theory consists of all instances of the inertial schema **IR**.

In the initial situation, Fred is alive and the pistol is unloaded.

IC 1 $\mathit{Holds}(\texttt{Alive}, s_0)$

IC 2 $\neg\mathit{Holds}(\texttt{Loaded}, s_0)$

The monotonic theory W of the scenario consists of (1) the action axioms **Load**, **Shoot 1**, and **Shoot 2** and (2) the initial conditions **IC 1** and **IC 2**.

Let $s_1 = \textsc{Result}(\texttt{load}, s_0)$, $s_2 = \textsc{Result}(\texttt{wait}, s_1)$, and $s_3 = \textsc{Result}(\texttt{shoot}, s_2)$.

The Yale shooting anomaly consists of the fact that the theory allows an extension in which the actions are `load`; `shoot`; `wait`, and in the final situation s_3, the pistol is unloaded and Fred is alive. The initial situation in the anomaly and the three actions, with their resulting situations, can be pictured as follows.

$$s_0 \xrightarrow{\texttt{load}} s_1 \xrightarrow{\texttt{wait}} s_2 \xrightarrow{\texttt{shoot}} s_3$$

The natural, expected outcome of these axioms is that the pistol is loaded and Fred is alive after waiting, so that shooting yields a final outcome in which

Fred is not alive and the pistol is unloaded. There is no problem in showing that this corresponds to an extension; the problem is the presence of the other, anomalous extension, which looks like this.

`Alive`		`Alive`		`Alive`		`Alive`
`¬Loaded`	$\xrightarrow{\texttt{load}}$	`Loaded`	$\xrightarrow{\texttt{wait}}$	`¬Loaded`	$\xrightarrow{\texttt{shoot}}$	`¬Loaded`
s_0		s_1		s_2		s_3

Here is a narrative version of this extension. At first, Fred is alive and the pistol is unloaded. After loading, the pistol is loaded and Fred remains alive. After waiting, the pistol becomes unloaded and Fred remains alive. Shooting is then vacuous since the pistol is unloaded, so finally, after shooting, Fred remains alive and the pistol remains unloaded. The best way to see clearly that this is an extension is to work through the proof. Less formally, though, you can see that the expected extension violates just one default: The frame default for `Alive` is violated when Fred changes state in the last step. But the anomalous extension also violates only one default: The frame default for `Loaded` is violated when the pistol spontaneously becomes unloaded while waiting. So, if you just go by the number of defaults that are violated, both extensions are equally good.

The Yale shooting anomaly represents a major obstacle in developing a theory of predictive reasoning. A plausible, well-motivated logical solution to the frame problem runs afoul of a simple, crisp example in which it clearly delivers the wrong results. Naturally, the literature concerning the Yale shooting problem is extensive. Surveys of some of this work, with bibliographical references, can be found in Morgenstern (1996) and Shanahan (1997).

4.6. Some Emergent Frameworks

Many formalisms have been proposed to deal with the problems surveyed in the previous section. Some are more or less neglected today. Several are still advocated and defended by leading experts; some of these are associated with research groups who are interested in not only developments of logical theory but in applications in planning and cognitive robotics.

The leading approaches provide solutions to the main problems mentioned in section 4.5, and to many of the scenarios designed to test and illustrate theories of reasoning about action and change. It is commonly agreed that good solutions need to be generalizable to more complex cases than the early planning formalisms, and that in particular the solutions they offer should be deployable even when continuous time, concurrent actions, and various kinds of ignorance are allowed. Also, it is generally agreed that the formalisms should support several kinds of reasoning, and, in particular, not only prediction and plan verification but *retrodiction*, that is, construction of a sequence of states and actions given partial information in the form of a narrative.

I will describe four approaches here: (1) features and fluents (Sandewall), (2) motivated action theory (Morgenstern and Stein), (3) state minimization in the event calculus (Shanahan), and (4) causal theories (Lifschitz and others). My accounts of the first three will be fairly brief; fortunately, each approach is well documented in a single reference. I believe that the fourth approach is likely to be most interesting to philosophers, and that it contains elements that will be of lasting importance whatever changes future developments in this area may bring.

Features and Fluents This approach, described in Sandewall (1994), uses preference semantics as a way to organize nonmonotonic solutions to the problems of reasoning about action and change. Rather than introducing a single logical framework, Sandewall considers a number of temporal logics, including ones that use discrete, continuous, and branching time. The properties of the logics are systematically tested against a large suite of test scenarios.

Motivated Action Theory This theory grew out of direct consideration of the problems in temporal reasoning described in section 4.5, and especially the Yale shooting scenario. Morgenstern and Stein (1994) seek to find a general, intuitively motivated logical framework that solves the difficulties. They settle on the idea that unmotivated actions are to be minimized, where an action ("actions" construed generally enough to include any change) can be motivated directly, for example, by an axiom, or indirectly, through chains of motivations. The key technical idea of the paper is a (rather complicated) definition of motivation in an interval-based temporal logic. Morgenstern (1996) presents a summary of the theory, along with reasons for rejecting its causal rivals. The most important of these reasons is that these theories, based on the situation calculus, do not appear to generalize to cases allowing for concurrency and ignorance. She also cites the failure of early causal theories to deal with retrodiction.

State-Based Minimization in the Event Calculus Baker (1989) presented a solution to the version of the Yale shooting problem in the situation calculus, using a circumscriptive inertial axiom. The very brief account of circumscription in section 3 indicated that circumscription uses preferred models in which the extensions of certain predicates are minimized. In the course of this minimization, a set of parameters (including, of course, the predicates to be minimized) is allowed to vary; the rest are held constant. Which parameters vary and which are held constant is determined by the application.

In the earliest circumscriptive solutions to the frame problem, the inertial rule **CIR** is stated using an abnormality predicate:

CIR $\forall \mathrm{f} \forall \mathrm{s} \forall \mathrm{a}[\neg \mathit{Ab}(\mathrm{f}, \mathrm{a}, \mathrm{s}) \rightarrow [\mathit{Holds}(\mathrm{f}, \mathrm{s}) \leftrightarrow \mathit{Holds}(\mathrm{f}, \textsc{Result}(\mathrm{a}, \mathrm{s}))]]$.

This axiom uses a biconditional, so that it can be used for retrodiction; this is typical of the more recent formulations of common sense inertia.

In circumscribing, the abnormality predicate is minimized while the *Holds* predicate is allowed to vary and all other parameters are fixed. This formalization succumbs to the Yale shooting anomaly in much the same way that default logic does. (Circumscription does not involve multiple extensions, so the problem emerges as the nonderivability of the conclusion that Fred is alive after the occurrence of the shooting.)

In Baker's reformulation of the problem, separate axioms ensure the existence of a situation corresponding to each Boolean combination of fluents, and the RESULT function is allowed to vary, while the *Holds* predicate is held constant. In this setting, the RESULT function needs to be specified for "counterfactual" actions—in particular, for shooting as well as for waiting in the Yale shooting anomaly. It is this feature that eliminates the incorrect model for that scenario; for details, see Baker (1989) and Shanahan (1997, chapter 6).

This idea, which Shanahan calls "state-based minimization," is developed and extended in Shanahan (1997), in the context of a temporal logic deriving from the event calculus of Kowalski and Sergot (1986). Shanahan's formalism has the advantage of being closely connected to implementations using logic programming.

Causal Theories Recall that in the anomalous model of the Yale shooting scenario the gun becomes unloaded after the performance of the `wait` action, an action that has no conventional effects—the unloading, then, is uncaused. In the context of a nonmonotonic logic—and without such a logic, the Yale shooting anomaly would not arise—it is very natural to formalize this by treating uncaused eventualities as abnormalities to be minimized.

This strategy was pursued by Hector Geffner (1990, 1992), and he formalizes this simple causal solution to the Yale shooting anomaly. But the solution is presented in the context of an ambitious general project in nonmonotonic logic that not only develops properties of the preferred model approach and shows how to apply it to a number of reasoning problems but relates nonmonotonic logic to probabilities, using ideas deriving from Adams (1975). In Geffner (1992), the causal theory is sketched; it is not developed to show its adequacy in dealing with the battery of problems presented, and in particular the ramification problem is left untouched.

The work beginning with Lifschitz (1987) has contributed to a sustained line of research in the causal approach—not only by Lifschitz and students of his such as Enrico Giunchiglia and Hudson Turner but by researchers at other sites. For work in this area, and further references, see Haugh (1987), Elkan (1991), Baral (1995), Lin (1995), McCain and Turner (1995, 1997), Gustaffson and Doherty (1996), Thielscher (1989, 1996), Gelfond and Lifschitz (1998), Lifschitz (1997, 1998), Nakashima et al. (1997), Giunchiglia and Lifschitz (1998), and Turner (1999).

Here, I briefly describe some of the features of the theory presented in Turner (1999), which returns to the ideas of Geffner (1992), but places them in a simpler logical setting and applies them to the formalization of more complex scenarios that illustrate the interactions of causal inertia with other considerations, especially the ramification problem.

The idea is to treat `Caused` as a modal operator $[c]$, making this the basis of a modal nonmonotonic logic. In the preferred models of this logic, the caused propositions coincide with the propositions that are true, and this must be the only possibility consistent with the extensional part of the model. To make this more explicit, recall that in the possible-worlds interpretation of **S5**, it is possible to identify possible worlds with *state descriptions*, which we can represent as sets I of literals (atomic formulas and their negations). Making this identification, then, we can think of a model as a pair $\langle I, S\rangle$, where S is a set of interpretations including I. The modal operator $[c]$ is given the standard semantics: Let S be a set of interpretations. Then, where $I \in S$, $S \models_I [c]A$ if and only if $S \models_{I'} A$ for all $I' \in S$. $\langle I, S\rangle$ *satisfies* a set of formulas T if and only if $S \models_I A$ for all $A \in T$.

Turner's preferred models of T are the pairs $\langle I, S\rangle$ such that (1) $\langle I, S\rangle$ satisfies T, (2) $S = \{I\}$, and (3) $\langle I, S\rangle$ is the unique interpretation $\langle I', S'\rangle$ meeting conditions (1) and (2) with $I' = I$. Condition (2) guarantees the "universality of causation"; it validates $A \leftrightarrow [c]A$. Condition (3) "grounds" causality in noncausal information (in the models in which we are interested, this will be information about the occurrence of events), in the strongest sense: It is uniquely determined by this information.

Although it is not evident from the formulation, Turner's account of preferred models is related to the constructions of more general nonmonotonic logics, such as default logic. Consult Turner (1999) for details.

The axioms that specify the effects of actions treat these effects as caused; for instance, the axiom schema for loading would read as follows.

Causal-Load $[c]\mathit{Holds}(\mathtt{load}, \textsc{Result}(\mathtt{load}, \mathrm{s}))$[31]

Ramifications of the immediate effects of actions are also treated as caused. And the nonmonotonic inertial axiom schemata take the form

$$[[[c]\mathit{Holds}(\mathrm{f}, \mathrm{s})] \wedge \mathit{Holds}(\mathrm{f}, \textsc{Result}(a, \mathrm{s}))] \rightarrow [c]\mathit{Holds}(\mathrm{f}, \textsc{Result}(a, \mathrm{s}))$$

and

$$[[[c]\neg\mathit{Holds}(\mathrm{f}, \mathrm{s})] \wedge \neg\mathit{Holds}(\mathrm{f}, \textsc{Result}(a, \mathrm{s}))] \rightarrow [c]\neg\mathit{Holds}(\mathrm{f}, \textsc{Result}(a, \mathrm{s})).$$

Thus, a true proposition can be caused either because it is the direct or indirect effect of an action, or because it involves the persistence of a caused proposition. Initial conditions are also considered to be caused, by stipulation.

To illustrate the workings of this approach, let's consider the simplest case of inertia: We have a language with just one constant denoting a fluent, f, and one action-denoting constant, `wait`. As in the Yale shooting problem, there are no axioms for `wait`; this action can always be performed and has no associated effects. Let s_1 be $\textsc{Result}(\mathtt{wait}, \mathrm{s}_0)$. The theory T contains an initial condition for f, $Holds(\mathrm{f}, \mathrm{s}_0)$ and a statement that the initial condition is caused, $[c]Holds(\mathrm{f}, \mathrm{s}_0)$, as well as the inertial schemata.

Two models of T satisfy conditions (1) and (2): $M_1 = \langle I_1, \{I_1\}\rangle$ and $M_2 = \langle I_2, \{I_2\}\rangle$, where $I_1 = \{Holds(\mathrm{f}, \mathrm{s}_0), Holds(\mathrm{f}, \mathrm{s}_1)\}$ and $I_2 = \{Holds(\mathrm{f}, \mathrm{s}_0), \neg Holds(\mathrm{f}, \mathrm{s}_1)\}$.

M_1 is the intended model, in which nothing changes. It satisfies condition (3), since if $\langle I_1, S\rangle$ satisfies T it satisfies $[c]Holds(\mathrm{f}, \mathrm{s}_1)$ by the inertial axiom

$$[[c]Holds(\mathrm{f}, \mathrm{s})] \wedge Holds(\mathrm{f}, \mathrm{s}_1))] \rightarrow [c]Holds(\mathrm{f}, \mathrm{s}_1).$$

Therefore, $S = \{I_1\}$.

M_2 is an anomalous model, in which the fluent ceases spontaneously. This model does not satisfy condition (3), since $M_3 = \langle I_2, \{I_1, I_2\}\rangle$ also satisfies T; in particular, it satisfies the inertial axiom for f because it fails to satisfy $Holds(\mathrm{f}, \mathrm{s}_1)$. So, while M_1 is a preferred model, M_2 is not.

The apparent usefulness of a "principle of universal causality" in accounting for a range of problems in qualitative commonsense reasoning will be tantalizing to philosophers. And the causal theory, as initiated by Geffner and developed by Turner, has many interesting detailed features. For instance, while philosophical work on causality has concentrated on the causal relation, this work in logical AI shows that a great deal can be done by using only a nonrelational causal predicate.

Morgenstern's two chief criticisms of the causal approach to reasoning about actions are that it does not give an adequate account of explanation[32] and that the logical context in which it works (the situation calculus) is limited. As work on the approach continues, progress is being made in these areas. But the constraints that a successful logic of action and change must meet are so complex that it is a reasonable research methodology to concentrate initially on a restricted logical setting.

4.7. Action Formalisms and Natural Language

Although for many AI logicists, the goal of action formalisms is to illuminate an important aspect of common sense reasoning, most of their research is uninformed by an important source of insights into the commonsense view of time—namely, natural language. Linguists concerned with the semantics of temporal constructions in natural language, like the AI community, have begun with ideas from philosophical logic but have discovered that these ideas need to be modified to deal with the phenomena. A chief discovery of the AI logicists has been the importance of actions and their relation to change. Similarly, an important discovery of the "natural language logicists" has been the importance of different kinds of events (including structured composite events) in interpreting natural language. From work such as this the idea of "natural language metaphysics" (see, for instance, Bach 1989) has emerged.

The goal of articulating a logical framework tailored to a representational system that is motivated by systematic evidence about meanings in natural languages is not acknowledged by all linguistic semanticists. Nevertheless, it is a significant theme in the linguistic literature. This goal is remarkably similar to those of the commonsense logicists, but the research methodology is entirely different.

Can the insights of these separate traditions be reconciled and unified? Is it possible to constrain theories of temporal representations and reasoning with the insights and research methodologies of both traditions? In Steedman (1995, 1998) these important questions are addressed, and a theory is developed that extends action formalisms like the situation calculus, and that incorporates many of the insights from linguistic semantics. The project reported in Steedman (1998) is still incomplete, but the results reported there make a convincing case that the event-based ideas from linguistics can be fruitfully combined with the action-centered formalisms in the AI literature. The possibility of this unification is one of the most exciting logical developments in this area, bringing together as it does two independent descendants of the earlier work in the logic of time.

5. Causal Reasoning

In section 4.6, we traced the reasons for the development of theories incorporating causality in work on reasoning about action and change. This is not the only area of AI in which causality has emerged. Causality figures in qualitative reasoning about devices; for Herbert Simon's important work in this area, which goes back to the 1950s, see Simon (1952, 1977) and Iwasaki and Simon (1986). Both these traditions are important. But the most robust and highly developed program in AI relating to causality is that of Judea Pearl and his students and associates, which derives from the use of causal diagrams in the formalism for reasoning about probabilities known as *Bayesian belief networks*.

Pearl's program has developed into a far-reaching campaign to rehabilitate causality in statistical thinking. I do not discuss this topic here. For one thing, this survey omits probabilistic reasoning in AI. For another, Pearl's views on causality are systematically and comprehensively presented in a recent book-length study (Pearl 2000).

But I do wish to point out that the work on causality discussed in section 4.6 and Pearl's ideas do share some common themes. On both approaches: *Action* is central for causality. Also there is a focus on causality as a tool in reasoning that is necessitated in part by limited resources. Another important theme is the deployment and systematic study of formalisms in which causality is related to other constructs (in particular, to probability and to qualitative change) and a variety of realistic reasoning problems are addressed.

These commonalities provide reason to hope that we will see a science of causality emerging from the AI research, unifying the contributions of the probabilistic, the qualitative physics, and the nonmonotonic traditions, and illuminating the various phases of causal reasoning.

Whether you take causality to be a fundamental construct in natural science, or a fundamental natural phenomenon, depends on whether you have in mind an idealized nature described by differential equations or you have in mind the view of nature we have to take to act, either in everyday situations, or for that matter in designing experiments in the laboratory. The fact that, as Bertrand Russell noted (1957), causality is not to be found as a theoretical primitive in contemporary physical theories is at odds with its seeming importance in so many familiar areas of reasoning. The rigorous theories emerging in AI that are beginning to illuminate the workings of causality are important not only in themselves but in their potentiality to illuminate wider philosophical issues.

6. Spatial Reasoning

The precomputational literature in philosophical logic relating to spatial reasoning is very sparse in relation, for instance, to the temporal literature. The need to support computational reasoning about space, however, in application areas such as motion planning and manipulation in physical space, the indexing and retrieval of images, geographic information systems, diagrammatic reasoning, and the design of high-level graphics programs has led to new interest in spatial representations and spatial reasoning. Of course, the geometrical tradition provides an exceptionally strong mathematical resource for this enterprise. But as in many other AI-related areas, it is not clear that the available mathematical theories are appropriate for informing these applications, and many computer scientists have felt it worthwhile to develop new foundations. Some of this work is closely related to the research in qualitative reasoning mentioned in section 2.2, and in some cases has been carried out by the same individuals.

The literature in spatial reasoning is extensive; for references to some areas not discussed here, see Forbus et al. (1991), Kapur and Mundy (1988), Renz and Nebel (1999), Stock (1997), Wilson (1998), Yeap and Jeffries (1999), Allwein and Barwise (1996), Glasgow et al. (1995), Hammer (1995), Kosslyn (1990), Osherson and Lasnik (1990), Burger and Bhanu (1992), and Chen (1990). Here, I discuss only one trend, which is closely connected with parallel work in philosophical logic.

Qualitative approaches to space were introduced into the logical literature early in the twentieth century by Leśniewski; see Leśniewski (1916), which presents the idea of a *mereology*, or qualitative theory of the part-whole relation between physical individuals. This idea of a logical theory of relations among regions or the objects that occupy them, which does not depend on construing regions as sets of points, remained an active area of philosophical logic, even though it attracted relatively few researchers. More recent work in the philosophical literature, especially Clarke (1981, 1985), Simons (1987), and Casati and Varzi (1996, 1999), was directly influential on current computational work.

The regional connection calculus (RCC), developed by computer scientists at the University of Leeds, is based on a primitive C relating regions of space: The intended interpretation of $C(x, y)$ is that the intersection of the closures of the values of x and y is nonempty. (See Cohn et al. 1997; Cohn 1996 for details and references.) One area of research concerns the definability of shapes in RCC. The extent of what can be defined with this simple primitive is surprising, but the technicalities quickly become complex; see, for instance, Gotts (1994, 1996). The work cited in Cohn et al. (1997) describes constraint propagation techniques and encodings in intuitionistic propositional logic as ways of supporting implemented reasoning based on RCC and some of its extensions. More recent work based on RCC addresses representation and reasoning about motion, which of course combines spatial and temporal issues; see Wolter and Zakharyaschev (2000). For more information about qualitative theories of movement, with references to other approaches, see Galton (1997).

7. Reasoning about Knowledge

Epistemic logic is another area in which strong influences from philosophical logic can be traced on logic in computer science. The classical source for epistemic logic is Hintikka (1962), in which Jaakko Hintikka showed that a modal approach to single-agent epistemic attitudes could be informative and rewarding. This work discusses at length the question of exactly which constraints are appropriate for knowledge and belief, when these attitudes are viewed as explicated by a model theoretic relation over possible worlds; in both cases, Hintikka argues for **S4** type operators.

In several papers (including McCarthy 1979), McCarthy has recommended an approach to formalizing knowledge that uses first-order logic, but that quantifies explicitly over such things as individual concepts. In this section I'll discuss the approach taken by most computer scientists, however, who use a modal language to formalize propositional attitudes.

The logical aspects of modal epistemic logic were not significantly developed after Hintikka's 1962 presentation; instead, the philosophical literature (which is not extensive, compared with many other topics in the area) concentrates on the issue of *hyperintensionality*, that is, whether epistemic attitudes should be closed under logical consequence. This topic is especially challenging, turning out to be closely related to the semantic paradoxes, and the philosophical literature is inconclusive. Intuitions seem to conflict, and it is difficult to find ways to model the important phenomena using logical techniques.[33]

Fagin et al. (1984) begins a tradition in computational logic that revives the modal approach to epistemic logic, developing generalized logical foundations and applications that had not occurred to the philosophers. The technical idea is to simplify the modality, using **S5** (or deontic **S5** for belief), but to introduce multiple agents and concentrate on reasoning having to do with agents' attitudes about one another's attitudes. Such logics have direct applications in the analysis of *distributed systems*, dynamic systems in which change is effected by message actions, which change the knowledge of agents according to rules determined by a *communications protocol*.

As such, this work belongs to a separate area of computer science, but one that overlaps to some extent with AI. Later, this work has interacted with a research tradition in economics that is concerned with the role of knowledge in games and bargaining; see, for instance, Geanakopolos (1994) and Osborne and Rubenstein (1994, chapter 5).

For some reason, the multiagent case did not occur to philosophical logicians.[34] This is another example of the way in which need for an application (in this case, the need for a theory of distributed systems) provided the inspiration for an important logical development. I will not present details concerning the logic here, since they are extensively and systematically recorded in Fagin et al. (1995); this is essential reading for anyone seriously interested in this topic.

Much of the interdisciplinary work in applications of the logic of knowledge is reported in the proceedings of a series of conferences initiated in 1986 with Halpern (1986). These conferences record one of the most successful collaborations of philosophers with logicians in computer science, although the group of involved philosophers has been relatively small. The focus of the conferences has gradually shifted from computer science to economics.

AI applications deal with with knowledge in the form of stored representations, and the tradition in AI with which we are concerned here thinks of reasoning as the manipulation of symbolic representations. Also, it is mainly due to AI that the problem of limited rationality has become a topic of serious interest, providing a counterbalance to the idealizations of philosophy and economics.[35] So you would think that a logical model of propositional attitudes that is committed to closure under logical consequence would be highly unpopular in AI. But this is not so; the possible worlds approach to attitudes is not only the leading theory in the areas discussed in Fagin et al. (1995) but has even been advocated in robotics applications; see Rosenschein (1989) and Rosenschein and Kaelbling (1995).

Nevertheless, the issue of hyperintensionality has been investigated in the AI literature; see Konolige (1986), Lakemeyer (1997), Levesque (1984), and Perlis (1985). Though there are some new positive results here, the AI work in this area, in my opinion, has been as inconclusive as that in philosophy.

The philosophical literature on a related topic, the logic of perception, has not been extensive; the main reference is Hintikka (1970).[36] But sensation is addressed in recent work in the AI literature which is concerned with developing logical frameworks for general-purpose applications in Robotics. The main idea in this area is to add sensing actions to the repertoire of a planning formalism of the sort discussed in section 4. The earliest work in this area was carried out in the 1980s by Robert Moore; see R. C. Moore (1985, 1995). For some of the contemporary work in cognitive robotics, see Bacchus et al. (1999), Baral et al. (2000), Golden and Weld (1996), Pirri and Finzi (1999), and Thielscher (2000).

8. Logical Approaches to Natural Language and Communication

Over the last 25 years or so, many profound relations have emerged between logic and grammar. Computational linguistics (or natural language processing) is a branch of AI, and it is fairly natural to classify some of these developments under logic and AI. But many of them also belong to an independent tradition in logical foundations of linguistics; in many cases it is hard (and pointless) to attempt a classification. This sketch concentrates on developments that focus on reasoning; other applications of logic to linguistics are described in van Benthem and ter Meulen (1996).

9. Parsing and Deduction

Grammar formalisms—special-purpose systems for the description of linguistic systems and subsystems—can be thought of as logics designed to axiomatize the association of linguistic structures with strings of symbols. You might be able to infer from such a system, for instance, that "assignments" is the plural form of the nominalization of the verb "assign." So you can look at the process of *parsing* a string of words—of finding the linguistic structures, if any, that are associated with it—as a search for a proof in a certain logical system.

This approach has been highly successful as an analytic tool. It makes model-theoretic techniques applicable to linguistic reasoning, This makes the underlying reasoning problems much more transparent, and makes it possible to apply many well-developed areas of logic to grammar formalisms. For more information on these topics, see Shieber (1992) and Buszkowski (1996).

10. Feature Structure Logic

The usefulness and scope of logical techniques in relation to linguistics is greatly increased by the development of techniques for analyzing the way information is attached to linguistic units. It is very natural to represent the information attaching, say, to a lexical item in the form of a set of functions (or attributes) that produce values in some linguistic domain. A pronoun x may have a number, a person, and a case: if $x =$ "we" then

$$\begin{aligned} &number(x) = \text{plural},\\ &person(x) = \text{first},\\ &case(x) = \text{nominative}. \end{aligned}$$

In more general cases, the values of these functions may themselves be linguistic units that take on values for certain attributes.

Allowing these functions to be partial provides a useful informational representation of the stages of a linguistic parse; much of the work of parsing involves completing this partial information, subject to constraints imposed by linguistic agreement conditions. Feature structures have a natural algebraic treatment, and there is an elegant treatment of their logic. For more information and references, see Rounds (1996).

11. Logic and Discourse

The reasoning associated with discourse is the probably the least well understood area of computational linguistics. Although logical techniques do not yet play a major role in discourse, they seem to offer one of the most promising ways of providing a uniform account of the many forms of reasoning that are involved in generating and interpreting language in interactive conversation.

I briefly mention three contributions to this area. Building on the fact that the rules governing conversation are exception-ridden, Alex Lascarides and Nicholas Asher have developed techniques for formalizing discourse phenomena based on nonmonotonic logic; see Asher and Lascarides (1994, 1997). Jerry Hobbs and various co-workers look at the inference processes used in discourse as *abductive*, and propose to formalize abduction as a search for a proof in which certain “low-cost” assumptions may be made that serve as data or additional axioms for the proof. Hobbs et al. (1993) shows how an impressive range of discourse phenomena can be formalized using this idea. In practice, this abductive account looks rather similar to that of Lascarides and Asher, because it involves deploying axioms about discourse (in the form of Horn clause rules supplemented with weights giving the assumption costs of premises) that in effect are nonmonotonic.

In more recent work, Matthew Stone (1998) shows how modal logic can inform the complex reasoning involved in natural language generation. Generating a coherent, appropriately phrased text that usefully performs a task-oriented communication task is difficult to formalize because it requires the integration of complex and sophisticated domain information with discourse planning, user modeling, and linguistic constraints. Stone shows that modal logic can be used to modularize the formalization of the information required in this task; he also shows how modal theorem proving can be used to implement the reasoning.

12. Taxonomic Representation and Reasoning

12.1. Concept-Based Classification

Traditionally, the task of representing large amounts of domain information for general-purpose reasoning has been one of the most important areas of knowledge representation. Systems that exploit the intuitive taxonomic organization of domains are useful for this purpose; taxonomic hierarchies not only help to organize the process of knowledge acquisition but provide a useful connection to rule-based reasoning.[37]

For domains in which complex definitions are a natural way to organize information, knowledge engineering services based on definitions of concepts have been extremely successful. Like variable-free versions of first-order logic (see, for instance, Quine 1960), these systems are centered on concepts or first-order predicates, and provide a number of mechanisms for their definition. The fundamental algorithm associated with these *taxonomic logics* is a classifier that inputs a system of definitions and outputs the entailment relations between defined and primitive concepts. For background on these systems, see Brachman et al. (1991) and Woods and Schmolze (1992).

The simplest taxonomic logics can be regarded as subsystems of first-order logic with complex predicates, but they have been extended in many ways, and the issues raised by many of these extensions overlap in many cases with topics in philosophical logic.

12.2. Nonmonotonic Inheritance

Much more complex logical issues arise when the organization of a domain into hierarchies is allowed to have exceptions. One way to approach this topic is to explore how to make a taxonomic logic nonmonotonic in its own right; but *nonmonotonic inheritance* is a topic in its own right. Although there are strong affinities to nonmonotonic logic, nonmonotonic logic relies more heavily on graph-based representations than on traditional logical ideas, and seems to provide a much finer-grained approach to nonmonotonic reasoning that raises entirely new issues, which quickly becomes problematic. For this reason, systems of nonmonotonic inheritance tend to be expressively weak, and their relations to the more powerful nonmonotonic logic has never been fully clarified. For background on this topic, see Thomason (1992) and Horty (1994).

13. Contextual Reasoning

In the tradition in philosophical logic dealing with contextual effects on the interpretation of expressions, as well as in the more recent tradition in dynamic logic, context is primarily formalized as an assignment of values to variables, and the language is designed to make explicit reasoning about context either very limited or outright impossible.

Concern in AI about the representation of large and apparently heterogeneous domains and about the integration of disparate knowledge sources, as well as interests in formalizing common sense of the sort discussed in section 2.2, have led to interest in the AI community in formalizing languages that take context into account more explicitly.

McCarthy (1993b) recommends the study of languages containing a construct

$$ist(c, \phi),$$

where *ist* is read "is-true." This is analogous to the *Holds* construct of the situation calculus—but now c stands for a context, and ϕ is a possibly complex propositional representation, which many (including McCarthy) take to refer to a sentence.

There are analogies here both to modal logic and to languages with an explicit truth-predicate. But the applications that are envisioned for a logic of context create opportunities and problems that are in many ways new. For more about the logic of context, see Guha (1991), McCarthy and Buvač (1998), and the papers in Bouquet et al. (1999) and Akman et al. (2001).

14. Prospects for a Logical Theory of Practical Reason

I believe there is reason to hope that the combination of logical methods with planning applications in AI can enable the development of a far more comprehensive and adequate theory of practical reasoning than has heretofore been possible. As with many problems having to do with commonsense reasoning, the scale and complexity of the formalizations that are required are beyond the traditional techniques of philosophical logic. However, with computational methods of implementing and testing the formalizations and with areas such as cognitive robotics to serve as laboratories for developing and testing ideas, we can hope to radically advance a problem that has seen little progress since it was first proposed by Aristotle: the problem of devising a formalization of practical reasoning that is genuinely applicable to commonsense reasoning problems.

The classical work in deontic logic that was begun by von Wright (1983) is one source of ideas; see van der Torre (1997) and Horty (2001). In fact, as the more recent work in deontic logic shows, nonmonotonic logic provides a natural and useful way to modify the classical deontic logic.

An even more robust account of practical reasoning begins to emerge when these ideas are supplemented with work on the foundations of planning and reasoning about action that were discussed in section 4. But this development can be pursued even further, by extending the formalism to include preferences and intentions.[38]

Ultimately, what is needed is a model of an intelligent reasoning and acting agent. Developing such a model need not be entirely a matter of logic, but according to one school of thought, logic has a central role to play in it; see, for instance, Baral and Gelfond (2000), Burkhard et al. (1998), Rao and Georgeff (1991), and Wobcke et al. (1998).

Acknowledgments

I am grateful to John McCarthy, who read a draft of this article and provided extensive and helpful comments.

Notes

1. See, for instance, Nilsson (1995).

2. For two debates, see volume 3, number 3 of *Computational Intelligence*, devoted to McDermott (1987), and the later exchange Nilsson (1991), Birnbaum (1991).

3. For some of the historical background, see Davis (1988).

4. See Stefik (1995) for general background on expert systems. For information concerning explanation, see Clancey (1983) and J. Moore (1995).

5. For a good example of the use of these intuitions to motivate a system of logic, see the extended argument in Hintikka (1962) that the modal logic **S4** is the correct logic of belief.

6. The submissions to the 1989 conference were unclassified as to topic; I sampled every other article, a total of 522. The 1989 conference divided its contributed articles into 26 topical sessions; I sampled the first paper in each of these sessions.

7. In the decade from 1990 to 1999 I counted one *JPL* publication by an AI researcher, Boutilier (1996), and five papers showing some AI influence; all of these dealt with nonmonotonic logic.

8. This includes robots (or "softbots") that navigate artificial environments such as the Internet or virtual worlds as well as embodied robots that navigate the physical world.

9. I was surprised at first to hear the AI community refer to its logical advocates as logicists. On reflection, it seems to me much better to think of logicist projects in this general sense, as proposals to apply what Alonzo Church called "the logistic method" in seeking to understand reasoning in various domains. It is far too restrictive to narrowly associate logicism with Frege's program.

10. Data integration is one such area. See Levy (2000). Large-scale knowledge representation is another. See Lenat and Guha (1989).

11. See Reiter (2001) for an extended contribution to cognitive robotics, with references to some of the other literature in this area. Reiter's book also contains self-contained chapters on the situation calculus and the problems of formalizing reasoning about action and change. I recommend these chapters to anyone wishing to follow up on the topics discussed in section 4. Another extended treatment of action formalisms and issues is Shanahan (1997).

12. Much of the work in this last area has not made heavy use of logical techniques. Qualitative physics and the formalization of other forms of qualitative reasoning is an independent specialty in AI, different in many ways from logical AI. Nevertheless, the two specialties have certainly influenced each other. For information concerning qualitative reasoning, consult Forbus (1988), Weld and de Kleer (1990), and Kuipers (1993).

13. For further details concerning this approach, see section 3.4.

14. John McCarthy makes a similar point, illustrating it with an example, in McCarthy (1993a).

15. This very difficult and not particularly well-defined problem was very much on the minds of many AI researchers in the area that later became knowledge representation, but it has not proved to be a productive focus for the field. Natural language interpretation has developed into a separate field that is largely concerned with less sweeping problems, such as automated speech-to-speech discourse, data mining, and text summarization. Logical techniques have been used with some success in this area, but it is fair to say that natural language interpretation has

not been the best showcase for logical ideas. Even the problem of providing an adequate semantic interpretation of generic constructions—a natural application of nonmonotonic logic—has turned out to be problematic. See Krifka et al. (1995) for a general discussion of the issues.

16. This use of the word "frame" is unconnected to the use of the term in the "frame problem," and is not to be confused with that problem.

17. The analogy to modal logics of provability inspired by Gödel's work (Boolos 1993) has, of course, been recognized in later work in nonmonotonic logic. But it has not been a theme of major importance.

18. See Konolige (1988).

19. Readers interested in the historical aspects of the material discussed in this section might wish to compare it to Ohrstrom and Hasle (1995)

20. For additional historical background on Prior's work, see Copeland (1996).

21. In retrospect, the term "situation" is not entirely fortunate, since it was later adopted independently and in quite a different sense by the situation semanticists (see, for instance, Seligman and Moss 1996). In the AI literature, the term "state" is often used interchangeably with "situation" and this, as far as I can see, causes no confusion: The connections with physical states, as well as with the more general states of any complex dynamic system are entirely appropriate.

22. The early versions of the situation calculus were meant to be compatible with concurrent cases, that is, with cases in which there are multiple planning agents, possibly acting simultaneously. But most of the logical analyses have been devoted to the single-agent case.

23. Carnap's attempts to formalize dispositional terms and inductive methods are classical examples of the problems that emerge in the formalization of empirical science.

24. For information about planning under uncertainty, see, for instance, Bacchus et al. (1999), Boutilier et al. (1996), and DeJong and Bennett (1989).

25. Examples are Dennett (1987) and Fodor (1987).

26. See Schubert (1990) and Reiter (1993).

27. See Carlson and Pelletier (1995).

28. This way of putting it is a little misleading for the situation calculus, since there is no robust notion of performing an action; instead, you consider the results of performing hypothetical action sequences. Even so, the point that the theory of unsuccessful actions has not been explored holds up.

29. Effects of actions that are delayed in time are a separate problem, which, as far as I know, no one has solved.

30. The relationship between an action and the occurrence of its conventional consequences is complicated, of course, by the "imperfective paradox" (see Dowty 1977; Lascarides 1992). Some of the work on AI theories of action and change is informed by these complexities; see Steedman (1998, 1995). But for the most part, they have not been taken into account in the AI literature.

31. Turner uses a discrete temporal logic other than the situation calculus. But for uniformity of presentation I have used the situation calculus to present the ideas.

32. In explanation problems, one is reasoning backward in time. Here, information is provided about a series of occurring states and the problem is to provide actions that account for the occurrences.

33. For information about the philosophical tradition, see Hintikka (1986). Also, see Laux and Wansing (1995).

34. A personal recollection: I was certainly aware of this case in the early 1970s, but did not devote much attention to it because it seemed to me that the generalization from the single-agent case was relatively trivial and did not pose any very interesting logical challenges.

35. See, for instance, Simon (1982a,b) and Russell and Wefald (1991).

36. Although this topic has received attention more recently in situation theory, the logical issues, in my opinion, have not been illuminated by this work.

37. See Stefik (1995) for background on considerations having to do with knowledge engineering.

38. For background on quantitative models of preference and decision, see Doyle and Thomason (1999). For work in AI on intentions, see, for instance Cohen and Levesque (1990), Konolige and Pollack (1993), Pollack (1992), and Sadek (1992).

References

Adams, Ernest W. 1975. *The Logic of Conditionals*. Dordrecht: D. Reidel.

Aiello, Luigia Carlucci, Jon Doyle, and Stuart Shapiro, eds. 1996. *KR'96: Principles of Knowledge Representation and Reasoning*. San Francisco: Morgan Kaufmann.

Akman, Varol, Paolo Bouquet, Richmond Thomason, and Roger A. Young, eds. 2001. *Modeling and Using Context*. Berlin: Springer-Verlag.

Alcourrón, Carlos E. 1995. Defeasible logics: Demarcation and affinities. In *Conditionals: From Philosophy to Computer Science*, eds. Gabriella Crocco, Luis Fariñas del Cerro, and A. Herzig, 67–102. Oxford: Oxford University Press.

Allen, James F., Richard Fikes, and Erik Sandewall, eds. 1991. *KR'91: Principles of Knowledge Representation and Reasoning*. San Mateo, Calif.: Morgan Kaufmann.

Allwein, Gerard, and Jon Barwise, eds. 1996. *Logical Reasoning with Diagrams*. Oxford: Oxford University Press.

Amarel, Saul. 1968. On representations of problems of reasoning about actions. In *Machine Intelligence 3*, ed. D. Mitchie, 131–171. Chichester, England: Ellis Horwood.

Antoniou, Grigoris. 1997. *Nonmonotonic Reasoning*. Cambridge, Mass.: MIT Press.

Arlo-Costa, Horacio, and Scott Shapiro. 1992. Maps between nonmonotonic logic and conditional logic. In *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, eds. Bernhard Nebel, Charles Rich, and William Swartout, 553–564. San Mateo, Calif.: Morgan Kaufmann.

Asher, Nicholas. 1995. Commonsense entailment: A conditional logic for some generics. In *Conditionals: From Philosophy to Computer Science*, eds. Gabriella Crocco, Luis Fariñas del Cerro, and A. Herzig, 103–145. Oxford: Oxford University Press.

Asher, Nicholas, and Alex Lascarides. 1994. Intentions and information in discourse. In *Proceedings of the Thirty-Second Meeting of the Association for Computational Linguistics*, ed. James Pustejovsky, 35–41. San Francisco: Association for Computational Linguistics, Morgan Kaufmann.

Asher, Nicholas, and Alex Lascarides. 1997. Lexical disambiguation in a discourse context. In *Lexical Semantics: The Problem of Polysemy*, eds. James Pustejovsky and Brian Boguraev, 69–108. Oxford: Oxford University Press.

Asher, Nicholas, and Michael Morreau. 1991. Commonsense entailment: A modal theory of nonmonotonic reasoning. In *Proceedings of the Twelfth International*

Joint Conference on Artificial Intelligence, eds. J. Mylopoulos and R. Reiter, 387–392. Los Altos, Calif.: Morgan Kaufmann.

Austin, John L. 1961. A plea for excuses. In *Philosophical Papers*, eds. J. O. Urmson and G. J. Warnock. Oxford: Oxford University Press.

Bacchus, Fahiem, Joseph Y. Halpern, and Hector J. Levesque. 1999. Reasoning about noisy sensors and effectors in the situation calculus. *Artificial Intelligence* 111(1–2): 171–208.

Bach, Emmon. 1989. *Informal Lectures on Formal Semantics*. Albany: State University of New York Press.

Baker, Andrew B. 1989. A simple solution to the Yale shooting problem. In *KR'89: Principles of Knowledge Representation and Reasoning*, eds. Ronald J. Brachman, Hector J. Levesque, and Raymond Reiter, 11–20. San Mateo, Calif.: Morgan Kaufmann.

Baral, Chitta. 1995. Reasoning about actions: Non-deterministic effects, constraints, and qualification. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, ed. Chris Mellish, 2017–2023. San Francisco: Morgan Kaufmann.

Baral, Chitta, and Michael Gelfond. 2000. Reasoning agents in dynamic domains. In *Logic-Based Artificial Intelligence*, ed. Jack Minker, 257–279. Dordrecht: Kluwer.

Baral, Chitta, Sheila McIlraith, and Tran Cao San. 2000. Formulating diagnostic reasoning using an action language with narratives and sensing. In *KR2000: Principles of Knowledge Representation and Reasoning*, eds. Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, 311–322. San Francisco: Morgan Kaufmann.

Belnap, Nuel D. Jr. 1996. Agents in branching time. In *Logic and Reality: Essays on the Legacy of Arthur Prior*, ed. Jack Copeland, 239–271. Oxford: Oxford University Press.

Benferat, Salem, Didier Dubois, and Henri Prade. 1997. Nonmonotonic reasoning, conditional objects, and possibility theory. *Artificial Intelligence* 92(1–2): 259–276.

Besnard, Philippe. 1992. *Default Logic*. Berlin: Springer-Verlag.

Birnbaum, Lawrence. 1991. Rigor mortis: A response to Nilsson's 'Logic and artificial intelligence'. *Artificial Intelligence* 47(1–3): 57–77.

Boolos, George. 1993. *The Logic of Provability*. Cambridge: Cambridge University Press.

Bouquet, Paolo, Luigi Serafini, Patrick Brézillon, Massimo Benerecetti, and Francesca Castellani, eds. 1999. *Modeling and Using Contexts: Proceedings of the Second International and Interdisciplinary Conference, CONTEXT'99*. Berlin: Springer-Verlag.

Boutilier, Craig. 1992. Conditional logics for default reasoning and belief revision. Technical Report KRR–TR–92–1, Computer Science Department, University of Toronto, Ontario.

Boutilier, Craig. 1996. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic* 25(3): 263–305.

Boutilier, Craig, Thomas Dean, and Steve Hanks. 1996. Planning under uncertainty: Structural assumptions and computational leverage. In *New Directions in AI Planning*, 157–171. Amsterdam: IOS Press.

Brachman, Ronald J., Hector J. Levesque, and Raymond Reiter, eds. 1989. *KR'89: Principles of Knowledge Representation and Reasoning*. San Mateo, Calif.: Morgan Kaufmann.

Brachman, Ronald J., Deborah L. McGuinness, Peter F. Patel-Schneider, and Lori A. Resnik. 1991. Living with Classic: When and how to use a Kl-One-like language. In *Principles of Semantic Networks*, ed. John F. Sowa, 401–456. San Mateo, Calif.: Morgan Kaufmann.

Brewka, Gerhard. 1991. *Nonmonotonic Reasoning: Logical Foundations of Commonsense*. Cambridge: Cambridge University Press.

Brewka, Gerhard, Jürgen Dix, and Kurt Konolige. 1997. *Nonmonotonic Reasoning: An Overview*. Stanford: CSLI Publications.

Buchanan, Bruce, and Edward H. Shortliffe. 1984. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Reading, Mass.: Addison-Wesley.

Burger, Wilhelm, and Bir Bhanu. 1992. *Qualitative Motion Planning*. Dordrecht: Kluwer.

Burgess, John P. 1984. Basic tense logic. In *Handbook of Philosophical Logic, Volume II: Extensions of Classical Logic*, eds. Dov Gabbay and Franz Guenther, 89–133. Dordrecht: D. Reidel.

Burkhard, Hans-Dieter, Markus Hannebauer, and Jan Wendler. 1998. Belief-desire-intention deliberation in artificial soccer. *AI Magazine* (3): 87–93.

Buszkowski, Wojciech. 1996. Mathematical linguistics and proof theory. In *Handbook of Logic and Language*, eds. Johan van Benthem and Alice ter Meulen, 683–736. Amsterdam: Elsevier.

Carlson, Greg N., and Francis Jeffrey Pelletier, eds. 1995. *The Generic Book*. Chicago: Chicago University Press.

Carnap, Rudolph. 1955. Meaning and synonymy in natural languages. *Philosophical Studies* 7: 33–47. Reprinted in Carnap (1956), 233–247.

Carnap, Rudolph. 1956. *Meaning and Necessity*. Chicago: Chicago University Press, 2nd ed. (first edition published in 1947).

Casati, Roberto, and Achille C. Varzi. 1996. *Holes and Other Superficialities*. Cambridge, Mass.: MIT Press.

Casati, Roberto, and Achille C. Varzi. 1999. *Parts and Places: The Structures of Spatial Representation*. Cambridge, Mass.: MIT Press.

Chellas, Brian. 1975. Basic conditional logic. *Journal of Philosophical Logic* 4(2): 133–154.

Chen, Su-Shing, ed. 1990. *Advances in Spatial Reasoning*, vol. 1. Norwood, N.J.: Ablex.

Clancey, William J. 1983. The epistemology of a rule-based expert system: a framework for explanation. *Artificial Intelligence* 20: 215–251.

Clark, Keith L. 1978. Negation as failure. In *Logic and Data Bases*, eds. H. Gallaire and Jack Minker, 293–322. New York: Plenum Press.

Clarke, Bowman L. 1981. A calculus of individuals based on 'connection'. *Notre Dame Journal of Formal Logic* 22(3): 204–218.

Clarke, Bowman L. 1985. Individuals and points. *Notre Dame Journal of Formal Logic* 26(1): 61–75.

Cohen, Philip R., and Hector J. Levesque. 1990. Intention is choice with commitment. *Artificial Intelligence* 42(3): 213–261.

Cohn, Anthony G. 1996. Qualitative spatial representation and reasoning techniques. In *KI-97, Advances in Artificial Intelligence*, eds. Gerhard Brewka, Christopher Habel, and Bernhard Nebel, 1–30. Berlin: Springer-Verlag.

Cohn, Anthony G., Brandon Bennett, John Gooday, and Nicholas M. Gotts. 1997. Representing and reasoning with qualitative spatial relations. In *Spatial and Temporal Reasoning*, ed. Oliviero Stock, 97–134. Dordrecht: Kluwer.

Cohn, Anthony G., Fausto Giunchiglia, and Bart Selman, eds. 2000. *KR2000: Principles of Knowledge Representation and Reasoning*. San Francisco: Morgan Kaufmann.

Cohn, Anthony G., Lenhart Schubert, and Stuart C. Shapiro, eds. 1998. *KR'89: Principles of Knowledge Representation and Reasoning*. San Francisco: Morgan Kaufmann.

Copeland, B. Jack. 1996. Arthur Prior's life and legacy. In *Logic and Reality: Essays on the Legacy of Arthur Prior*, ed. Jack Copeland, 1–40. Oxford: Oxford University Press.

Davis, Ernest. 1991. *Common Sense Reasoning*. San Francisco: Morgan Kaufmann.

Davis, Martin. 1988. Mathematical logic and the origin of modern computers. In *The Universal Turing Machine: A Half-Century Survey*, ed. Rolf Herkin, 149–174. Oxford: Oxford University Press.

DeJong, Gerald D. and Scott W. Bennett. 1989. Permissive planning: Extending classical planning to uncertain task domains. *Artificial Intelligence* 89(1–2): 173–217.

Delgrande, James P. 1998. Conditional logics for defeasible logics. In *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 2*, eds. Dov M. Gabbay and Philippe Smets, 135–174. Dordrecht: Kluwer.

Dennett, Daniel. 1987. Cognitive wheels: The frame problem of AI. In *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, ed. Zenon Pylyshyn, 41–64. Norwood, N.J.: Ablex.

Dowty, David R. 1977. Toward a semantic analysis of verb aspect and the English 'imperfective' progressive. *Linguistics and Philosophy* 1(1): 45–77.

Doyle, Jon. 1979. A truth maintenance system. *Artificial Intelligence* 12(3): 231–272.

Doyle, Jon, Erik Sandewall, and Pietro Torasso, eds. 1994. *KR'94: Principles of Knowledge Representation and Reasoning*. San Francisco: Morgan Kaufmann.

Doyle, Jon, and Richmond H. Thomason. 1999. Background to qualitative decision theory. *AI Magazine* 20(2): 55–68.

Elkan, Charles. 1991. Reasoning about action in first-order logic. In *Proceedings of the Conference of the Canadian Society for Computational Studies of Intelligence (CSCSI)*, 221–227. Canadian Society for Computational Studies of Intelligence, San Francisco: Morgan Kaufman.

Elkan, Charles. 1995. On solving the qualification problem. In *Working Notes of the AAAI Spring Symposium on Extending Theories of Action: Formal Theories and Applications*. Menlo Park, Calif.: American Association for Artificial Intelligence.

Fagin, Ronald, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. 1995. *Reasoning about Knowledge*. Cambridge, Mass.: MIT Press.

Fagin, Ronald, Joseph Y. Halpern, and Moshe Y. Vardi. 1984. A model theoretic analysis of knowledge. In *Proceedings of the Twenty-Fifth Annual Symposium on the Foundations of Computer Science*, 268–278.

Fensel, Dieter, Fausto Giunchiglia, Deborah McGuinness, and Mary-Anne Williams, eds. 2002. *KR2002: Principles of Knowledge Representation and Reasoning*. San Francisco: Morgan Kaufmann.

Fikes, Richard. 1996. Ontologies: What are they, and where's the research? In *KR'96: Principles of Knowledge Representation and Reasoning*, eds. Luigia Carlucci Aiello, Jon Doyle, and Stuart Shapiro, 652–654. San Francisco: Morgan Kaufmann.

Finger, Jeffrey J. 1987. *Exploiting Constraints in Design Synthesis*. Ph.D. diss., Department of Computer Science, Stanford University, Stanford, California.

Fodor, Jerry A. 1987. Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, ed. Zenon Pylyshyn, 139–149. Norwood, N.J.: Ablex.

Forbus, Kenneth D. 1988. Qualitative physics: Past, present, and future. In *Exploring Artificial Intelligence*, ed. Howard E. Shrobe, 239–296. San Mateo, Calif.: Morgan Kaufmann.

Forbus, Kenneth D., Paul Nielsen, and Boi Faltings. 1991. Qualitative spatial reasoning: The CLOCK project. *Artificial Intelligence* 51(1–3): 417–471.

Ford, Kenneth M., and Zenon Pylyshyn, eds. 1996. *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*. Norwood, N.J.: Ablex.

Gabbay, Dov M. 1995. Conditional implications and non-monotonic consequence. In *Conditionals: From Philosophy to Computer Science*, eds. Gabriella Crocco, Luis Fariñas del Cerro, and A. Herzig, 337–359. Oxford: Oxford University Press.

Gabbay, Dov, Christopher J. Hogger, and J. A. Robinson, eds. 1994. *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*. Oxford: Oxford University Press.

Galton, Anthony. 1997. Space, time, and movement. In *Spatial and Temporal Reasoning*, ed. Oliviero Stock, 321–352. Dordrecht: Kluwer.

Gärdenfors, Peter, and David Makinson. 1994. Nonmonotonic inferences based on expectations. *Artificial Intelligence* 65(2): 197–245.

Geanakopolos, John. 1994. Common knowledge. In *Handbook of Game Theory*, vol. 2, eds. Robert Aumann and S. Hart. Amsterdam: Elsevier.

Geffner, Hector. 1990. Causal theories of nonmonotonic reasoning. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, eds. Thomas Dietterich and William Swartout, 524–530. American Association for Artificial Intelligence, Menlo Park, Calif.: AAAI Press.

Geffner, Hector. 1992. *Default Reasoning: Causal and Conditional Theories*. Cambridge, Mass.: MIT Press.

Gelfond, Michael, and Vladimir Lifschitz. 1998. Action Languages. *Electronic Transactions on AI*, available at http://www.ep.liu.se/rs/cis/1998/016.

Genesereth, Michael, and Nils J. Nilsson. 1987. *Logical Foundations of Artificial Intelligence*. San Mateo, Calif.: Morgan Kaufmann.

Ginsberg, Matthew L., ed. 1987. *Readings in Nonmonotonic Reasoning*. Los Altos, Calif.: Morgan Kaufmann.

Giunchiglia, Enrico, G. Neelakantan Kartha, and Vladimir Lifschitz. 1997. Representing action: Indeterminacy and ramifications. *Artificial Intelligence* 95(2): 409–438.

Giunchiglia, Enrico, and Vladimir Lifschitz. 1998. An action language based on causal explanation. In *Proceedings of the Fourteenth National Conference on Artificial Intelligence and the Ninth Innovative Applications of Artificial Intelligence Conference*, eds. Ted Senator and Bruce Buchanan, 623–628. Menlo Park, Calif.: AAAI Press.

Glasgow, Janice, N. Hari Narayanan, and B. Chandrasekaran, eds. 1995. *Diagrammatic Reasoning*. Cambridge, Mass.: MIT Press.

Golden, Keith, and Daniel Weld. 1996. Representing sensing actions: The middle ground revisited. In *KR'96: Principles of Knowledge Representation and Reasoning*, eds. Luigia Carlucci Aiello, Jon Doyle, and Stuart Shapiro, 174–185. San Francisco: Morgan Kaufmann.

Goodman, Nelson. 1946. *Fact, Fiction and Forecast*. Cambridge, Mass.: Harvard University Press, 4th ed.

Gotts, N. M. 1994. How far can we 'C'? Defining a doughnut using connection alone. In *KR'94: Principles of Knowledge Representation and Reasoning*, eds. Jon Doyle, Erik Sandewall, and Pietro Torasso, 246–257. San Francisco, Calif.: Morgan Kaufmann.

Gotts, N. M. 1996. Topology from a single primitive relation: Defining topological properties and relations in terms of connection. Technical Report 96.24, School of Computer Studies, University of Leeds.

Guha, Ramanathan V. 1991. Contexts: A formalization and some applications. Technical Report STAN-CS-91-1399, Stanford Computer Science Department, Stanford, California.

Gustaffson, Joakim, and Patrick Doherty. 1996. Embracing occlusion in specifying the indirect effects of actions. In *KR'96: Principles of Knowledge Representation and Reasoning*, eds. Luigia Carlucci Aiello, Jon Doyle, and Stuart Shapiro, 87–98. San Francisco: Morgan Kaufmann.

Halpern, Joseph Y., ed. 1986. *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the First Conference* (*TARK 1986*). Los Altos, Calif.: Morgan Kaufmann.

Halpern, Joseph Y., and Yoram Moses. 1985. Towards a theory of knowledge and ignorance. In *Logics and Models of Concurrent Systems*, ed. Krzysztof R. Apt, 459–476. Berlin: Springer-Verlag.

Hammer, Eric M. 1995. *Logic and Visual Information*. Stanford, Calif.: CSLI.

Hanks, Steven, and Drew McDermott. 1985. Temporal reasoning and default logics. Technical Report YALEU/CSD/RR#430, Department of Computer Science, Yale University, New Haven, Connecticut.

Hanks, Steven, and Drew McDermott. 1986. Default reasoning, nonmonotonic logics and the frame problem. In *Proceedings of the Fifth National Conference on Artificial Intelligence*, eds. Tom Kehler and Stan Rosenschein, 328–333. American Association for Artificial Intelligence, Los Altos, Calif.: Morgan Kaufmann.

Hanks, Steven, and Drew McDermott. 1987. Non-monotonic logics and temporal projection. *Artificial Intelligence* 33(3): 379–412.

Haugh, Brian. 1987. Simple causal minimization for temporal persistence and projection. In *Proceedings of the Seventh National Conference on Artificial Intelligence*, eds. Kenneth Forbus and Howard Shrobe, 218–223. Menlo Park, Calif.: AAAI Press.

Hintikka, Jaakko. 1962. *Knowledge and Belief*. Ithaca, N.Y.: Cornell University Press.

Hintikka, Jaakko. 1970. Objects of knowledge and belief: Aquaintances and public figures. *Journal of Philosophy* 67(21): 869–883.

Hintikka, Jaakko. 1986. Reasoning about knowledge in philosophy: The paradigm of epistemic logic. In *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the First Conference*, ed. Joseph Y. Halpern, 63–80. Los Altos, Calif.: Morgan Kaufmann.

Hobbs, Jerry, Mark Stickel, Douglas Appelt, and Paul Martin. 1993. Interpretation as abduction. *Artificial Intelligence* 63(1–2): 69–142.

Horty, John F. 1994. Some direct theories of nonmonotonic inheritance. In *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*, eds. Dov Gabbay, Christopher J. Hogger, and J. A. Robinson, 111–187. Oxford: Oxford University Press.

Horty, John F. 2001. *Agency and Deontic Logic*. Oxford: Oxford University Press.

Israel, David J. 1991. A short sketch of the life and career of John McCarthy. In *Artificial Intelligence and Mathematical Theory of Computation: Papers in Honor of John McCarthy*, ed. Vladimir Lifschitz. California: Academic Press.

Iwasaki, Yumi and Herbert Simon. 1986. Causality in device behavior. *Artificial Intelligence* 29(1): 3–32.

Kapur, Deepak and Joseph L. Mundy. 1988. Geometric reasoning and artificial intelligence: Introduction to the special volume. *Artificial Intelligence* 37(1–3): 1–11.

Konolige, Kurt. 1986. What awareness isn't: A sentential view of implicit and explicit belief. In *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the First Conference*, ed. Joseph Y. Halpern, 241–250. Los Altos, Calif.: Morgan Kaufmann.

Konolige, Kurt. 1988. On the relation between default and autoepistemic logic. *Artificial Intelligence* 35(3): 343–382. (See also errata, *Artificial Intelligence* 41(1): 115).

Konolige, Kurt. 1994. Autoepistemic logic. In *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 3: Nonmonotonic Reasoning and Uncertain Reasoning*, eds. Dov Gabbay, Christopher J. Hogger, and J. A. Robinson, 217–295. Oxford: Oxford University Press.

Konolige, Kurt, and Martha Pollack. 1993. A representationalist theory of intention. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, ed. Ruzena Bajcsy. San Mateo, Calif.: Morgan Kaufmann.

Kosslyn, Stephen Michael. 1990. Visual cognition: Introduction. In *An Invitation to Cognitive Science*. Vol. 2: *Visual Cognition and Action*, eds. Daniel N. Osherson and Howard Lasnik, 3–4. Cambridge, Mass.: MIT Press.

Kowalski, Robert A., and Marek J. Sergot. 1986. A logic-based calculus of events. *New Generation Computing* 4: 67–95.

Krifka, Manfred, Francis Jeffrey Pelletier, Gregory Carlson, Alice ter Meulen, Gennaro Chierchia, and Godehard Link. 1995. Genericity: An introduction. In *The Generic Book*, eds. Gregory Carlson and Francis Jeffrey Pelletier, 1–124. Chicago: University of Chicago Press.

Kuipers, Benjamin. 1993. Reasoning with qualitative models. *Artificial Intelligence* 59(1–2): 125–132.

Lakemeyer, Gerhard. 1997. Limited reasoning in first-order knowledge bases. *Artificial Intelligence* 71(2): 213–255.

Lascarides, Alex. 1992. The progressive and the imperfective paradox. *Synthèse* 87(6): 401–447.

Laux, Armin, and Heinrich Wansing, eds. 1995. *Knowledge and Belief in Philosophy and Artificial Intelligence*. Berlin: Akedemie Verlag.

Lehmann, Daniel and Menachem Magidor. 1992. What does a conditional knowledge base entail? *Artificial Intelligence* 55(1): 1–60.

Lenat, Douglas B., and R. V. Guha. 1989. *Building Large Knowledge-Based Systems: Representation and Inference in the CYC Project*. Reading, Mass.: Addison-Wesley.

Leśniewski, Stanisław. 1916. Podstawy ogólnej teorii mnogości I. English title: Foundations of a general set theory I.

Levesque, Hector J. 1984. A logic of implicit and explicit belief. In *Proceedings of the Fourth National Conference on Artificial Intelligence*, 198–202. American Association for Artificial Intelligence.

Levesque, Hector J. 1987. Taking issue: Guest editor's introduction. *Computational Intelligence* 3(3): 149–150.

Levesque, Hector, and Gerhard Lakemeyer. 2000. *The Logic of Knowledge Bases*. Cambridge, Mass.: MIT Press.

Levy, Alon Y. 2000. Logic-based techniques in data integration. In *Logic-Based Artificial Intelligence*, ed. Jack Minker, 575–595. Dordrecht: Kluwer.

Lifschitz, Vladimir. 1987. Formal theories of action: Preliminary report. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, ed. John McDermott. Los Altos, Calif.: Morgan Kaufmann.

Lifschitz, Vladimir, ed. 1990a. *Formalizing Common Sense: Papers by John McCarthy*. Norwood, N.J.: Ablex.

Lifschitz, Vladimir. 1990b. Understanding common sense: McCarthy's research in artificial intelligence. In *Formalizing Common Sense: Papers by John McCarthy*, ed. Vladimir Lifschitz, 1–8. Norwood, N.J.: Ablex.

Lifschitz, Vladimir. 1997. On the logic of causal explanation. *Artificial Intelligence* 96(2): 451–465.

Lifschitz, Vladimir. 1998. Situation calculus and causal logic. In *KR'98: Principles of Knowledge Representation and Reasoning*, eds. Anthony G. Cohn, Lenhart Schubert, and Stuart C. Shapiro, 536–546. San Francisco, Calif.: Morgan Kaufmann.

Lin, Fangzhen. 1995. Embracing causality in specifying the indirect effects of actions. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, ed. Chris Mellish, 1985–1991. San Francisco: Morgan Kaufmann.

Lormand, Eric. 1996. The holorobophobe's dilemma. In *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*, eds. Kenneth M. Ford and Zenon Pylyshyn, 61–88. Norwood, N.J.: Ablex.

Łukasiewicz, Jan. 1970. *Jan Łukasiewicz, Selected Writings*, ed. Ludwik Borowski. Amsterdam: North-Holland.

Lukaszewicz, Witold. 1990. *Non-Monotonic Reasoning: Formalization of Commonsense Reasoning*. New York: Ellis Horwood.

Marek, Victor, and Mirosław Truszczyński. 1991. Autoepistemic logic. *Journal of the Association for Computing Machinery* 38(3): 588–619.

Marek, Wictor, and Mirosław Truszczyński. 1989. Relating autoepistemic and default logics. In *KR'89: Principles of Knowledge Representation and Reasoning*, eds. Ronald J. Brachman, Hector J. Levesque, and Raymond Reiter, 276–288. San Mateo, Calif.: Morgan Kaufmann.

Marek, Wictor, and Mirosław Truszczyński. 1994. *Nonmonotonic Logic: Context-Dependent Reasoning*. Berlin: Springer-Verlag.

McCain, Norman, and Hudson Turner. 1995. A causal theory of ramifications and qualifications. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, ed. Chris Mellish, 1978–1984. San Francisco: Morgan Kaufmann.

McCain, Norman, and Hudson Turner. 1997. Causal theories of action and change. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence* *and the Eighth Innovative Applications of Artificial Intelligence Conference*, eds. Howard Shrobe and Ted Senator, 460–465. Menlo Park, Calif.: AAAI Press.

McCarthy, John. 1959. Programs with common sense. In *Proceedings of the Teddington Conference on the Mechanization of Thought Processes*, 75–91. London: Her Majesty's Stationary Office.

McCarthy, John. 1979. First order theories of individual concepts and propositions. In *Machine Intelligence 9*, eds. J. E. Hayes, D. Mitchie, and L. I. Mikulich, 129–148. Chichester, UK: Ellis Horwood.

McCarthy, John. 1980. Circumscription: A form of non-monotonic reasoning. *Artificial Intelligence* 13: 27–39.

McCarthy, John. 1983. Situations, actions, and causal laws. Technical Report Memo 2, Stanford Artificial Intelligence Project, Stanford University.

McCarthy, John. 1986. Applications of circumscription to formalizing common sense knowledge. *Artificial Intelligence* 13: 27–39.

McCarthy, John. 1987. Epistemological problems of artificial intelligence. In *Readings in Nonmonotonic Reasoning*, ed. Matthew L. Ginsberg, 46–55. Los Altos, Calif.: Morgan Kaufmann.

McCarthy, John. 1993a. History of circumscription. *Artificial Intelligence* 59: 23–26.

McCarthy, John. 1993b. Notes on formalizing contexts. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, ed. Ruzena Bajcsy, 555–560. San Mateo, Calif.: Morgan Kaufmann.

McCarthy, John. 1999. Elaboration tolerance. Available at http://www-formal.stanford.edu/jmc/elaboration.html.

McCarthy, John, and Saša Buvač. 1998. Formalizing context (expanded notes). In *Computing Natural Language*, eds. Atocha Aliseda, Rob van Glabbeek, and Dag Westerståhl, 13–50. Stanford, Calif.: CSLI.

McCarthy, John, and Patrick J. Hayes. 1969. Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence 4*, eds. B. Meltzer and D. Michie, 463–502. Edinburgh: Edinburgh University Press.

McDermott, Drew. 1982. Nonmonotonic logic II: Nonmonotonic modal theories. *Journal of the Association for Computing Machinery* 29(1): 33–57.

McDermott, Drew. 1987. Critique of pure reason. *Computational Intelligence* 3(3): 149–160.

McDermott, Drew, and Jon Doyle. 1980. Non-monotonic logic I. *Artificial Intelligence* 13: 41–72.

Meyer, John-Jules C., and Wiebe van der Hoek. 1995. *Epistemic Logic for AI and Computer Science*. Cambridge: Cambridge University Press.

Minker, Jack. 1997. Logic and databases: Past, present and future. *AI Magazine* 18(3): 21–47.

Minker, Jack. 2000a. Introduction to logic-based artificial intelligence. In *Logic-Based Artificial Intelligence*, ed. Jack Minker, 3–33. Dordrecht: Kluwer.

Minker, Jack, ed. 2000b. *Logic-Based Artificial Intelligence*. Dordrecht: Kluwer.

Minsky, Marvin. 1974. A framework for representing knowledge. Technical Report 306, Artificial Intelligence Laboratory, MIT.

Moore, Johanna. 1995. *Participating in Explanatory Dialogues*. MIT Press.

Moore, Robert C. 1985. A formal theory of knowledge and action. In *Formal Theories of the Commonsense World*, eds. Jerry R. Hobbs and Robert C. Moore, 319–358. Norwood, N.J.: Ablex.

Moore, Robert C. 1993. Autoepistemic logic revisited. *Artificial Intelligence* 59(1–2): 27–30.

Moore, Robert C. 1995. *Logic and Representation.* Cambridge: Cambridge University Press.

Morgenstern, Leora. 1996. The problem with solutions to the frame problem. In *The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence*, eds. Kenneth M. Ford and Zenon Pylyshyn, 99–133. Norwood, N.J.: Ablex.

Morgenstern, Leora, and Lynn Stein. 1994. Motivated action theory: A formal theory of causal reasoning. *Artificial Intelligence* 71(1): 1–42.

Nakashima, Hideyuki, Hitoshi Matsubara, and Ichiro Osawa. 1997. Causality as a key to the frame problem. *Artificial Intelligence* 91(1): 37–50.

Nebel, Bernhard, Charles Rich, and William Swartout, eds. 1992. *KR'92: Principles of Knowledge Representation and Reasoning.* San Francisco: Morgan Kaufmann.

Nilsson, Nils J. 1991. Logic and artificial intelligence. *Artificial Intelligence* 47(1–3): 31–56.

Nilsson, Nils J. 1995. Eye on the prize. Available at http://robotics.stanford.edu/~nilsson.

Ohrstrom, Peter, and Per F. V. Hasle. 1995. *Temporal Logic from Ancient Ideas to Artificial Intelligence.* Dordrecht: Kluwer.

Osborne, Martin J., and Ariel Rubenstein. 1994. *A Course in Game Theory.* Cambridge, Mass.: MIT Press.

Osherson, Daniel N., and Howard Lasnik, eds. 1990. *An Invitation to Cognitive Science.* Vol. 2: *Visual Cognition and Action.* Cambridge, Mass.: MIT Press.

Pearl, Judea. 1994. From Adams' conditionals to default expressions, causal conditionals, and counterfactuals. In *Probability and Conditionals: Belief Revision and Rational Decision*, eds. Ellery Eells and Brian Skyrms, 47–74. Cambridge: Cambridge University Press.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.

Perlis, Donald. 1985. Languages with self-reference I: Foundations. *Artificial Intelligence* 25: 301–322.

Pirri, Fiora, and Alberto Finzi. 1999. A preliminary approach to perception in theory of agents. In *Proceedings of the IJCAI-99 Workshop on Nonmonotonic Reasoning, Action and Change*, ed. Michael Thielscher, 49–56. Murray Hill, N.J.: IJCAI, International Joint Conference on Artificial Intelligence.

Pollack, Martha. 1992. The uses of plans. *Artificial Intelligence* 57(1): 43–68.

Pollock, John L. 1995. *Cognitive Carpentry: A Manual for How to Build a Person.* Cambridge, Mass.: MIT Press.

Prior, Arthur. 1956. *Time and Modality.* Oxford: Oxford University Press.

Prior, Arthur. 1967. *Past, Present and Future.* Oxford: Oxford University Press.

Prior, Arthur. 1968. *Papers on Time and Tense.* Oxford: Oxford University Press.

Pylyshyn, Zenon, ed. 1987. *The Robot's Dilemma: The Frame Problem in Artificial Intelligence.* Norwood, N.J.: Ablex.

Quine, Willard V. O. 1960. Variables explained away. In *Selected Logic Papers*, ed. Willard V. Quine, 227–235. Cambridge, Mass.: Harvard University Press.

Rao, Anand S., and Michael P. Georgeff. 1991. Modeling rational agents within a BDI-architecture. In *KR'91: Principles of Knowledge Representation and Reasoning*, eds. James F. Allen, Richard Fikes, and Erik Sandewall, 473–484. San Mateo, Calif.: Morgan Kaufmann.

Reiter, Raymond. 1978. On closed world data bases. In *Logic and Data Bases*, eds. H. Gallaire and J. Minker, 55–76. New York: Plenum Press.

Reiter, Raymond. 1980. A logic for default reasoning. *Artificial Intelligence* 13: 81–32.

Reiter, Raymond. 1993. Proving properties of states in the situation calculus. *Artificial Intelligence* 64: 337–351.

Reiter, Raymond. 2001. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. Cambridge, Mass.: MIT Press.

Renz, Jochen, and Bernhard Nebel. 1999. On the complexity of qualitative spatial reasoning: A maximal tractable fragment of the region connection calculus. *Artificial Intelligence* 108(1–2): 69–123.

Rosenschein, Stanley J. 1989. Synthesizing information-tracking automata from environment descriptions. In *KR'89: Principles of Knowledge Representation and Reasoning*, eds. Ronald J. Brachman, Hector J. Levesque, and Raymond Reiter, 386–393. San Mateo, Calif.: Morgan Kaufmann.

Rosenschein, Stanley J. and Leslie Pack Kaelbling. 1995. A situated view of representation and control. *Artificial Intelligence* 73(1–2): 149–173.

Rounds, William C. 1996. Feature logics. In *Handbook of Logic and Language*, eds. Johan van Benthem and Alice ter Meulen, 475–533. Amsterdam: Elsevier.

Russell, Bertrand. 1957. *Mysticism and Logic*. Garden City, N.Y.: Doubleday.

Russell, Stuart J., and Eric Wefald. 1991. *Do the Right Thing*. Cambridge, Mass.: MIT Press.

Sadek, M. D. 1992. A study in the logic of intention. In *KR'92. Principles of Knowledge Representation and Reasoning: Proceedings of the Third International Conference*, eds. Bernhard Nebel, Charles Rich, and William Swartout, 462–473. San Mateo, Calif.: Morgan Kaufmann.

Sandewall, Eric. 1972. An approach to the frame problem, and its implementation. In *Machine Intelligence 7*, eds. D. Michie and B. Meltzer, 195–204. Edinburgh: Edinburgh University Press.

Sandewall, Erik. 1994. *Features and Fluents: A Systematic Approach to the Representation of Knowledge About Dynamical Systems*. Oxford: Oxford University Press.

Schlechta, Karl. 1997. *Nonmonotonic Logics*. Berlin: Springer-Verlag.

Schubert, Lenhart. 1990. Monotonic solution of the frame problem in the situation calculus; an efficient method for worlds with fully specified actions. In *Knowledge Representation and Defeasible Reasoning*, eds. Henry Kyburg, Ronald Loui, and Greg Carlson, 23–67. Dordrecht: Kluwer.

Seligman, Jerry, and Lawrence S. Moss. 1996. Situation theory. In *Handbook of Logic and Language*, eds. Johan van Benthem and Alice ter Meulen, 239–307. Amsterdam: Elsevier.

Shanahan, Murray. 1997. *Solving the Frame Problem*. Cambridge, Mass.: MIT Press.

Shieber, Stuart. 1992. *Constraint-Based Grammar Formalisms: Parsing and Type Inference for Natural and Computer Languages*. Cambridge, Mass.: MIT Press.

Shoham, Yoav. 1988. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence*. Cambridge, Mass.: MIT Press.

Simon, Herbert. 1952. On the definition of the causal relation. *Journal of Philosophy* 49: 517–528.

Simon, Herbert. 1966. On reasoning about action. Technical Report Complex Information Processing Paper #87, Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

Simon, Herbert A. 1977. *Models of Discovery*. Dordrecht: D. Reidel.

Simon, Herbert A. 1982a. *Models of Bounded Rationality, Volume 1*. Cambridge, Mass.: MIT Press.

Simon, Herbert A. 1982b. *Models of Bounded Rationality, Volume 2*. Cambridge, Mass.: MIT Press.

Simons, Peter. 1987. *Parts: A Study in Ontology*. Oxford: Oxford University Press.

Stalnaker, Robert C. 1993. A note on non-monotonic modal logic. *Artificial Intelligence* 64(2): 183–196. (Widely circulated in manuscipt form, 1980 to 1992.)

Steedman, Mark. 1995. Dynamic semantics for tense and aspect. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, ed. Chris Mellish, 1292–1298. San Francisco: Morgan Kaufmann.

Steedman, Mark. 1998. The productions of time. Unpublished manuscript, University of Edinburgh. Available from http://www.cogsci.ed.ac.uk/~steedman/papers.html.

Stefik, Mark J. 1995. *An Introduction to Knowledge Systems*. San Francisco: Morgan Kaufmann.

Stock, Oliviero, ed. 1997. *Spatial and Temporal Reasoning*. Dordrecht: Kluwer.

Stone, Matthew. 1998. Modality in Dialogue: Planning, Pragmatics and Computation. Ph.D. dissertation, Computer Science Department, University of Pennsylvania, Philadelphia, Penn.

Thielscher, Michael. 1989. Ramification and causality. *Artificial Intelligence* 89(1–2): 317–364.

Thielscher, Michael. 1996. Causality and the qualification problem. In *KR'96: Principles of Knowledge Representation and Reasoning*, eds. Luigia Carlucci Aiello, Jon Doyle, and Stuart Shapiro, 51–62. San Francisco: Morgan Kaufmann.

Thielscher, Michael. 2000. Representing the knowledge of a robot. In *KR2000: Principles of Knowledge Representation and Reasoning*, eds. Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, 109–120. San Francisco: Morgan Kaufmann.

Thomason, Richmond H. 1992. NETL and subsequent path-based inheritance theories. In *Semantic Networks in Artificial Intelligence*, ed. Fritz Lehmann, 179–204. Oxford: Pergamon Press.

Turner, Hudson. 1999. A logic of universal causation. *Artificial Intelligence* 113(1–2): 87–123.

van Benthem, Johan. 1983. *The Logic of Time*. Dordrecht: D. Reidel.

van Benthem, Johan, and Alice ter Meulen. 1996. *Handbook of Logic and Language*. Amsterdam: Elsevier.

van der Torre, Leendert W. N. 1997. *Reasoning about Obligations: Defeasibility in Preference-Based Deontic Logic*. Amsterdam: Thesis Publishers.

von Wright, Georg Henrik. 1983. *Practical Reason: Philosophical Papers, Volume 1*. Ithaca, N.Y.: Cornell University Press.

Weld, Daniel S., and Johan de Kleer, eds. 1990. *Qualitative Reasoning about Physical Systems*. San Mateo, Calif.: Morgan Kaufmann.

Wilson, Randall H. 1998. Geometric reasoning about assembly tools. *Artificial Intelligence* 98(1–2): 237–279.

Wobcke, Wayne, Maurice Pagnucco, and C. Zhang, eds. 1998. *Agents and Multi-Agent Systems—Formalisms, Methodologies, and Applications*. Berlin: Springer-Verlag.

Wolter, Frank, and Michael Zakharyaschev. 2000. Spatio-temporal representation and reasoning based on RCC-8. In *KR2000: Principles of Knowledge Representation and Reasoning*, eds. Anthony G. Cohn, Fausto Giunchiglia, and Bart Selman, 3–14. San Francisco: Morgan Kaufmann.

Woods, William A., and James G. Schmolze. 1992. The KL-One family. In *Semantic Networks in Artificial Intelligence*, ed. Fritz Lehmann, 133–177. Oxford: Pergamon Press.

Yeap, Wai K., and Margaret E. Jeffries. 1999. Computing a representation of the local environment. *Artificial Intelligence* 107(2): 265–301.

18

Indian Logic

1. Introduction

J. N. Mohanty and Amita Chatterjee

In India, logic was never developed as a distinct discipline but was embedded in epistemology. The roots of Indian logic can be traced to two slightly distinct traditions—the *vāda* and the *pramāṇa* traditions. While the former lays down the principles of argumentation to be used in debates, the latter discusses inference as a source of sound knowledge of the world. "On account of this genesis, Indian logic imbibed an epistemological character which was never removed throughout the history" (Matilal 1970).

Indian theories of inference have evoked two diametrically opposite sentiments in the Western mind. On the one hand, we find concerted efforts to translate Indian "logic" into Anglo-European formal logics, like syllogistics or the first-order predicate calculus, and on the other hand, we find statements to the effect that Indian logic is no logic at all. Both these groups are, however, construing "logic" in a very narrow sense. "Logic" to them means the formal deductive theory of inference, which revolves around "the consequence relation" and its properties. But no Indian philosophical system takes a purely formal approach to inference or inferential knowledge. Yet we consider these theories of inference as logic insofar as these are theories of human reasoning and tell us how to distinguish good arguments from bad arguments, acceptable arguments from unacceptable ones.

Since theory of inference in India has been steeped in epistemology, it has developed some unique features not found in Western logic. First, over and above its truth-preserving aspect, an inference must possess a truth-giving aspect. That is, for inference as a way of knowing, it is not enough that the conclusion follows from the premises, the conclusion must also be true. It is expected of such a theory of inference that it lays down conditions of validity/consistency as well as conditions of soundness. In Western formal logic, conditions of validity do not depend in any way on conditions of truth or soundness. So, an inference can yield a true conclusion without being valid, and an inference can be valid even without yielding a true conclusion. In a formal system, syntax can be developed without any reference to semantics. Hence in Western formal logic, semantic considerations are brought in after all syntactic rules are laid down and syntactic consequences are derived. The connection between the syntactic and the semantic sides of a logical system is made by two metatheorems of completeness and soundness. But in Indian theory of inference, syntax always remains hyphenated with semantics. For inference as an accredited source of knowing the world, validity is not enough; soundness and epistemic progress also need to be guaranteed. Besides, in Indian theories of inference the notion of validity/invalidity of an inference usually presupposes a host of background information and often essentially hinges on them. Two necessary corollaries of this stance are (a) no constituent of an inference can have zero-information content like a tautology, and (b) validity of an inference cannot be delinked from soundness, hence an inference is valid only if it yields a true conclusion.

Second, inference as a cognitive process admits of causal analysis. In Indian theories of inference we find elaborate discussions on how inference results from a number of cognitive states, and what conditions give rise to cognitive certainty. Here the relation between premise and conclusion is viewed not as an abstract logical relation but as a psychocognitive relation of causal sequence. Consequently, inferential necessity is understood in terms of a deterministic knowledge machine (Sarkar 1997) of which human cognitive processes, including logical thinking, is a representative model. Thus, we find, psychology enters the territory of logic. Yet no system of Indian logic is open to the charge of "psychologism," because psychological conditions involved in the causal process are not unique to any individual, but are possessed by all those who infer.

Third, all Indian logicians adopted a grammar-based model of logical analysis, while in Western logic the geometrico-mathematical model is in use. But interestingly, this grammar-based model has led the Naiyāyika-s to the insight of mathematical logicians. They understand general sentences as containing two predicates and not in the manner of traditional Aristotelian logic. Even then, this logic was never considered useful in scientific endeavors because of its inextricable reference to the knowing subject.

Fourth, Western formal logic is extensional. Indian logic, it has been said, is basically logic of properties and hence intensional. But what Indian logicians mean by "property" is somewhat different from its meaning in English. The term "property" here signifies any locatee, be it an abstract property or a concrete object, which resides in a locus. So the basic combination in Indian logic is not a straightforward subject-predicate proposition but a Sanskrit sentence of the locus-locatee model, for example, "a has f-ness." But as the sentences of the form "a has f-ness" can be easily correlated with the sentence of the form "a is f," it is possible to read such logic of properties extensionally.

Fifth, there has been lots of debate in recent times over the question of whether Indian philosophers admit propositions and what should be taken as the truth-bearer. Though no general consensus has been reached, we think that the content of episodic cognition is the most suitable to be the truth-bearer. Indian logicians generally view the content of cognition as a relational complex referring to a complex object and not to a fact.

Sixth, though ontological and epistemological commitments underlying different Indian logical systems are diverse and there are controversies centering the number of constituents (*avayava*) in an inference, every system accepts at least three inferential components. These are *sādhya* (the provable property, the probandum, the signified), *hetu* (the ground of inference, the reason, the sign, the probans), and *pakṣa* (the locus of inference). It will be evident from the following sections that Indian systems of logic were basically logic of terms but essentially different from syllogistic. But interestingly, there are fragments indicating their interpretation of the so-called propositional connectives or the sentence-forming operators, of which different varieties of negation have been the subject of the most elaborate discussions. The subtle distinctions that Indian logicians have made between absence and difference and different readings of the principle of double negation that they have offered are being appreciated by modern logicians and computer scientists.

Finally, Indian theories of inference cannot be neatly categorized as deductive or inductive in the standard sense. Ancient Indian logicians were, in fact, trying to formulate conditions of human reasoning in general. They were trying to determine under what conditions an inferential leap from the known to the unknown would be warranted. As Sarkar (1997) rightly points out,

> The problem that Indian logicians were concerned with was neither the development of a formal theory of deductive or syllogistic reasoning, nor was it the problem of induction as we understand it now, nor even the problem of how to make a palatable cocktail of the two. Their real concern centered around the problem of selecting the right sort of projection-base and of framing appropriate rules for distinguishing between projectable and nonprojectable properties.

So Indian theories of inference are primarily theories of adequate evidence, but they may also be viewed as systems of nonmonotonic reasoning, which is being used in modern computer simulation of actual human reasoning processes.

2. Nyāya (Old) Logic

S. R. Saha

Nyāya philosophy is based on the aphorisms called *Nyāya-sūtra* of Gautama who is believed to have compiled them in the second century after Christ.[1] In his first aphorism he mentions 16 kinds of things beginning with *pramāṇa*. By *pramāṇa* Gautama understands the means or instrument for *pramā*. *Pramā* is knowledge and is obtainable through either perception or inference or comparison or testimony. *Prameya* or knowable, strictly speaking, is that thing whose ignorance or misconception leads to suffering and whose knowledge yields freedom from suffering. The most important of these knowable entities is *ātman* or self. The other knowables are related to inference, which is the second *pramāṇa* of the four admitted by Gautama. One of them is "limb" (*avayava*) or constituent of an argument (*nyāya*). In a sense Gautama has developed a new science (*śāstra*) about *nyāya* in fullness and that gives his system of philosophy the name "Nyāya"; this suggests that of the various logical and epistemological concerns in Nyāya, varieties of inference, structure of *nyāya* or arguments and pseudo-probans constitute the most essential features of Nyāya (old) logic. We shall not cover that part of Gautama's philosophy that deals with distinction between the different forms of debate (*vāda*, *jalpa*, and *vitaṇḍā*) and the related topics of quibbles (*chala*), futile rejoinders (*jāti*), and defeat situation (*nigrahasthāna*).[2]

2.1.

Gautama devotes three *Nyāyasūtras* (NS 1/1/5 and 2/1/37–38) to inference. NS 1/1/5 gives both a definition and a classification of inference. In Sanskrit the aphorism runs as the following.

Atha tatpūrvakam trividham anumānam pūrvavat śeṣavat sāmānyatodṛṣṭaṃca. The second word [the first word *Atha* (meaning after perception) is not considered here] of NS 1/1/5 is *tatpūrvakam* and is supposed to give Gautama's definition of *anumāna*. The remaining words are *trividham anumānam* (the word *anumāna* occurs only once) *pūrvavat śeṣavat sāmānyatodṛṣṭaṃca*. The word *anumānam* in this cluster is intended to be understood in relation to the word for the definiens (*tatpūrvakam*) and also for enabling us to get the full proposition of the classification by relating it to the word *trividham*. The word *tatpūrvakam* may be understood as that which has that (*tat*, that is, perception that has been discussed in the preceding *sūtra*) as its (causal) antecedent (*pūrvakam*). The definition thus states that inference (*anumāna or anumiti*) is that cognition (*jñāna*) which is due to (*pūrvakam*) the former (*tat*) as its causal antecedent and which is veridical in nature (*avyabhicāri*) (the words *jñāna* and *avyabhicāri* being brought forward from the preceding aphorism).

Inference, according to the master, is directly based on perception in the paradigmatic cases, but indirectly on it in special cases where it may be based on other perception-based types of knowledge. But perception of what kinds of things can act as such a basis? There is no mention of this in Gautama's aphorism. Vātsyāyana thus offers clarification in his commentary. He uses two words in this context, which require some explanation. The words are *liṇga* and *liṇgin*. The etymological meaning of *liṇga* is that a *liṇga* is a (natural) sign and the *liṇgin* is what is signified by it. In the familiar example of inference from smoke to fire, smoke is the *liṅga* and fire is the *liṅgin* because perception of the *liṅga* in a given place leads to knowledge of fire in that place. Perception of the *liṅga* is called *liṅga*-darśana, the latter component *darśana* here means perception. This causal factor, however, is inadequate by itself. It can lead to the desired inferential knowledge, if it is also known to the person that there is an appropriate kind of relationship (*sambandha*) between the *liṅga* and the *liṅgin*. Such a relationship might have been, as a matter of fact, known much earlier through perception. But on the occasion of the inference when a person comes to have the knowledge of the *liṅga*, that earlier knowledge of the relationship between the *liṅga* and the *liṅgin* gets revived leading to memory of that relationship.[3] Thus, Vātsyāyana thinks that Gautama has offered a causal definition of inference in NS 1/1/5 as he has given us a causal definition of perception in NS 1/1/4.

Gautama explicitly mentions that inference is threefold (*trividha*) and also names its three varieties in NS 1/1/5, although he does not give their examples. In NS 2/1/37 he records an objection against the claim of soundness of inference and he offers his reply in the following aphorism. He uses words there, which give us an idea of the inference whose soundness is being challenged. Vātsyāyana did certainly have these aphorisms in mind while writing his commentary on the earlier aphorism (NS 1/1/5). But he gave therein examples of the three varieties, although all of these cannot be constructed from the words used by Gautama in the two aphorisms (NS 2/1/37–38) under reference. In other words, Vātsyāyana has certainly offered innovative interpretation of NS 1/1/5.

The words for the three varieties mentioned in NS 1/1/5 are (a) *pūrvavat*, (b) *śeṣavat*, and (c) *sāmānyatodṛṣṭa*. These three words are all compound words. Compound words in Sanskrit are often amenable to different interpretations. The clue to the meaning of a word ending with *vat* [as in (a) and (b)] lies (a) in deciphering the meaning of (i) *vat* and also of (ii) the other component of the compound and (b) in determining how the two are to be related to one another in the compound. The resultant multiplicity of alternatives makes it difficult to understand the exact meanings of the compound words under reference. Added to this ambiguity in respect of the component *vat* there is another problem, which makes the situation more complicated, and that problem relates to the fact that the first components of the compound words in question are also quite ambiguous in meaning. Thus, the word *pūrva* in *pūrvavat* may be understood simply as the preceding or in the stronger sense as the cause. Again, the word *śeṣa* in *śeṣavat* may be understood as the succeeding or as the remainder (after elimination of other possible alternatives in a given context).

So far as the third word *sāmānyatodṛṣṭa* is concerned, the problem may not be that acute. But this word may be understood in the negative sense to include all inferences, which are neither *pūrvavat* nor *śeṣavat*, as we shall see later. Or should we assign a positive meaning to the compound as Vātsyāyana seems to have done? If we assign negative meaning to the third compound, the scheme of classification promises to be exhaustive or complete in the sense that if an instance of inference does not merit inclusion under any of the first two types, it can surely be accommodated under the third type. But in case we assign a positive meaning to the third compound while doing the same for the first two, the classification will pose a problem because a given instance may seem to defy the types thus threatening the claim of exhaustive character of the scheme of classification.

Vātsyāyana proposed two interpretations, in both of which all the three compound words figuring in Gautama's aphorism seem to have been assigned positive meanings. His expectation was perhaps that if the first interpretation leaves out any example outside the scheme then the second interpretation would not. But it is advisable now to take a look at the examples themselves and at Vātsyāyana's interpretations.[4]

Vātsyāyana's first interpretation treats both *pūrvavat* and *śeṣavat* as adjectives with respect to *anumāna* (inference) and the word *pūrva* is to be understood in the sense of cause and the word *śeṣa* in the sense of effect. Hence, an inference is of the *pūrvavat* type in which from the perception of the cause one infers the effect, for example, when one infers that there will be rain by perceiving clouds in the sky. An inference is of the *śeṣavat* type in which from the perception of the effect one infers its cause, for example, when one infers that there has been rain earlier by noticing that the river is full of water and the current is moving very fast. In both the types of inference, the relationship between the sign (*liṅga*) and the signified is a matter of regular sequence amenable to observation and the members of the sequence themselves are such that they belong to different segments of time. Thus, from the perception of a present phenomenon one infers the future in *pūrvavat* inference (from cause to effect). In the *śeṣavat* inference one infers the existence of the cause, which belongs to the past on the basis of the effect, which belongs to the present. It is to be noted that the sign and the signified are not both things that belong to the present. Therefore, if in an inference both are things of the present time it cannot be either of the *pūrvavat* or of the *śeṣavat* variety, such an inference then is of the third variety. Thus Uddyotakara in his gloss on Vātsyāyana's commentary on NS 1/1/5 mentions an example in which one infers the presence of water nearby on seeing cranes in the vicinity. In explanation of the negative meaning Tarkavagisa, in his *Nyāyadarśana* has quoted from Gadādhara's gloss on Raghunātha's *Anumitidīdhiti* in support of this negative meaning. He has himself offered an example in which on the basis of knowledge that earthen-ness is pervaded by substance-hood one infers about an earthen object that it is a substance—an inference in which both the sign and the signified are eternal entities and hence are not events of a sequence.[5]

Vātsyāyana is, however, silent about this negative meaning of *sāmānyato-dṛṣṭa* and also about whether the foregoing examples should be subsumed under this variety. In fact, he makes an altogether different stipulation here, as indicated by the example chosen by him in which one infers that the sun moves. But that the sun moves is never a datum to anybody's perception, although the sign that may lead one to that inference is something, which is given in perception. Therefore, the relationship between the possible sign and the movement of the sun also will not be a perceptual datum. If one believes that the sun moves as people of that time used to believe, his belief will have to be the result of some inference, the *liṇga* or sign which must be such that it is something perceptible and must be so in respect of the sun. It is suggested that since position of the sun in the sky is something one can observe, as we all see that the sun rises in the eastern horizon and sets in the western horizon and stays just above our head at noon, we can search for an explanation of this phenomenon. But the varying positions of the sun in the sky that are perceptible and the movement of the sun (which is not perceptible) cannot be shown to be related on the basis of perception. Hence, the relationship between the sign and the signified has to be conceived and formulated for the purpose of the inference in such a way that it does not involve any reference to the sun. Instead of taking the movement of the sun as one of the terms of the relationship we shall have to take it in a general way (*sāmānyataḥ*) as the movement of a body for making the relationship amenable to perception (*dṛṣṭa*). Consequently, we shall have to drop the reference to the sun on the side of the sign too, thus changing it to "varying positions of a body" from "varying positions of the sun." The required relationship between the sign and the signified may accordingly be stated as that the fact of varying positions of a body must be associated with the movement of that body. This relationship, it is claimed, can be seen with respect to many known things, for example, an animal. On the strength of this and on the basis of perception of varying positions of the sun in the sky, one can legitimately infer that the sun moves from one place to another. We thus see that Vātsyāyana here clearly suggests that we assign a positive meaning to the type, and this he shows to be consistent with the etymological meaning of the word. It may be noted here that the etymological meaning is restricted only to the fact that the knowledge of relationship warranting the inference is about a relationship not between a sign conceived as a particular instance and a similarly conceived instance of the signified but between the sign and the signified conceived in a general way. If *sāmānyatodṛṣṭa* is understood only in this etymological sense, it would have to accommodate within its scope the familiar example of inference of fire from smoke. If the relationship were presented as obtaining between a given instance of smoke and a given instance of fire and not between smoke in general and fire in general then it would not have been justified for a person to infer fire in a place where he has perceived a different instance of smoke. In fact a later logician, Nārāyaṇa Bhaṭṭa in his treatise *Mānomeyodaya*, which is a textbook of Mīmāṃsā philosophy, has sought to include this example under the type by accepting only its etymological meaning (*sāmānyata*—in general) as the guide.[6] But as Vātsyāyana makes the additional stipulation that the signified (*liṇgin*) in such a type of inference is not amenable to perception (like the motion of the sun, which is imperceptible to us), we cannot include the inference of fire from smoke under this type. And because of similar considerations we shall not be allowed to include the other two examples (of crane/water and earthen-ness/substance-hood) noted under this variety. Vātsyāyana does of course offer a second interpretation of the tripartite classification, which makes room for such examples.

According to Vātsyāyana's second interpretation, the word *pūrvavat* is not to be treated as adjective of *anumāna* (inference) as in the first interpretation but as an adjective of *liṅga* (sign) and *liṅgin* (signified) although the word *śeṣavat* is to be understood as before as the adjective of *anumāna*. Under this new interpretation the component *vat* in *pūrvavat* means "similar." The relationship between the sign and the signified is something that might have been learned earlier as illustrated in a given case or cases of presence of particular instances of the sign and the signified. If subsequently a thing of the kind of the sign that figured in the previous knowledge of the relationship comes to be noticed at a place then the person can infer the presence of a thing of the kind to which the instance of the signified known earlier belongs. Vātsyāyana illustrates such a *pūrvavat* case by reference to inference of fire from smoke. It seems to us that under this interpretation of *pūrvavat* the relationship is in fact between the sign in general and the signified in general. So understood the *pūrvavat* variety can even include all cases of inference where relationship is not understood as obtaining between the cause and the effect. It will also include all cases of causal type inference either from cause to effect or from effect to cause that have been mentioned in respect of the first two varieties of the earlier interpretation. A given instance of effect may be due to a given instance of cause, but when it comes to the matter of formulating the causal relationship involved in the case, the relationship has to be conceived as that obtaining between types and not between particular instances. If this explanation looks acceptable, the *pūrvavat* variety under the second interpretation will certainly be more inclusive, leaving room for cases of inference under the now vacant *śeṣavat* variety if these do not merit inclusion under the *pūrvavat* or the type.

Vātsyāyana picks up an example from Kaṇāda to which he tags the label *śeṣavat*. He describes the process involved as one in which many alternatives are probable ones in a given context, although only one of them will be true while others will be false. Now it may be the case that after considering the alternatives all but one are found unacceptable in the situation. Then the remaining one can be inferred to be the case not on the basis of an autonomous positive reason but on the strength of the supposition that the alternatives are exhaustive in the given context and on finding all but the last (*śeṣa*) one to be untenable. Following Kaṇāda, Vātsyāyana gives here the illustration of the following Vaiśeṣika argument.

Sound is an event in time and is not eternal in nature. It must then be distinct from the types of eternal entities like universal, the relation of inherence and also *viśeṣa* that distinguishes one particular eternal substance from all others. Sound is consequently either a substance or a motion or a quality, all of which may be noneternal in nature. Of these three alternatives, the first one is not tenable as sound is inherent in a single and simple thing, which is not true of a created substance because it has parts in which it inheres. Nor can it be the case that sound is motion. Sound may be the cause of a second similar sound in a series, which the Vaiśeṣika philosopher admits for explaining audibility of sound in all directions by persons with their respective auditory sense organs located in their respective ear cavities while sound originates initially at a given place outside. Although sound can beget sound, motion cannot be said to generate a second motion in a series. Motion is causally associated with separation of the moving body from some substance it is attached to. Such a separation being achieved by the first motion, there is no need to stipulate that a second motion will be necessary for the same purpose of separation of the moving body from that other substance. Hence, two of the three alternatives being eliminated, the third one, namely, that sound is a quality, gets established.

The inference of the type just noted was known to the philosophers of that time who employed many such inferences. Vātsyāyana's second interpretation signifies his recognition of soundness of such inference, and it goes to his credit to have made a place for it within Gautama's scheme of threefold classification. After having rendered the *pūrvavat* type more inclusive and the *śeṣavat* one adequate for the disjunctive type inference, Vātsyāyana treats the variety identically in both the interpretations for making room for inference of a property in a locus on the basis of perception of the sign in that thing and knowledge of relationship between the sign in general and the signified in general. Vātsyāyana, however, gives a different example of the second interpretation of the scheme. For proving the existence of the self as distinct from the body and the senses it is argued that the features of cognition, conation, and so on are qualities, and as quality belongs to a substance, these features also must be features of a substance. This substance is nothing other than the self. Many later thinkers have challenged the claim of an independent and distinct character of such an inference as it involves, at least in part, *śeṣavat* model. But many others come to the defense of Vātsyāyana and hold that there is something in the type of argument under consideration, which is not reducible to any other variety. Phanibhusana in his *Nyāyadarśana* notes that the argument that seeks to prove that the locus of cognition, conation, and so on is the self has been shown by Uddyotakara to be based on the principle of elimination as this locus must be shown to be different from other substances than the self. In reply to the criticism, Phanibhusana mentions that the fact that there is a locus of qualities like cognition and conation cannot be said to have been established by the *śeṣavat* type of inference and this justifies us to say that variety is very much necessary.[7] Matilal has discussed some recalcitrant examples from Buddhist sources, which we are not taking into consideration here.[8]

2.2.

Gautama's theory of inference is closely related with his theory of *Nyāya* (understood here not in the sense of a system of philosophy but as the set of propositions that constitute an argument). In fact, according to Vātsyāyana and others, of the 16 items that constitute the subject of Nyāya philosophy and which have been mentioned by Gautama in his first *Nyāyasūtra*, 14 items are essential elements of Gautama's theory of *Nyāya*, these being mentioned by him after the first 2 items of *pramāṇa* and *prameya*. According to all Nyāya philosophers, *Nyāya*, which is referred to by Gautama by the word *avayava* (the seventh of his 16 items under reference) is related to *anumāna* or inference. The Naiyāyikas of later time admit two varieties of inference, namely, *svārthānumāna*, which is believed to be inference for oneself, and *parārthānumāna*, which is taken to be inference for others. According to them, employment of argument (*Nyāya*) is not necessary for inference for oneself. As thought is believed by some philosophers to be possible without language, inference, too, may be considered possible for oneself without employment of *Nyāya* (which is defined to be a sentence consisting of clauses [*avayavas*]). But when it comes to the question of persuading another (be that an opponent in a debate or some moderator acting as the judge in a debate), employment of an argument through the vehicle of language becomes indispensable.

In *Nyāyasūtra* 1/1/32 Gautama mentions the five *avayava-s* (lit., "limbs") or constituents of a *Nyāya* and proceeds to define them in the following seven aphorisms. In his commentary Vātsyāyana remarks that many logicians recognize five more constituents than those mentioned by Gautama. In later times, Buddhist and Jaina logicians have tried to make further reduction by admitting between them the first three while the Bhaṭṭas admit the first three or the last three of Gautama's five constituents of an argument. Parsimony apart, Gautama's proposal was accorded approval by these philosophers as between themselves they accepted all the five to be good and acceptable candidates as argument constituents. When we consider the fact that these five constituents have been proposed not as independent sentences but as ones that ought to be taken in relationship with one another, the differences between Nyāya and other systems may not be as great as these have been made to appear. After all if a perspicuous articulation of the steps involved in an argument is considered desirable, the explicitly stated constituents of *Nyāya* seem to be quite all right. We shall be following Vātsyāyana in giving illustrations of the argument constituents.

The first step in the argument is called *pratijñā*. It is the statement of the proposed thesis. A Naiyāyika, who believes in the impermanence of sound as against other philosophers, according to whom it is eternal may announce the thesis he proposes:

(1) Sound is noneternal.

That sound is noneternal is to be proved or established (i.e., *sādhya*). This thesis itself is thus called *sādhya* or probandum. In the context of a debate this thesis has also been referred to by Gautama as *pakṣa* (thesis or position) to be proved and its opposite by the word *pratipakṣa* (counterthesis). Vātsyāyana however draws our attention to the fact that Gautama also uses the word *sādhya* for the property that is sought to be established in the thing figuring as the subject of the proposition (in our example the property of being noneternal will be such a provable property).

Similarly, in Gautama there is an ambiguity also in the use of the word *hetu*. It designates the second sentence in the argument, and it also seems that he would not object to the use of the word to mean a thing the sentence is about. In later writings on logic in India, this word is almost universally used in the sense of *liṅga* or sign that we have come across earlier. Gautama rightly regards the second constituent, which he calls *hetu*, as recording the reason for the proposed thesis. This relationship between the two propositions *pratijña* and *hetu* (the first one requiring proof and the second one offering the reason) is very much essential for the names to be applied to the steps. The second step, however, in relation to the proposed thesis under consideration is as follows.

(2) Because of its being a product.

The proposition as expressed by the sentence has two contents, namely, (i) sound is a product, and (ii) a product is noneternal. The first part involves reference to the first step and represents what was mentioned earlier in NS 1/1/5 as perception of the *liṅga*; the second part, on the other hand, involves reference to the third step that records co-presence of the sign and the signified in some known instance or co-absence of both in some other instance. If the announcement of the proposed thesis in the first step was a matter of mere statement (or what Vātsyāyana calls *āgama* or testimony in a loose sense) the statement of the reason in the second step points toward the causal factors of inference as we have noted. This second argument constituent may thus be compared to the *pramāṇa* or evidential support known as *anumāna* (literally, that which leads to inferential knowledge). But what we have taken to be the second part of the meaning of this argument constituent must receive confirmation in a positive or a negative instance. This confirmation thus points to the perceptual base of the general proposition at work in inference. Gautama defines the *hetu* as the statement of the reason involving reference to the confirming instance and he records next two possible ways of confirmation. Thus, the third step for the argument under consideration is either:

(3) (A product is noneternal, thus) an (earthen) dish (*sthāli*) is a product and is also noneternal;

or

(3′) (A product is noneternal, thus) the soul which is not a product is admitted to be eternal.

This third step is called example or *udāharaṇa*, (3) being *sādharmya udāharaṇa* or positive example recording co-presence of the probans and the probandum, and (3′) being the *vaidharmya udāharaṇa* or negative example in which both are absent. Vātsyāyana explains the fourth step by noting two formulations, one being made consistent with (3) and another with (3′). The formulations are:

(4) Like the (earthen) dish, sound also is a product;

or

(4′) Unlike the soul, sound is a product.

This fourth step records in (4) similarity between the subject of the inferential knowledge and the positive example where both the probandum and the probans have been noticed, or it states in (4′) dissimilarity between this and the negative example where both are absent. Vātsyāyana notes that this step represents knowledge of similarity or of dissimilarity, which is believed to be related to knowledge by comparison or *upamiti*. Hence, in addition to the three accredited sources of testimony, reasoning (*anumāna*) and perception which are taken to be at work in the previous three steps the other accredited source of knowledge (knowledge of similarity or dissimilarity known as *upamāna pramāṇa* is at work in the fourth step, although no knowledge of the nature of *upamiti* results here). This completes the process of convergence of all the sources of knowledge (*pramāṇasamavāya*), and naturally enough if these backings are really there the argument presented through the steps should really be a dependable one not worthy of contradiction.[9] Hence the steps culminate in asserting over again in the fifth step what was initially proposed as the thesis to be proved, beginning with the word "therefore" (*tasmāt*) showing its following or transition (*nigamana*) from the earlier ones. Thus, the subject of the inference is taken to be the seat of the probandum being warranted in its journey by steps (1), (2), (3), and (4) or by steps (1), (2), (3′), and (4′) as the case may be. The fifth step is thus:

(5) Therefore, sound is noneternal (because of its being a product).

Vātsyāyana is of the opinion that this theory of *Nyāya* is to be seen and interpreted in the light of the theory of pseudo-probans (*hetvābhāsa*). Exposing the unsound nature of an argument by showing the presence of pseudo-probans is permissible in each of these three types—debate called *vāda* (dispassionate discussion), *jalpa* (defense of a position) and *vitaṇḍā* (polemical argument), while futile rejoinders and employment of other techniques of refutations such as quibbles (*chala*) and (*jāti*) showing defeat situations (*nigrahasthāna*) are apt only for the last two varieties.[10]

2.3.

Vātsyāyana regards a putative probans as a pseudo-probans (*hetvābhāsa*) if it only looks like a genuine probans but is not really so. Philosophers in the Nyāya tradition believe that the features of a genuine probans are five in number and must be such that if these are all present in a probans it cannot be subsumed under any of the pseudo-probans. But normally every pseudo-probans should have at least someone or the other of such features.

It is customary among philosophers of all schools to regard the following three as properties of a genuine probans:[11]

i. it must be present in the subject of inference (*pakṣavṛttitva*),

ii. it must be present in at least one place where the probandum is known to be present (*sapakṣavṛttitva*), and

iii. it must be absent in all places where the probandum is known to be absent (*vipakṣa-avṛttitva*).

The followers of Nyāya, however, suggest that a genuine probans must have two more marks or properties such as:

iv. the absence of the probandum is not established in the subject of inference on the basis of some other probans (the property called *asatpratipakṣitatva*), and

v. it cannot be the case that an absence of the probandum is known in the subject of inference on the basis of some veridical perception or reliable testimony (the mark called *abādhitatva*).

But the Masters Gautama and Vātsyāyana themselves did not mention these properties. Because absence of these properties can be mapped to someone or the other of the five pseudo-probans admitted by them, it seems to us that they attached primacy to the topic of pseudo-probans.

The Sanskrit names used by Gautama for these pseudo-probans along with English terms within parentheses that we shall be using following B. K. Matilal[12] are (a) *savyabhicāra* (the deviating = D) representing the absence of the property indicated in (iii); (b) *viruddha* (the contradictory = C) representing the absence of the property indicated in (ii); (c) *prakaraṇasama* (the counterbalanced = CB) representing the absence of the property indicated in (iv); (d) *sādhyasama* (the unestablished or the unproven = UE) representing absence of the property indicated in (i); and (e) *kālātīta* (the untimely = U) representing absence of the property indicated in (v). The more familiar words for (c) is *satpratipakṣita*, for (d) is *asiddha*, and for (e) is *bādhita*. Using the abbreviations mentioned within parentheses we give some examples of each of the five types.

D1: Sound is eternal as it is devoid of touch.

D2: This is a cow as it has horns.

C1: Sound is eternal as it is an event in time.

C2: This has no fire in it as it has smoke in it.

CB1: Sound is eternal as it is audible.

CB2: Sound is noneternal as it is an event.

U1: Fire is not hot as it is a substance.

U2: The skull of a dead man is sacred as it is a part of the body of an animal.

UE1: Sound is eternal as it consists of parts.

UE2: The golden mountain has fire as it has smoke.

In later development of Nyāya, philosophers have come up with newer examples, which led to classification of several of the pseudo-probans. Gaṅgeśa who is regarded as the founder of the new movement called Navya Nyāya offered a general definition of pseudo-probans and also definitions of each specific variety.

3. Buddhist Logic

Amita Chatterjee

3.1.

Exploring Buddhist logic is a twentieth-century phenomenon. It started with Satish Chandra Vidyabhusana's studies in the first decade of the twentieth century (1921) and received a boost from Stcherbatsky's *Buddhist Logic*, which appeared in 1930. Ever since, there is a steady flow of articles on the subject.

I begin the exposition of the Buddhist logic proper with their theory of inference as we find in the Buddhist epistemology. Due to space constraint, it will not be possible for me to discuss their science of debate at all.[13] Buddhist logic developed over a period of 1500 years. In course of interactions with other Indian systems, the Buddhists had to change the contour of their logic many times. I shall, however, mainly confine myself to the Diṅnāga-Dharmakīrti tradition covering only a period of 300 years, highlighting en passant the points of difference between the stalwarts.

The Buddhists divided knowable objects into two groups—particular and general—and accordingly admitted two different ways of knowing—perception and inference. Unlike the Naiyāyika-s, they thought that objects of perception cannot be known through inference, nor is it possible to perceive objects of inference. Perception is the means of directly apprehending particulars in their pristine purity, undistorted by the play of fanciful imagination. Inference, on the other hand, is the means of indirectly apprehending objects in general after the mind imposes names, forms and the so-called universals on them. Both perception and inference are supposed to yield certain and correct knowledge of objects. But of these two, perception is considered the means of correct awareness par excellence, because it is free from all possible errors. Error creeps in with conceptualization, with play of imagination. Inference, however, yields certainty when based on adequate evidence. Hence the Buddhist theory of inference is primarily a theory of adequate evidence.

The Buddhists also divide inference into two types—inference for oneself (*svārthānumāna*) and inference for others (*parārthānumāna*). I use SA and PA to refer to these two types of inference henceforth. This distinction was first introduced by Diṅnāga and was adopted by all later logicians. All inferences, according to the Buddhists too, must possess three terms: (a) a logical sign (*hetu*/probans), (b) the signified (*sādhyadharma*/probandum) and (c) the subject-locus (*pakṣa*). When, for example, one infers fire on a hill, seeing smoke coming out of the hilltop, smoke is generally taken to be the logical sign, the hill the subject-locus and fire the signified.

Dharmakīrti (600–660 A.D.), following Diṅnāga, defines SA as knowledge of the signifier in the subject-locus originating from a logical sign having triple character. PA is defined as the expression or statement of the logical sign having triple character. Diṅnāga's definition of PA had two additional clauses, namely, the logical sign must express reality and must be experienced by the person making an inference. The first clause is related to the soundness of inference and the second clause emphasizes the inadequacy of mere hearsay evidence.

The major difference between SA and PA lies in the fact that while SA deals with the psychological conditions, that is, causally connected cognitive states leading to one's own inferential knowledge, PA essentially deals with the proper linguistic expression of this inference with a view to convincing others.

But before taking up SA and PA separately, let me address two questions related to both. These are: When do we infer and what do we infer? The older Naiyāyika-s maintain that doubt or absence of certainty is a necessary precondition of any inference. Diṅnāga and Dharmakīrti point out that only those doubtful facts, which are admitted by the disputants for proof or disproof, lead to inference. This way of explaining the motivation of inference clearly hints at the genesis of inference from the debating tradition.

Logicians differ among themselves regarding what it is that we are supposed to infer. Diṅnāga asks, what actually are we supposed to prove in an inference like "the yonder hill is fiery, since it has smoke?" Is it the hill, the fire, the fire and the hill, the relation between the fire and the hill, the fire characterized by the hill, or the hill characterized by the fire? He rejects five of these alternatives and admits the last one to be the object of inference in this particular case.

3.2.

A person infers for himself to have certain awareness of some object, which he cannot directly apprehend through sense perception. So the SA theory specifies conditions that yield certainty whenever one infers something (*sādhya*) on the basis of an adequate sign (*hetu*) in a particular subject-locus (*pakṣa*). Any property can be a sign for another property, says Diṅnāga, provided (a) the first property is observed at least once and (b) if no instance has been observed where the first is present but the second is absent. A sign to be adequate must, therefore, possess triple characters as follows.

i. A sign must be present in the subject-locus (*pakṣa*) where the signified would be inferred.

ii. It must be present in similar locations (*sapakṣa*) or homologs.

iii. It must not be present in dissimilar locations (*asapakṣa*/*vipakṣa*) or heterologs.[14]

The significance of the theory of triple character will be clear if we apply it in the case of an actual inference. Consider the condensed inference: Sound is impermanent because it is produced by human effort. Here sound is the subject-locus, impermanence is the signified or the object of inference, and the property of being produced by human effort is the logical sign. A similar location or a homolog is similar to the subject-locus in respect of the presence of the signified. So, in this case, a homolog is any location other than sound where impermanence is present, say, a pot. A dissimilar location or a heterolog here would be any permanent entity, for example, an atom,[15] because a dissimilar location has been described as the contradictory or "the other" of the similar location. The Naiyāyika defines a heterolog as that where the absence of the probandum is known to be present.

The language of Diṅnāga's original formulation of the triple character was quite objective, but actually in all three conditions "must be present" should be read as "must be known to be present." Besides, in Diṅnāga's original formulation, we do not get any hint regarding how to quantify *pakṣa*, *sapakṣa*, and *vipakṣa*. He was severely criticized by Uddyotakara for that. Dharmakīrti, therefore, tells us how we are to quantify three different locations in (i), (ii), and (iii). It becomes obvious from *Hetucakraḍamaru* that this reading is not contrary to Diṅnāga's intention. Hence the final formulation of the triple character as pointed out by Dharmottara is as follows.

I. A sign must be known to be present in the entire class corresponding to the subject-locus.

II. A sign must be known to be present in at least one similar location.

III. A sign must never be known to be present in any dissimilar location.

Diṅnāga thinks that these three conditions taken together constitute the necessary condition of a projectible sign. This means *pakṣadharmatā*[16] and *vyāpti*[17] together constitute the ground of a legitimate inference. Logically speaking, if all cases of H are cases of P, then only cases of P could be cases of H, that is, all cases of non-H are cases of non-P, which means no case of H is a case of non-P. In other words, the sign surely leads to the signified, if it is present in the subject-locus and if it is backed by the relation of universal concomitance between the sign and the signified, which is obtainable through positive and negative concomitance.

Dharmakīrti, on the contrary, thinks that either (I) and (II) or (I) and (III) should be sufficient for arriving at an acceptable conclusion. Dharmakīrti's view accords well with the Nyāya position. Besides, strictly from the logical point of view, where the interpretation of negation is standard, (II) and (III) are equivalent and should have the same meaning. It thus appears that Diṅnāga did not subscribe to the principle of double negation; his negation was rather nonstandard.[18]

Diṅnāga's theory of the triple-charactered sign was questioned by the Indian logicians long ago. Śāntarakṣita in his *Tattvasaṃgraha*[19] discusses the objections of a Jaina logician named Pātrasvāmī. Pātrasvāmī says that even a triple-character sign is inadequate because it cannot lend any guarantee that the resulting inference will be acceptable. So he proposes the view that a sign is adequate if it satisfies just one character, that is, "being otherwise impossible" (*anyathānupapannatva*). In this context, he mentions an inference in which the sign allegedly possesses three characters, but the inference is unacceptable. The inference is of the form: He must be dark because he is a son of so and so—like all other sons of his who are found to be dark. Śāntarakṣita answers that the sign in this case does not really fulfill the third condition. It may be true that whoever is a son of this person is dark, but it cannot be shown that whoever is not a son of this person is not dark. Thus Śāntarakṣita reinforces Diṅnāga's view that fulfilment of the first two characters only does not make a sign predictable.

Besides providing epistemic certainty, the third character of a sign has another special significance for Diṅnāga. Dharmakīrti did not consider the third character from this perspective. Diṅnāga extended his theory of signs beyond the theory of inference to the context of language, especially to the famous Buddhist doctrine of *apoha*. Let me tell you very briefly wherein lies the similarity between an inferential sign and a linguistic sign. When someone infers fire on a hill, seeing smoke there, he does not know for sure any particular fire on the hill. He is only sure about one thing—that the hill does not lack fire. Similarly, when somebody hears the word "fire," the word acts as a sign for the object which is not nonfire, that is, "knowledge of the word 'fire' " leads to our knowing the object of reference as excluded from nonfire.[20] That means the sign leads to the knowledge of the signified if we know that the former is excluded from whatever excludes the latter.

3.3.

We have seen that the second and the third character of a logical sign are intimately connected with the ascertainment of universal concomitance between the sign and the signified (*vyāpti*), which everybody admits to be the ground of inference. The Cārvāka-s deny inference as a means of knowing, because they think that the knowledge of the invariable connection between the sign and the signified is not obtainable. The Buddhists assert that one can be sure about the inferential relation, if any of the two following conditions hold: (a) when there is essential identity between the sign and the signified (*tādātmya*), that is, when the inferential relation is a relation of class-inclusion, as in "It is a tree, because it is a *śiṃśapā*"[21]; and (b) when there is a causal connection between the signified and the sign (*tadutpatti*), as in "There is fire here, because there is smoke." Since "All *śiṃśapā*-s are trees" is a necessary identity statement, which not many will try to deny, Dharmakīrti concentrates on the second type of universal statement. In fact, in his *Pramāṇa-vārttika* and *Hetubindu*, Dharmakīrti discusses elaborately on causality. He has written at length to establish that (1) every event has a cause, (2) the same cause produces the same effect, and (3) the relation of causality holds without exception. If one reads between the lines of Dharmakīrti's texts, one gets the impression that Dharmakīrti, unlike the Naiyāyika-s, wants to treat causality as a necessary relation.

Diṅnāga was in favor of the view that causality can be apprehended if one can rule out the contrary possibilities with the help of negative concomitance. It appears from Dharmottara's commentary that Dharmakīrti agreed with Diṅnāga on this point. But in *Pramāṇa-vārttika* Dharmakīrti categorically states that the necessary inferential relation cannot be known only by observation of positive or negative concomitance between the sign and the signified.[22] Prajñākaragupta, therefore, suggests in his commentary that causality is to be apprehended with the help of some extraordinary perception. But this is not Dharmakīrti's view. Dharmakīrti insists that both class inclusion and causality are conceived in imagination (*vikalpabuddhi*). Once the mind constructs this concept, we impose them to derive an orderly world out of our perception of real particulars. So these two relations are unlikely to be derived from our experience; rather, they precede our experience of the orderly world.[23]

3.4.

Logical signs possessing the triple character can be divided into three types: (1) sign in the form of nonapprehension (*anupalabdhi*), (2) sign in the form of essential identity (*svabhāva*), and (3) sign in the form of effect (*kārya*). The first type of sign establishes a negative conclusion, but the second and the third are meant for establishing positive conclusions. I give a list of inferences involving three types of sign to make their connection with the necessary inferential relation clear.

A. There is no pot in this location, because no pot is perceived here (nonapprehension).

B. This is a tree, because it is a *śiṃśapā* (identity).

C. There is fire in this location, because there is smoke here (effect).

Essential identity acts as a sign for deducing a signifier, the nature of which is not different from that of the sign. Sign in the form of an effect is sufficient to deduce the signified in the form of a cause. But the question remains, how is (A) relevant to the ascertainment of the inferential relation? (A) is relevant because according to the Buddhists, the contrary supposition that there is no universal concomitance is disproved by contrary evidence (*viparyaye bādhakapramāṇa vṛttiḥ*). The absence of the pervading property (*vyāpakābhāva*) is, therefore, the logical sign for the absence of the pervaded property (*vyāpyābhāva*). This is, in fact, nothing but the restatement of the third character of a sign.

It is possible to offer other reasons, both ontological and epistemological, behind the Buddhists' acceptance of the sign in the form of nonapprehension. The Nyāya-Vaiśeṣika philosophers admit absence as a separate ontological category. The Naiyāyika-s uphold that the absence of a perceptible thing is perceptible. The Buddhists do not admit absence as a separate ontological category, nor do they admit a separate means of apprehending absence like the Bhāṭṭa Mīmāṃsaka-s. They think that perception of an empty locus leads to the nonperception of a character of the locus, which acts as a sign in the inference of nonexistence of the signified in that particular locus.

Dharmakīrti in *Nyāyabindu* classifies sign in the form of nonapprehension into 11 types, but asserts almost immediately that all the different types are in fact reducible to one, namely, nonapprehension of the essential nature (*svabhāvānupalabdhi*) of an entity. He also mentions that differences are applicational and not essential.

3.5.

According to Dharmakīrti, communicating the sign with the triple character to others is to be called *Parārthānaumāna* (PA). Because PA is meant for convincing others, it has to be expressed in language in a structured way. That is, statements constituting a PA must have specific forms and follow a specific order. *Yuktidīpikā*, a Sāṃkhya text, mentions PA with 10 constituent members. The Naiyāyika-s have reduced the number of members to five, but the Buddhists have gone in for further reduction. They recommend just two constituents—*hetu* or the Reason (the statement containing the logical sign in the subject locus) and *dṛṣṭānta* or the Example (the statement containing a general rule of universal concomitance together with one application):

PA 1. R: The hill has smoke

E: Wherever there is smoke, there is fire, as in a kitchen.

The Buddhists don't even feel the necessity of stating the conclusion. They think that if the sign is a projectible one, then after hearing R and I, the hearer will draw the conclusion right away without waiting for a prompt.[24]

Dharmakīrti first divides PA into two types: (a) those based on similarity (between the example and the subject-locus in respect of the logical sign) and (b) those based on dissimilarity (between the example and the subject-locus in respect of the logical sign). The distinction between these two types of example will be evident from the following examples.

PA 2. R: Sound is a product
E: Whatever is a product is impermanent, for example, a pot.

PA 3. R: Sound is a product
E: Whatever is permanent is not a product, for example, the sky.

In PA 2 there is similarity between sound and a pot in respect of being a product, whereas in PA 3, there is dissimilarity between sound and the sky in the same respect (i.e., sound is a product but the sky is not). PA 2 depends on positive concomitance, and PA 3 depends on negative concomitance. So the former contains a co-instance and the latter contains a counterinstance.

Here Dharmakīrti makes an interesting observation. There is no difference in the intent of these PAs; the difference between them is in formulation only. This observation shows that Dharmakīrti, unlike Diṅnāga, had no qualms over contraposition. That is why, he upholds that

Whatever is a product, is impermanent

and

Whatever is permanent is not a product

mean the same thing and vice versa.

Each of these two types of PA is again subdivided into three subtypes depending on the nature of the logical sign—in the form of essential identity, or in the form of an effect, or in the form of nonapprehension. A detailed discussion of all types of PAs together with suitable examples is available in *Nyāyabindu*.

3.6.

In this section I propose to discuss the fallacies of inference. In the Buddhist texts, the discussion of fallacies is confined to the chapter on PA. But we should not, therefore, think that in case of SAs there is no possibility of error. SAs are also equally vulnerable. In the *Svārthānumānapariccheda* of *Pramāṇa-vārttika*, Dharmakīrti first defines a pseudo-sign or a pseudo-probans. Still all discussions of fallacies have been pushed to the end of the discussion on PA probably because one needs to be more careful while communicating to others. At least this is Dharmottara's explanation.

Inferences, think the Buddhists, may be vitiated by fallacies relating either to the Subject locus, or to the Reason or to the Example. Of these three categories, the Naiyāyika-s are mainly concerned with the fallacies of Reason, for logic or theory of inference is, according to them, the theory of adequate evidence or of legitimate signs. Yet the Buddhists and the Vaiśeṣika-s mentioned possible defects of thesis and example.

In *Nyāyapraveśa*, a sixth-century Buddhist text, allegedly written by Śaṇkarasvāmin, we find a list of possible defects of the thesis that the arguer wants to establish. Later, Dharmakīrti also endorsed them. A subject-locus is considered defective (*paksābhāṣa*/pseudo-minor) under the following conditions:

a. When a thesis is contradicted by perception as in "Sound is inaudible";
b. When a thesis is contradicted by inference, such as, "a pot is permanent";
c. When a thesis contradicts the tradition, such as, when a Vaiśeṣika would pose the thesis that sound is permanent;
d. When a thesis is contradicted by common knowledge, for example, "a human skull is pure" (because it is a part of a living organism like a conch shell);
e. When a thesis is contradicted by one's own statement, such as, "My mother is barren";
f. When one wants to establish a thesis that does not need any proof, such as, "sound is audible."

Fallacies that occur in Example (*dṛṣṭāntābhāsa*/pseudo-exemplar), that is, the statement containing an example are as follows. The first set of fallacies relates to inferences based on similarity. An example is faulty,

1. when the sign is not found in the example, such as, sound is permanent because it is incorporeal like an atom;
2. when the object of inference is not found in the example, for example, sound is permanent because it is incorporeal like cognition;
3. when neither the sign nor the signified is found in the example, such as, sound is permanent because it is incorporeal like a pot;
4. when positive concomitance between the sign and the signified is not properly mentioned in the Illustration, for example, impermanence and the property of being produced are known to reside in a pot;
5. when positive concomitance is mentioned in the wrong order, such as, whatever is impermanent is known to be produced.

We now mention the second set of fallacies of Example that relates to inferences based on dissimilarity. An Example is defective,

6. when the signified is not excluded from the example, for example, sound is permanent because it is incorporeal and whatever is impermanent is known to be corporeal like an atom;
7. when the sign is not excluded from the example, such as, sound is permanent because it is incorporeal and whatever is impermanent is known to be corporeal like an action;
8. when neither the signified nor the sign is excluded from the example, for example, sound is permanent because it is corporeal and whatever is impermanent is known to be corporeal like the sky;
9. when negative concomitance between the sign and the signified is not properly stated, such as, corporeality and impermanence are seen to reside in a pot;

10. when negative concomitance is expressed in wrong order, for example, whatever is corporeal is known to be impermanent.

Let us now look closely at the defects of Reason (*hetvābhāsa*/pseudo-sign/pseudo-probans). We have seen that a sign, according to the Buddhists, is adequate for projection into unknown cases if and only if it possesses the triple character. Hence all defects of sign arise from the absence of one or two of the said triple character. These three characters thus act both as conditions of validity as well as conditions of soundness of an inference. When the first character is violated, we have the pseudo-signs of the unestablished (*asiddha*) variety. The violation of the third character leads to inconclusive (*aniścita*) pseudo-signs, and when both the second and the third are violated we get hostile (*viruddha*) pseudo-signs.

A sign involves defects of the unestablished variety when something is wrong with the relation between the sign and the subject-locus. The first type of unestablished pseudo-sign arises when the sign is known not to be present in the subject locus both by the proponent and by the opponent (*ubhayāsiddhi*), for example, word is impermanent because it is visible.

The second type of unestablished pseudo-sign arises when the opponent does not admit that the sign resides in the subject-locus (*prativādyasiddhi*), for instance, tree is conscious because it dies, when its bark is removed. The Jainas propose this inference and admits the death of a tree due to removal of the bark, but not the Buddhists.

The third type of unestablished pseudo-sign arises when the relation between the sign and the subject-locus is not admitted by the proponent (*vādya-siddhi*), such as, when the Sāṃkhya argues against his opponent that sound is impermanent because it has an origin. For the Sāṃkhya does not believe in the origination of effects but only in manifestation. Diṅnāga and Śaṇkarasvāmin have combined the second and the third varieties together, but Dharmakīrti mentions them separately. Recognition of all these three defects is again a legacy of the debating tradition.

The fourth type of unestablished pseudo-sign arises when the arguer himself has some doubt regarding the nature of the sign actually residing in the subject-locus (*sandigdhāsiddhi*). For example, when a person has doubts whether it is smoke or mist he sees on the hill, he cannot establish the presence of fire on the hill.

The fifth type of pseudo-sign of the same group is known as unestablished due to the nonexistence of the subject-locus (*dharmyasiddhi*/*āśrayāsiddhi*) because there is no possibility of the sign residing in the subject-locus due to its nonexistence. For example, when one who does not admit that the soul exists argues that the soul is a substance because it is a substratum of qualities. The contraposed version of the Buddhists' doctrine of momentariness is also likely to get vitiated by this type of pseudo-sign.

Pseudo-signs of another variety, which we discuss now, are called inconclusive. A sign may be inconclusive under any of the following circumstances:

I.1. When a sign resides both in similar and dissimilar locations, for example, sound is permanent because it is knowable;

I.2. When the sign is unique and resides neither in similar nor in dissimilar locations, for instance, sound is permanent because it is audible;

I.3. When a sign resides in some similar and in all dissimilar locations, such as, sound is not produced by human effort because it is impermanent;

I.4. When a sign resides in all similar and some dissimilar locations, for example, sound is produced by human effort because it is impermanent;

I.5. When a sign resides in some similar and some dissimilar locations, for instance, sound is permanent because it is not amenable to touch.

The next variety of pseudo-signs is called hostile. A sign is hostile if it proves the opposite of the signified in any of the following manners:

H.1. When a sign resides in all dissimilar locations and in no similar location, for example, sound is permanent because it is produced;

H.2. When a sign resides in some dissimilar locations but not in any similar location, for instance, sound is permanent because it is produced by human effort.

In *Nyāyapraveśa*, which was modeled after Diṅnāga's *Nyāyamukha*, we find another entry in the list of inconclusive pseudo-signs. It reads like this: An inconclusive pseudo-sign is that which establishes a set of contradictory results, for example, sound is impermanent because it is produced like a pot; sound is permanent because it is audible like soundness. Dharmakīrti calls this pseudo-sign *viruddhavyabhicārin* which the Naiyāyika-s name *satpratipakṣa*. Literally *viruddhavyabhicārin* means a sign, which is not without a countersign.[25] In this case two signs warrant two different facts, which are opposed to one another. Whenever a person faces such a situation he is in a doubtful frame of mind. Both alternatives being equally cogent, he does not know which one to support. Dharmakīrti objects to this very point. He says that two mutually opposed phenomena cannot hold at the same time, hence two logical signs proving them cannot be of equal strength. One of the signs is bound to be illegitimate, and an illegitimate sign cannot nullify a legitimate sign. Hence Dharmakīrti does not include it in his list. Consequently, he does not think that *asatpratipakṣitatva*, that is, the property of a sign not being contradicted by a countersign, is a necessary prerequisite for any logical sign's being legitimate, which is the Nyāya view. Nor do the Buddhists admit that *bādhita* or a contradictory sign should be listed as a separate pseudo-sign. The example of a contradictory pseudo-sign as found in the Nyāya literature is: Fire is nonhot because it is a product. The Buddhists point out that if a person is in a position to assert that fire is not hot, then he is not aware of the invariable concomitance between fire and hotness. That means, this type of defect originates from the violation of the second and/or the third character(s) of a logical sign and can be subsumed under some other pseudo-sign already mentioned. Hence they do not even include noncontradictoriness as necessary.

Later Buddhists, maybe due to the Nyāya influence, overcame Diṅnāga's reservations about contraposition as a logical operation. In Ratnakīrti we find a defense of the Buddhist doctrine of momentariness along with its contraposed version. Thus Ratnakīrti argues not only for the thesis that "all existents are momentary" but also for "all nonmomentary things are nonexistent." Acceptance of the contraposed version of the doctrine of momentariness leads to two important philosophical issues. Let me address them one by one.

A nonmomentary or a permanent object, according to the Buddhists, is as unreal as a sky lotus or a hare's horn. Now as the subject-locus is unreal, the sign becomes a pseudo-sign of the unestablished type. Hence, urges the Naiyāyika, the contraposed version remains unestablished even by the Buddhist standard. Not only that, Ratnakīrti here ascribes some unreal attribute to an unreal subject, which is anathema to the Naiyāyika. The Naiyāyika-s will not allow statements like "The hare's horn is sharp" or "The hare's horn is not sharp." Besides, it appears that the Buddhists are ready to treat "existence" and "nonexistence" as predicates, which is not a valid move within first-order logic. The Naiyāyika has problems with ascribing an unreal predicate to an unreal object because they are committed to a realist ontology. But some Buddhists do not fight shy of positing a Meinongian domain as their universe of discourse. In fact, they treat all conceivable things at par, real or fictitious. That is how Ratnakīrti could avoid the charge of using an inconclusive pseudo-sign.

There is also another apparently insurmountable difficulty with these two inferences involving "whatever is existent is momentary" and "whatever is nonpermanent is nonexistent." Most Indian logicians demand that in an acceptable inference, one should be able to cite a supporting example to which both the proponent and the opponent agree; the supporting example must not be identical with or included in the subject-locus. Now, if the subject locus includes everything existent and real, then it is not possible to cite a supporting example in favor of the thesis because a fictitious or an unreal example will not do.

Such a predicament can be avoided by introducing the notion of internal concomitance (*antarvyāpti*) as opposed to the notion of external concomitance (*bahirvyāpti*). The distinction between external and internal concomitance may be formulated as follows. In case of external concomitance, this concomitance is apprehended in a corroborative example, but in case of internal concomitance, it is apprehended in the subject-locus (Bhattacharya 1986). So if the concomitance between the sign and the signified is taken as internal, that is, as one apprehended in the subject-locus of the inference itself, then it is not necessary to cite a supporting example. That is why, Ratnākaraśānti, a Buddhist logician after Ratnakīrti, has given an elaborate defense of internal concomitance following the Jaina philosopher Siddhasena.

Buddhist logicians earlier than Ratnākaraśānti were much more conservative. That is why we find that all of them including Dharmakīrti, Jñānaśrīmitra, and Ratnakīrti stuck to Diṅnāga's principle of the triple-charactered sign and remained committed to external concomitance. The *trairupya* theory is in favor of external concomitance, "since in accordance with it, the concomitance, positive and negative, between the sign and the signified has to be apprehended in external examples, homogenous and heterogenous" (Bhattacharya 1986). Probably, this is why Ratnakīrti tried to solve the problem by subscribing to a Meinong-like ontology of the unreal conceivables, so that he could take a nonmomentary entity as a "disagreeing example" for the universal concomitance between existence and momentariness.

A section on PA in Buddhism cannot be ended without some discussion of their reductio argument (*prasaṇgānumana*). The Buddhists, like the Naiyāyika-s, are for excluding them from inference proper. For inference proper is a valid means of acquiring knowledge, but reductio is not. In a reductio argument, there is a counterfactual premise, which the agent hypothetically offers, knowing fully well that such is not the case. Dharmakīrti in *Pramāṇa-vārttika* and Mokṣākaragupta in *Tarkabhāṣā* recommend reductio argument mainly for ascertaining the relation of universal concomitance. But Nāgārjuna has used his reductio argument to solve some metaphilosophical problems. Take for example the argument,

> If knowability had been the cause of nameability, the statement, "A thing which is nameable is knowable" would have been true.
>
> But it is impossible that the statement is true (rather it is nonsense).
>
> Therefore, it is impossible that the assumption "knowability is the cause of nameability" be true.
>
> The logical form of Nāgārjuna's Reductio is as follows.

$$L\ (p \to q)$$
$$L \sim q \,/ \therefore L \sim p.$$

Nāgārjuna has used his reductio argument to reject all philosophical views. But he insists that he is not committed to any view. He uses four-cornered negation for this purpose. It is like this:

A. He negates the thesis p.

B. He negates the thesis $\sim p$.

C. He negates the thesis $(p \,\&\, \sim p)$.

D. He negates the thesis $(p \vee \sim p)$.[26]

Surely Nāgārjuna's negation is non–truth-functional. For, under truth-functional interpretation, Neg.$(p \,\&\, \sim p)$ and Neg.$(p \vee \sim p)$ can never have the same truth value. Besides, Nāgārjuna maintains that all four theses are meaningless and should therefore be rejected, but their negations are not true. He thus upholds that the negation of a philosophical thesis is no philosophical thesis.

This is, in sum, an oversimplified overview of Buddhist logic. In this exposition I have deliberately refrained from all attempts at formalization. But if such an attempt is to be made at all, the best way will be to try our hand within a model-theoretic framework.

4. Jaina Logic

Tushar Kanti Sarkar

4.1. Definition of *Anumāna* and Its Constituents (*Avayavas*)

The Jaina logicians define inferential knowledge (hereafter "inference" for short) as "getting at the knowledge about the probandum (*sādhya*) on the basis of one's knowledge about the probans (*sādhana*)." This very definition of inference is repeated verbatim by Hemacandra (PM) and also by Yaśovijaya (JTB 12). Yaśovijaya adds that inference is of two kinds, namely, *svārtha* (for personal conviction of the inferer) and *parārtha* (for the conviction of others, that is, the public in general).

Originally (as in *Tattvārthasūtra*), the Jainas were inclined to accept the three component view of inference. This view was supported by Sāmantabhadra (AM 6, 17), Pujyapāda (SS 10/5, 6), and Siddhasena (NAv 13, 14). Bhadrabāhu in his *Daśavaikālikaniryukti* (DVN: 49–137) speaks of two, three, four, five, and ten components of inferences. However, the standard view of the Jainas is two-component view of inference. Hemacandra (PM 2/1/9), Māṇikyanandī (PMS 37), and Anantavīryācārya (PRM) (in his commentary on the above *sūtra* 37) support the two component view.[27]

Most supporters of two-constituent theory make it amply clear that the restriction of the number of components to two is *not* a logical requirement. One may go on adding more and more components, if needed, taking into consideration such contextual factors as the intellectual level and background knowledge of the person for whose conviction the inference is being made. Stripped down to its barest minimum, an inferential unit takes the following form: The hill is fiery because of having smoke (which would not be there otherwise). This is an enthymeme of the second order. An enthymeme does not really have a lesser number of premises, only a lesser number of premises explicitly stated. It makes use of the information background of the inferer. Similarly, the Jaina two-constituent theory tries to emphasize the fact that if inference is considered as a deterministic input-output sequence, then depending on how rich the database of the inferer is, the amount of input information to be fed for getting a certain output may be minimized gradually. With this background in mind, it becomes quite understandable why the Buddhists go even a step further and regard the Reason (*hetu*) as the only necessary constituent.[28] Actually, the controversy regarding the number of components that is required for an inference to be possible is not a logical controversy in the sense in which the question of consistency of an axiom system is. It is actually a controversy about how much background information may be taken for granted in drawing an inferential result. Hemacandra states that a probans (which cannot be explained otherwise) plus the declaration of what is to be proved (*pratijñā*) are the only two constituents needed. Because *pratijñā* does not have any assertional force, it acts only as a signpost for the desired conclusion. The one operative (*prayojaka*) constituent of an inferential unit is therefore probans alone. So the Buddhist and the Jaina views do not really differ on this point. The real and truly logical point of controversy comes to relief only when we seriously analyze the implications of the views of Jaina thinkers like Anantavīryācārya,[29] who rejects the three-constituent theory (*pratijñā*, *hetu*, *udāharaṇa*) of the Sāṃkhya-s and the four-constituent theory (*pratijñā*, *hetu*, *udāharaṇa*, *upanaya*) of the Mīmāṃsakas. As the Jainas deny any role to *pakṣadharmatā* and *udāharaṇa* is not required, (we shall presently say why) this third constituent is dropped by them. On the other hand, *upanaya* is dropped from the list as well because it actually is a *metalogical* step rather than a logical one. *Upanaya* only tells that the substitution instance say, "a" of the universal variable "x" in the sentence "$(x)(Lxh \rightarrow Lxs) \,\&\, (Lah \,\&\, Las)$" is a legitimate one. This is like the rule of E.G., which entitles one to infer "$(Ex)Fx$" from "Fa," *provided* "a" *is nonempty.* The role that the *upanayavākya* plays is that of explicitly stating that the said proviso (which is a metalogical rule) has been fulfilled.

Now, if *udāharaṇa* and *upanaya* be dropped, in this way, we are left with only three constituents out of the five. These are *pratijñā*, *hetu*, and *nigamana*. Of these, *pratijñā* is the provisionally entertained, but assertionally noncommitted, expression of the conclusion. As such, *pratijñā* acquires the "right to assertion," so to say, only on the basis of the *hetu* and thus becomes the *nigamana* at the fifth step. Hence, given the *pratijñā* and the *hetu* the conclusion need not be stated separately for anyone except a very dull fellow. Finally, therefore, only two components are retained, namely, *pratijñā* and *hetu*. But why may we not go further and drop even *pratijñā* from the list of five? From the same probans we may infer a lot of different things. From the presence of smoke on the hill not only can we infer presence of fire there, we could also infer with the same legitimacy the presence of wet fuel there. What we actually would infer depends on what we aim at establishing (*abhīpsita*). Hence *pratijñā*, acting as the signpost for the desired inferential cognition, must also be regarded as a nondispensable constituent of an inferential unit and should be mentioned at the very beginning.[30]

4.2. Svārtha and Parārthānumāna

Indian logicians classify inference patterns or types by using different principles of classification. One such scheme of classification about which the Jainas, the Buddhists, and the Naiyāyikas agree, classifies inferences into *svārthānumāna* (i.e., inferences where the subjective feeling of conviction of the inferer himself is the adequate criterion of justification) and *parārthānumāna* (i.e., inferences where the justifications need to be convincing to others, i.e., the public).[31] Although the unchallenged practice is to translate *svārthānumāna* and *parārthānumāna* as "inference for oneself" and "inference for others" respectively,[32] I propose to use "privately justified inference" and "publicly justified inference" for *svārthānumāna* and *parārthānumāna* respectively. This deviant translation, as we shall see, would help avoiding a number of misunderstandings. The way *svārthānumāna* and *parārthānumāna* is characterized in some texts (e.g., TS and TSD) tends to suggest, misleadingly, that

i. *svārtha* and *parārtha* inferences are inferences in the same sense[33] except that *svārthānumāna* is a less elaborate, covert, and compact version of *parārthānumāna*; and

ii. the role of *svārthānumāna* is to convince the inferer himself, while the aim of *parārthānumāna* is to convince others.[34]

The real import of the distinction, however, is quite different. The role of *svārthānumāna* according to our view, consists in showing how an associative bond between *hetu* and the *sādhya*, once it is established, automatically leads one to arrive, under appropriate initial conditions, at the relevant conclusions. *Svārthānumāna* therefore should be viewed as providing the required psychological grounding for a corresponding *parārthānumāna*, that is, an inferential argument fit to be presented in a public discursive context. A closer look at the characterizations of the two as given in various Jaina texts[35] clearly indicates that the Jaina logicians did understand the distinction between the two types of inference basically in the way suggested by us.

Viewing the natures of and the distinction between the two inference patterns in this way also helps one to tackle the following two vexing questions, namely:

i. how to understand and answer those who claim that, strictly speaking, justification of no inference can be public (i.e., there is no *parārthānumāna*)?[36]

and

ii. how to reconcile the standard view that *svārthānumāna* requires a lesser number of steps with the views of those Buddhist logicians or of those Jaina logicians (who hold that an inference needs to have just one component) but at the same time also accept the classification of inferences into *svārtha* and *parārtha*?

On the interpretation proposed here, the second of the two questions gets answered immediately because *svārthānumāna* is not an inference at par with *parārthānumāna* at all and hence cannot be regarded just as a condensed form of the latter. On the other hand, if *svārthānumāna* be viewed as providing the required psychological basis of *parārthānumāna*, as I indicated, then there is a sense in which the notion of an *exclusively public justification* of any inference (i.e., *parārthānumāna*) becomes a nonstarter—all empirical justificatory inferential grounds being, on ultimate analysis, based on the establishment of subjective psychological bonds of association.[37]

4.3. Nature of Vyāpti

Inference is the process of justified passage from an inferential ground or probans to its consequent, that is, the probandum. This entails that there has to be an unfailing concomitance between the probans and the probandum. Such an unfailing concomitance relation is called *vyāpti*. By what marks can tell an arbitrary relation from a concomitance relation (*vyāpti*) between a probans and a probandum? The Naiyāyikas, hold that when a probans has all the *five* characteristics (viz., presence in the *pakṣa* [locus of a probans], presence in *sapakṣa* [homologous] cases, absence in *vipakṣa* [heterologous] cases, not being countermanded by a stronger epistemic evidence [*abādhitatva*] and not being counterbalanced by opposing evidences [*asatpratipakṣatva*]) that indicates that such a probans has unfailing concomitance relation with its probandum. The Buddhist thinker Diṅnāga insisted that a probans must fulfill the three of these requirements to ensure that *vyāpti* relation holds between a probans and its corresponding probandum.[38] Vādibhasiṃha in *Syadvādasiddhi* shows that even when a probans is characterized by all the five required characteristics, it may still fail to ensure a *vyāpti* relation with its probandum. As a counterexample, he gives the following argument: "The as yet unborn offspring of Maitri will be of dark complexion because of being an offspring Maitri (like all of Maitri's other offsprings, who are dark complexioned)." Clearly, the probans of this inference (viz., being an offspring of Maitri) fulfills all the five requirements (and thus also the three conditions laid down by the Buddhists), but it still fails to ensure any universal concomitance relation (*vyāpti*) between itself and the probandum (viz., having dark complexion).[39] Hence, the Naiyāyikas' five-characteristic as well as the Buddhist's three-characteristic criterion fail as indicators of *vyāpti* relation.

The Jaina logician Pātrasvāmī holds that a probans to ensure unfailing concomitance with probandum need and must fulfill just one single condition, which is the sine qua non of unfailing concomitance. This single requirement is called *anyathānupapannatva* (probans being otherwise unexplainable without positing the relevant probandum as its explanans). It is also called *avinābhāva* (constant conjunction between the two). *Avinābhāva* may also be taken to mean (logically) inseparable conjunction, as opposed to a merely factual constant conjunction. The Jaina logicians were not very careful or consistent in their use of the term *avinābhāva*. The Jaina thinkers also reject repeated observation (*bhūyodarśana*) as capable of ensuring knowledge of unfailing concomitance, they ultimately have to fall back on the use of *avinābhāva* as a conceptually inseparable relation.[40] Although the Jaina logicians used *avinābhāva* and *anyathānupapanna* interchangeably, I think, it will be helpful to keep the two senses of constant conjunction (viz., merely empirical, that is, factual inseparability and conceptual inseparability) apart. We shall use *avinābhāva* to mean empirically ascertained constant conjunction and use *anyathānupapanna* to mean conceptual inseparability or inseparability in principle.

Thus in the Jaina texts like *Pramāṇa-parīkṣā*,[41] *Nyāyāvatāra*,[42] *Pramāṇa-saṃgraha*,[43] and *Saṭkhaṇḍāgama*,[44] single-feature criterion of probans (which is capable of ensuring appropriate kind of concomitant relation) has been given in terms of inseparability.

4.4. Ascertainment of Vyāpti

In their answers to the question "how is *vyāpti* relation ascertained?" different schools of Indian logicians differ widely from one another.

The Jaina logicians point out that elimination of *all possible* accidental factors from a postulated probans is possible neither by perception nor by some other inference.

Akalaṇka sums up in a verse the Jaina position about how *vyāpti* is known (*Nyāyaviniścaya*, 2/326). The verse says: Given observation of uncontradicted constant conjunction between a probans and a probandum, it would still have to be supported by *tarka* to yield the knowledge of an invariable universal concomitance.[45]

4.5. Nature of Tarka

The Jainas were also very clear and emphatic about one point, namely, the relation of invariable concomitance (*avinābhāva*) and conceptual inseparability (*anyathānupapannatva*) cannot be ascertained either by direct perception or by another inference. So they proposed that hypothetical reasoning (*tarka*) be taken as an independent *pramāṇa* (means of epistemic justification leading to knowledge). *Tarka* alone is capable of ascertaining invariable concomitance in the required sense. Not only does *tarka* help ascertaining *avinābhāva* (*tarkāt tanniścaya*),[46] but such ascertainment is the specific aim of *tarka*. Even a Naiyāyika like Udayana (in NVTP 1.1.5) had to admit that where invariable concomitance is ascertained merely on the basis of observation of instances of co-presence, it runs the risk of being vitiated by the presence of an accidental factor (*upādhi*) called *pakṣetaratva*.[47]

The idea of *tarka* as a mode of hypothetical reasoning starting with some tentatively entertained posits (posit-based hypothetical reasoning or PBHR for short) failed to be appreciated by other Indian logicians. The Jainas seem to have had a vague inkling of the methodological significance of *tarka* as a PBHR.[48] No direct textual evidence can be given for this claim. However, some circumstantial evidence can be given to show its plausibility.

i. The Jainas were unequivocal about the indispensability of universal law-like generalizations for an inference and also about the inadequacy of an inductivist approach to law-like generalizations.[49]

ii. They were sure that reliable law-like generalizations are possible.

iii. So they held that there is some noninductivist solution to the problem.

iv. *Bhūyodarśana* or observation of all cases was rejected by them and they also rejected perception and ordinary inference as means of ascertaining the relation.

v. Not mere constant conjunction but *anyathānupapannatva* (the probans's being unexplainable without the relevant probandum being posited) was substituted for "constant conjunction" as an essential mark of the relation.

vi. Some sort of analytical/conceptual link (*antarvyāpti*) between probans and probandum was considered to be the only real guarantee to have for the knowledge of their law-like regularity.

vii. Corroborating instances (*udāharaṇa*) and extrinsic concomitance (*vahirvyāpti*) were considered inessential for establishing a law-like generalization.

viii. *Tarka* was given the status of an independent full-fledged *pramāṇa* because, among other things, of its *avisaṃvāditva* (pragmatic coherence).

All these clearly suggest that *tarka* for the Jainas is a methodological tool crudely similar to what is nowadays called hypothetico-deductive method of theory construction. Moreover, if this PBHR view of *tarka* is correct, then one should also expect the notions of *tarka* and *antarvyāpti* to be closely related ones and that actually happens to be corroborated by textual evidence from Jaina sources.[50]

4.6. Antarvyāpti and Bahirvyāpti

The Nyāya school maintains that *vyāpti* can be ascertained by repeated observation of constant conjunction coupled with nonobservation of any exceptions to it. This implies that for the ascertainment of *vyāpti*, observation of corroborating instances is indispensable and hence all cases of *vyāpti* are cases of instance-based empirical generalization. This uniform model of *vyāpti* is accepted without qualms by most schools of Indian logic, except by the Jaina and the Buddhist schools of logic. The Jaina logicians classify *vyāpti* into *antarvyāpti* (intrinsic concomitance) and *bahirvyāpti* (extrinsic concomitance). Vādidevasūri says: If a given minor (*pakṣa*) is such that in it (i.e., within itself) concomitance between probans and probandum holds then that is a case of *antarvyāpti*: Elsewhere it is *bahirvyāpti*.[51]

According to Phanibhusana,[52] *antarvyāpti*, that is, "concomitance of a probans and its probandum holding internally in a given minor" should be taken to refer only to those cases where for the purpose of inferring "fire-on-the-hill" from "smoke-on-the-hill," one concentrates on the cognitive awareness of concomitance between "this hill-smoke" and "this hill-fire" to the exclusion of similar other cases of smoke-fire relation. This hardly makes any sense. Moreover the Jainas claim[53] that if *antarvyāpti* is capable of justifying the inferential transition from probans to probandum, then *bahirvyāpti* (extrinsic concomitance) is redundant; again, if *antarvyāpti* is not capable of it, then taking recourse to *bahirvyāpti* becomes useless. Pandit S. C. Nyāyacārya[54] maintains that by *antarvyāpti* the Jainas simply meant the type of *vyāpti* used in inferences that yield pan-inclusive universal conclusions (*kevalānvayī-anumāna*). This interpretation entails that *antarvyāpti* is a limiting case of *vahirvyāpti* having sole application to what the Naiyāyikas call *kevalānvayī-anumāna*. However, a more plausible interpretation of *antarvyāpti* would be to take it as a concomitance relation that is grounded on some internal conceptual link (roughly, a logico-semantic relation) between a probandum and its probans (which happen to be co-located/correlated in the minor, i.e., *pakṣa*). On this interpretation the MTNL claim about the adequacy and indispensability of *antarvyāpti* makes good natural sense. It is this: In the absence of any internal conceptual link between the probans and the probandum, extrinsic concomitance (universal generalization based solely on points of analogy in similar other instances) remains an unreliable ground of inference; while in the presence of such an internal conceptual link (with a priori ascertainability) an empirical generalization and its factual corroboration (i.e., *bahirvyāpti*) becomes redundant as the logical ground of a valid inference.[55] Matilal (1998, 12) seems inclined to accept such an interpretation.

The Buddhists hold that every *vyāpti* is either a case of identity relation (*tādātmya*) or else a case of cause-effect relation (*tadutpatti*).[56] Of these, the identity-based cases would ensure that there are conceptual links ascertainable a priori. On the other hand, the cause-effect based ones would account for pure empirical generalization. Accordingly, both the Jainas and the Buddhists had to rely on *antarvyāpti* to justify their respective pet metaphysical theses like "whatever is real is momentary" (Buddha) or "whatever is real has an infinite number of modes" (Jaina). The use of such metaphysically loaded claims has illustrative examples was unfortunate, as because of it, the ultimate shape that the controversy about *antarvyāpti* took could fool both its opponents as well as its proponents into believing that the only point for postulating *antarvyāpti* was to defend some pan-inclusive metaphysical claims or else clinch such fairly uninteresting issues like two-component theory of inference. Consequently the true logical significance of *antarvyāpti* was missed by all.

4.7. Hetvābhāsa

The logicians of the Nyāya school, as we have seen, held that a legitimate probans must be characterized by five features. A probans, in its turn, is considered illegitimate if it lacks any one or more of the five required features. Accordingly, the standard view of the Naiyāyikas is that there are five types of *hetvābhāsa* or defects of inference. Since, according to the Buddhists, any defect of inference is due to the fact that the probans used lacks one or other of the three legitimizing features mentioned by them, they admit of only the three kinds of *hetvābhāsa*, namely, *savyabhicara*, *asiddha*, and *viruddha*.[57]

The Jainas, as we have seen, hold that neither the five nor the three features can guarantee the legitimacy of a probans and ensure its ability to logically lead to probandum as a conclusion.[58] According to them, "not otherwise accountable" (i.e., *anyathānupapannatva*) is the sole requirement (both necessary and sufficient) for a probans being legitimate. So there is only one type of defect of inference, which arises from the violation of the requirement.[59]

Instead of calling all defects of inference *hetvābhāsa* the Jainas used a more inclusive term—*anumānābhāsa*—to mean defects of inference in general.

Now, though there should be only one type of *anumānābhāsa*, how is it that many Jaina writers speak of three or four types of them?[60] This apparent inconsistency in the Jaina position is explained by the Jaina thinkers following a general strategy, namely, by maintaining that all the so-called varieties of *hetvābhāsa* are but different ways of failing to satisfy the *anyathānupapannatva* requirement.[61] Akalaṇka goes even a step further and holds that all sorts of violation of the *anyathānupapannatva* requirement have a common feature, *akiñcitkaratva* (lack of inferential significance).[62] Vādidevasūri (in PNT 6/57) regards *akiñcitkara* as a defect that characterizes a pointless inference, (e.g., trying to inferentially justify a tautology or something that is obviously true).[63] It is easy to see why Yaśovijaya refuses to accept *akiñcitkara* as a distinct type of logical defect on top of the three mentioned by him.[64]

It is interesting to note that the nonorthodox Naiyāyika Jayanta Bhaṭṭa admitted *akiñcitkara* as the sixth type of *hetvābhāsa* (on top of the usual five admitted in *Nyāya* logic) but subsequently succumbed to the pressure of tradition and proposed to include it under *asiddha*.

4.8. Treatment of Existence in Jaina Logic

If logic, as the theory of inference, is always tied down to the actual world, then how is it possible, even for the sake of an argument, to make inferences about things, which are not known to exist? Yet any minimally powerful logic must be able to tackle this problem. The Jainas were aware of this problem and tried to find out a solution for it. First, they maintained that when it comes to existence proofs of ordinary things, it can be ascertained by one standard means of knowing (*pramāṇa*) or another, for example, sense perception. Second, when there is no such pramāṇa we need to use *vikalpa* proof. In *vikalpa it is legitimate to* hypothetically entertain as a posit the entity whose existence is sought to be proved. *Vikalpa* proof depends on the principle of self-consistency, which consists in the absence of a definite impossibility proof[65] (Cp. the proof that the square roots of -1 exist, or that there is no cardinality between Aleph-null and Aleph-one). Third, according to the Jainas, existence of something may be denied by changing the negative existential statement to its proper logical form. For example, "there are no hare-horns" is to be transformed into "there are no hares with horns." Fourth, the Jainas maintain that by means of *vikalpa*, only the existence or nonexistence of a thing with certain properties can be established. Now, by the second condition, *vikalpa* is applicable in proofs of only such entities which are not provable by any standard *pramāṇa*, such as perception (*pratyakṣa*), ordinary inference (*anumāna*), and so on. This coupled with the fourth condition gives the first rule of existence proof in Jaina logic.

This can be expressed in the form of a basic rule R_0.

Rule 0 A hypothesis H will come under the scope of *vikalpa* proof provided

a. H is an existence postulate of the form "$\Phi\, a$" where "a" denotes an individual (*dharmin*) and "Φ" is a putative *dharma*.

b. The truth-claim of H cannot be ascertained by any of the standard *pramāṇas*, for example, ordinary perception or inference or the like.

It is a rule about the range of applicability of *vikalpa*. Hence, R_0 is a meta-rule.

Rule 1 A *dharmi* with any property ϕ *possibly exists* if the assertion that "$\phi\, a$," that is, "a has ϕ" does not logically (necessarily) entail any contradiction. (Since we allow a to be possibly an *empty* name the rule of E.G. fails to hold, that is, here "$\phi\, a$" does not entail "$(\exists x)\,\phi\, x$.")

In symbols,

$$(\mathrm{R}_1) \qquad {\sim}L(\phi\, a \rightarrow C) \leftrightarrow M(E \mid \phi\, a).$$

R_1 gives a more tangible shape to the abstract principle of self-consistency as a *minimal* requirement for any existence claim being entertainable.

(Here, "L" and "M" are standard modal operators, "$\exists \mid \phi\, a$" means there is a "ϕ-characterized a.")

The next rule, Rule 2, adds flesh and blood to the skeletal claim authorized by R_1. Clearly from the fact that the hypothesis "perfect human knower exists" is

i. not ascertainable by any standard *pramāṇa*, and

ii. does not logically entail any self-contradiction,

it follows by virtue of R_0 and R_1 that $M(E \mid \phi\, a)$. However, from "$M(E \mid \phi\, a)$" we *can* logically infer neither "$(\exists y)(E \mid \phi\, y)$" nor "$(E \mid \phi\, a)$" nor even "$(\exists y)M(E \mid \phi\, y)$." Yet to be able to establish that *sarvajna asti* we must be able to derive at *least* "$(\exists y)M(E \mid \phi\, y)$" from $M(E \mid \phi\, a)$.

The rule that authorizes this passage/derivation is Rule 2. Before giving a precise formulation of R_2 we need to define a technical notion, namely, that of a *desirable property*.

Definition D_1 Ψ is a desirable property iff

i. Ψ is an instantiated property,

ii. Ψ is an original (*sahaja*) property (*dharma*),

iii. The grades or degrees of Ψ can be linearly ordered.

Let P be the set of all desirable properties and ϕ be the property of being a human knower.

Now, we can give a precise formulation of Rule 2.

Rule 2 If any *dharmi* a possess a *dharma* ψ such that (i) $\psi \in P$ and (ii) $M(E \mid \phi\, a)$ holds then ϕ has a maximal upper bound (m.u.b.), say, ϕ^* and there is some y such that

$$(\exists y) M(E \mid \phi^*\, y) \text{ holds.}$$

Now, since ϕ (being a human knower) satisfies all the *three* conditions (i)–(iii) of D_1, R_2 can be applied to "$\phi\, a$" to get the result $(\exists y) M(E \mid \phi\, y) \ldots (A)$.

Clearly, (A) is a logically stronger claim than both $M(E \mid \phi\, a)$ and $M(\exists y)(E \mid \phi\, y)$.

The clearest standard example of this kind of existence-proof is the Jaina proof for the existence of a most perfect human knower or omniscient being, called *kevalin.*[66] In short, the argument used by the Jainas to prove the existence of omniscient persons (*kevalin*) runs thus:

i. The idea of a perfect human knower (*kevalin*) cannot be established by any standard *pramāṇa* such as perception, ordinary inference, and so on.[67] So, it is a case fit for coming under the scope of *vikalpa* method of proof ... by R_0.

ii. Let "ϕ^*" stand for the property (*dharma*) of being a perfect human knower. The idea of a perfect knower is not provably inconsistent. Hence, such a knower possibly exists ... by R_1.

iii. Knowledge is an actually instantiated property, say ϕ, which admits of gradations in terms of its degree of perfection and these various degrees are linearly orderable.[68] Hence, $\phi \in P$, and thus ϕ has to have an m.u.b. ϕ^*, such that there exists some knower y who has ϕ^* and who is type-homogeneous to imperfect knowers like us, that is, who $(\exists y) M(E \mid \phi^*\, y)$ ... by R_2.[69]

But can we validly infer that there is a perfect smelling rose or a perfectly hot thing by rules R_0, R_1, and R_2? The Jainas would say that we cannot. But why not? How to block such undesirable existence proofs? The only way it can be done is by imposing some restrictions on the nature of P in R_2, which would ensure that only on properties of a very special kind the rules R_0, R_1, and R_2 can be applied and the resulting existence proof can be carried through. The Jainas called such special properties *sahaja dharma* (original properties).

Such restriction would exclude “perfect smell” or “a perfectly hot object,” etc., from the range of applicability of R_0, R_1, and R_2. This is precisely what the Jainas did.[70]

We may now look at the Nyāya-Jaina controversy about talks about nonexistents. The controversy centers around three main points:

i. The Jainas admit, while the Naiyāyikas do not, that indirect proof (*vikalpa*) is a legitimate way of proving existence in certain specific types of cases.

ii. The Jainas hold that prior to establishing the existence or otherwise of a purported entity we can significantly and meaningfully talk about it in a logically legitimate way (JTB, 14, line 10). The Naiyāyikas hold that all such provisional talks are meaningless and logically wrong.

iii. The Naiyāyikas follow the Paninian grammatical model of sentence analysis which leads them to treat sentences (about existence/nonexistence of something) as composite wholes, for example, “Mermaids-as-possessors-of-nonexistence” (= Mermaids are nonexistent = There are no mermaids.) and then, “Mermaids are charming” and also “Mermaids are not charming” become truth valueless (cp Strawson).

Thus, if lack of a truth value is the mark of nonsensicality of a sentence (as some logical positivists held), then in the light of that, the Naiyāyika claim (namely, prior to or in the absence of having a criterion of ascertaining the emptiness (or otherwise) of the subject of a discourse, all talks about such a subject are to be treated as meaningless (nonsensical)), becomes quite intelligible. The Jainas strongly reject this Nyāya view.

Russell’s TDD tackled the problem of empty description by using the five logical strategies, one of which was to logically recast the troublesome sentences in a prescribed format. Jaina logicians too prescribed such a strategy of recasting. For example, instead of the whole compound concept of “donkey’s-horn-as-a-locus-of-existence” being negativized, the Jainas recommended that the compound expression be componentially analyzed and be recast into “there is no horn which belongs to a donkey.”[71] This clearly avoids any commitment to any such thing as “existenceless-donkey-horn” (*khara-viṣāṇābhāva*) as a quasi-real entity.

The Jaina theory of *vikalpa* as a sort of existence proof, it should be noted, does not prove the actual existence of anything. It is more like a proof to the effect that there exists a limit to which a given infinite geometrical series (say, $1 + 1/2 + 1/4 + \cdots$) converges (without that limit being ever actually reached). The epsilon-delta definition of the limit of a function in terms of a “neighborhood” would be an even better analogy to express the Jaina position.

Quine had a very strong ontological view backed up by his logic. His is a classic case of logic dictating ontology. It is betrayed by Quine’s view on deviant logics. The Naiyāyikas, on the other hand, started by subscribing to a very rigid ontology. Moreover, their view about language-reality tie-up plus their view that everything knowable (*jñeya*) is necessarily linguistically expressible (*abhidheya*) committed them to draw a sharp and fixed line of demarcation between referring and nonreferring terms. So, theirs is a case of ontology dictating logic. Although, in a sense, Jaina logic was also shaped by their ontology, yet unlike Nyāya logic, Jaina logic was nonrigid and flexible because of their highly accommodative ontology of *anekāntavāda*, supported by an equally flexible logico-epistemic cum linguistic theory. Furthermore, the Jaina *nayavāda* (epistemic perspectivalism) ensured that all knowledge-claims are perspective-relative (and hence, incomplete or partial). Their *syadvāda* (logic of propositional expressions) ensured that all propositionally made truth claims are conditional. Jaina theory of language made its contribution by developing, among other things, a scheme of classification of propositional expressions (i.e., any meaningful sentence which either expresses a proposition or seeks a propositional answer) into truth-functional (*paryāpta*) and non–truth-functional (*aparyāpta*).[72] Truth-functional expressions are classified as "true" or "false." "Non–truth-functional" expressions are classified into quasi–truth-functional (*satyamṛṣā*) and pure non–truth-functional (*asatyamṛṣā*).[73]

The rule R_1, as we know, says that if a claim does not entail logical inconsistency, then it is provisionally entertainable as really being true. The question now is how to define inconsistency, particularly in the context of Jaina logic. For the Jainas even the notion of inconsistency is conditional, that is, context-relative. In short, there *is no absolute criterion* of logical inconsistency. If the notion of inconsistency itself be relative and conditional then how is it that the Jainas regard some claims, such as, "He is the son of a childless woman," as unconditionally inconsistent? Second, if the notion of inconsistency be always relative (and thus have multiple criteria) then which of those possible criteria is to be chosen for use in R_1 in the context of *vikalpa* proof? In the context of defining *fuzzy* or *graded* consequence relations, the choice of an appropriate notion of inconsistency becomes crucial.

4.9. Contradiction in Jaina Logic

Before one can answer these questions, some background information is required.

i. The Jainas, like all other Indian logicians, define contradiction/inconsistency in terms of *epistemic incompatibility* and not in terms of propositional incompatibility, as is done in Western logic.

ii. Contradiction is said to occur only when one epistemic state (*jñāna*) is blocked (*pratibaddha*) by another (*pratibandhaka*).

iii. Any given negated proposition, say not-q, does not form a contradictory pair with any arbitrary proposition p. A contradiction occurs only when both the assertion and the negation of the same proposition are conjoined. In the same way, an epistemic state J_1, can form an incompatible pair with another epistemic state J_2 only if J_1 and J_2 have exactly the same epistemic content (*viṣayatā*).

iv. *Viṣayatā* is a multicomponent composite notion, consisting of a specific set of relata and a specific relation between them. The *viṣayatā* of the knowledge "a pot is on the floor," consists of

a. the characterizer (*prakāra*) (here, it is the pot),

b. the characterized (*viśeṣya*) (here, it is the floor),

c. the specific floor-pot contact-relation of "being-on" (*saṃyoga sambandha*).

v. J_1 and J_2 cannot have the same epistemic content unless their respective *viṣayatās* have *exactly the same* componential cum relational specificity.

We may now proceed to analyze the notion of inconsistency in Jaina logic, keeping the above points in mind.

$$\text{Let } J_1 = \text{a pot is on the floor, and}$$
$$J_2 = \text{a pot is not on the floor.}$$

Are J_1, J_2 incompatible? It need not be, because one with knowledge J_2, may be referring to the pot (a different one) which he saw here yesterday, whereas J_1 has as its *prakāra* the pot which is now on the floor. Even a slight difference in any component of a *viṣayatā* can change its identity. Clearly, J_1, J_2 need not have the same *viṣayatā*. So they are not absolute incompatibles.

According to the Jainas, epistemic content of any *jñāna* has four dimensions of freedom (unspecificity), namely, *dravya* (substantiality), *kṣetra* (location), *kāla* (temporality), and *bhāva* (features). Each of which again admits of an infinite number of variations (JTB, 19).

It should be clear by now that any two epistemic claims like J_1, and J_2 can be shown to be nonincompatible (i.e., jointly entertainable) so long as even one dimension of freedom or a single degree of variations of their respective *viṣayatā* remain unfixed/unspecified. Exact specification of all dimensions and degrees of a specific *viṣayatā* is not possible (for obvious reasons) in normal cases.[74]

Now, if no epistemic claim is ever wholly untrue (as the Jainas seem to claim) then shouldn't we say that the claim "he is the son of a childless woman" is also partly true? The Jainas disagree. How to explain this? In Guṇaratna's commentary on Haribhadra's SDS we find an argument for the position that every real thing (*vastu*) has infinite number of characteristics (often apparently incompatible ones).[75] In that context it is said that whatever does not have an infinite number of characteristics is not a *vastu* (real thing). Let us use "$E \mid d, k, t, b, a$" to stand for "a exists relative to some d (*dravya*), k (*kṣetra*), t (time or *kāla*), and b (*bhāva*)," where d, k, t, and b are variables and a is a constant.

Using quantifiers on d, k, t, and b, we get the following proposition:

$$(\exists d)(\exists k)(\exists t)(\exists b)(E \mid dktba),$$

which expresses an existential claim about a. This claim can be contradicted only by $(d)(k)(t)(b){\sim}(E \mid dktba)$, that is, only when there is not a single substance, location, time or feature with respect to which it is true to say that "a exists." Clearly, if a is a four-sided triangle or the son of a childless woman then there is not a single time or location or feature in respect of which it is true to say that "a exists." Therefore, a claim to the effect that "the son of a childless woman exists," is unconditionally or absolutely false. Hence, it lacks the property of having infinite number of characteristics, but then by the Jaina criterion of reality it is an *avastu* (unreal) and hence, not subject to *anekāntavāda*, as expected.

The foregoing account gives a traditional interpretation and defense of the Jaina *anekāntavāda* and its underlying notion of mutual inconsistency. There are, however, other not so traditional interpretations of Jaina *anekāntavāda* proposed by some mathematicians (P. C. Mahalanobis)[76] and physicists (D. S. Kothari, Partha Ghosh, M. D. Srinivas),[77] and so on. The physicists prefer what I call a strong realistic interpretation of *anekāntavāda*, which holds that real, objective, contradictory properties can be and are simultaneously present in the reality without inconsistency, in a way similar to how an electron has the properties of a particle as well as of a wave[78] or in the way seven colors of the spectrum are simultaneously present in sun ray (cp. SDS). Some Jaina texts[79] also lend support to this strong realistic interpretation. These nontraditional interpretations are not to be discussed here. Elsewhere (Sarkar 1998), I have discussed the possibility of incorporating a strong realistic interpretation of *anekāntavāda*, which is based on an inconsistency tolerant paraconsistent logic. The possibility of using fuzzy logic and default logic has also been explored by some (Sarkar 1992; Brahma 1999) in an attempt to throw new lights on *Syadvāda* and some other aspects of Jaina logic.

5. An Introduction to Navya-Nyāya Logic

Sibajiban Bhattacharyya

5.1. Introduction

Navya-Nyāya logic is mainly a logic of cognitions. It is necessary to explain how "cognition" is used here.

1. "Cognition" is used in a very wide sense, to mean not merely propositional cognitions, like perceiving, inferring, remembering, imagining, assuming, introspecting, and so on, but also nonpropositional cognitions like sensing. But in developing the logical aspect of this system, we shall not have occasion to refer to the nonpropositional type of cognition.

2. "Cognition" is used only in the episodic sense, never in the dispositional sense. The propositional cognitions are called *qualificative cognitions*, the objects of which are always relational complexes of the form $a\ R\ b$. We now introduce three epistemic definitions.

Definition 1 Qualificand of a qualificative cognition: a in the cognized structure $a\ R\ b$, is called *the qualificand of the cognition* as well as *the qualicand of* b.

Definition 2 Qualifier (mode) of a qualificative cognition b in the cognized structure $a\ R\ b$ is called *the qualifier* (*mode*) *of the cognition* as well as *the qualifier* (*mode*) *of* a.

Definition 3 Qualification of a qualificative cognition R in the cognized structure $a\ R\ b$ is called *the qualification of the cognition.*

Here it is necessary to explain four points:

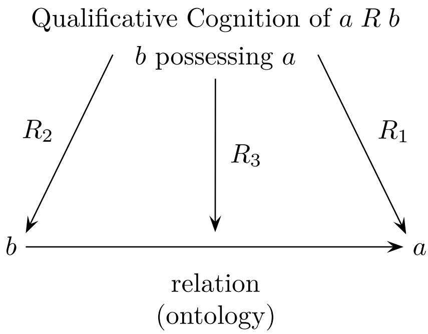
i. Every qualificative cognition of the structure $a\ R\ b$ is interpreted in Navya-Nyāya theory as cognition b in a by R. Thus "$a\ R\ b$" is a term, not a proposition in the usual Western sense. For in Navya-Nyāya theory, every qualificative cognition has to be analyzed as cognition of something in something, of something *possessing* something, of something as something, that is, something cognized under the mode of something. The structure $a\ R\ b$ is regarded as a complex object, and not as a proposition or fact in the sense of the *Tractatus.*

ii. The second point to be noted here is that the expression "$a\ R\ b$" is to be given a *de re* interpretation. Thus a, b, and R are objects in the real world.

iii. When the fact is that I perceive a table, that is, something under the mode of the property—being a table—what I say is just "a table" not that "I perceive a table." I do not perceive my perception. So also, generally, when I cognize an object, the fact is that I cognize it, but what I *say* is *what* I cognize, an expression for the object, *not* "I cognized the object." According to Navya-Nyāya, cognition is just belief, a true cognition is a true belief. We may note here that the Sanskrit term for "know" is *jña*, which is philologically cognate with the Greek "gnosis." Yet in spite of this philological affinity, semantically the Sanskrit noun *Jñāna* is very different from the English noun "knowledge."

iv. I have used "$a\ R\ b$" in symbols to mean the relational structure of a and b related by a dyadic relation R. But "$a\ R\ b$" is not its correct form, according to Navya-Nyāya philosophers. For (a) there are no relations admitted in the ontology, except a particular relation, inherence; (b) second, no relation can be expressed by a relation-word, for whatever is meant by a word becomes a term, which cannot be a relation. A relation is to be understood from the order of words occurring in an expression. To talk about a relational whole, it is usual in Navya-Nyāya logic to use the expression “possessing” to stand for relation in general. So the expression “b possessing a” is used, instead of “$a\ R\ b$.” Because relation is not admitted as a separate category in the ontology, it is held that any object of any category may function in a qualificative cognition.

What this function is, may be explained by a diagram:



a, as we have already pointed out (Def. 1) is the qualificand, b the qualifier (Def. 2), and “possessing” stands for the qualification (Def. 3). Any object to which the cognition is related by R_1 is the qualificand of the cognition; any object to which the cognition is related by R_2 is the qualifier; and any object to which the cognition is related by R_3 function in that cognition as the relation between a and b, that is, as the qualification in the qualificative cognition.

As R_1, R_2, R_3 are relations by which the qualificative cognition is related to a and b and R, the converse relations, $\breve{R}_1$, $\breve{R}_2$, $\breve{R}_3$, are the relations by which the three elements in the objective structure are related to the cognition. There is no difficulty in explaining how three elements in the objective structure can be related to the same cognition by three relations, $\breve{R}_1$, $\breve{R}_2$, $\breve{R}_3$. These relations, however, are cognized only in a second-order, introspective cognition, that is, although in every qualificative cognition there are always a qualificand, a qualifier and a qualification still they are cognized only in introspection.

I shall now state two objective definitions.

Definition 4 Property (superstratum) b is a property (superstratum) of a in the relation R in $a\ R\ b$ (b-possessing a).

Definition 5 Property-possessor (substratum, locus) a is the property-possessor (substratum, locus) of b in the relation R in $a\ R\ b$ (b-possessing a).

Now we explain how Navya-Nyāya language could apply to all serious discourse in the humanities. Navya-Nyāya language developed so that it could describe structures cognized on various occasions. Thus this language could be used in every sphere where cognition, belief, doubt, and other epistemic and doxastic factors play an essential role. This explains why this language could be used universally in the humanities, where epistemic factors predominate. In the sciences, however, propositions are stated in total or near total independence of the knowing subject. Hence Navya-Nyāya language, which is very helpful in the humanities, does not seem to be applicable in the sciences. Although a mastery of the Navya-Nyāya language is considered essential for studying medicine (*āyurveda*), still no work of medicine seems to have used this language. Modern Western logic, on the other hand, developed historically, so that it could be applied to mathematics, and is thus inappropriate for use in the humanities.

5.2. Some Technical Terms of Navya-Nyāya

5.2.1. Abstraction

Usually in Sanskrit, as in English, abstract terms are formed from given concrete terms, and concrete terms from given abstract terms. Thus, from the concrete terms, "wise," "honest," and "man" we have the following abstract terms by adding appropriate suffixes, making appropriate grammatical changes in the stem where necessary: "wis-dom," "honest-y," "man-ness," so also with abstract terms. But Navya-Nyāya philosophers form abstract and concrete terms from any terms. Thus, there will be higher-order abstract terms formed from given abstract terms; and, similarly higher order concrete terms from given concrete terms. The abstraction suffix is *tva* or its grammatical variant *tā* (translated as "ness") and the concretization suffix is *vat* or *mat* (translated as "possessing"). Abstraction and concretization are inverse processes in the sense that

(A1)
$$(t\text{-ness}) - \text{possessing} = t$$
$$(t\text{-possessing}) - \text{ness} = t,$$

where "t" is any term whatsoever, corresponding to the law of symbolic logic.

(A2)
$$[(\lambda x)Fx] = F$$

To explain a higher order abstract term, we may use in symbols:

(A3)
$$(\lambda y)[y = (\lambda x)(Fx)],$$

and so on.

Definition 6 Resident in: In $a\ R\ b$, a is the locus of b (Def. 5), and has, therefore, the property of being the locus of b. This "having a property" is stated in the form "the property of being the locus *resident* in a." So also b, which is the superstratum of a, has the property of being the superstratum of a. Hence the property of being the superstratum of a is *resident* in b.

Definition 7 Determiner-Determined: To say a is the locus of b, is to say that the property of being the locus (locus-ness) resident in a *determines* or *is determined by* the property of being the superstratum (superstratum-ness) resident in b. The determiner-determined relation is between *properties* of two correlatives. In symbols:

$$(\lambda x)\ x\ R\ y \text{ determines or is determined by } (\lambda y)\ x\ R\ y.$$

The structure $a\ R\ b$ (b-possessing a), which is interpreted as b in a by R is described in theoretical terms of the Navya-Nyāya system thus:

(S1) R is the limiting relation of superstratum-ness resident in b, which superstratum-ness determines the locus-ness resident in a.

5.2.2. The Concept of Limiter

Navya-Nyāya uses the "limiter" in many senses, extensively in its technical language. I explain here only two senses of the term: (a) Navya-Nyāya philosophers use the term to state explicitly the mode of cognition of any object; (b) they also use it to state explicitly the quantity of the cognition, that is, of the cognized structure.

a. This concept is needed because all expressions of cognitions have to be given a *de re* interpretation. For example, if I see what is objectively a cat, I may say (i) "a cat," or I might say (ii) "a furry white object." In (i) I see the object as possessing cat-ness; in (ii) as possessing the properties of being furry and being a white object. As the *thing* I see is the qualificand, in (i) cat-ness is the *limiting property* of being the qualificand. It is, therefore, a property under the mode of which the thing is cognized. In (ii) the two properties are the limiting properties of being the qualificand in the cognition. There are Navya-Nyāya philosophers who hold that in this instance, the limiting property of being the qualificand is the same as the qualifier of the cognition. But if we take a more complex case of cognition, of a brown cat (or "a cat is brown"), we have here three properties: (i) the thing that I cognize is under the mode of being a cat (cat-ness); (ii) the brown color as a qualifier of the thing, cat; and (iii) the color is cognized under the mode of being brown (brown-ness). Now in (i) the property of being a cat is the limiting property of *being the qualificand*. In (iii), the property—being brown (brown-ness) is the limiting property of *being the qualifier* (qualifier-ness). Here the limiting properties of qualificand-ness and qualifier-ness are different from the qualifier, the brown color.

The cognition of a brown cat is expressed in Navya-Nyāya as "a cat possessing brown color." Now "possessing" indicates a relation. This relation in this instance is inherence—the brown color *inheres* in the cat. So, inherence, which goes with the qualifier, is the limiting relation of being the qualifier (qualifier-ness) of the cognition.

It has to be noted that Navya-Nyāya has devised a technical language in which the structure cognized is stated fully, so that the cognition itself is specified. To develop this technical language, Navya-Nyāya has to first develop a method of forming abstract terms, then the concept of determiner-determined, and last the concept of limiting properties and limiting relations. It has to be noted here that these limiting properties cannot be meant by words in ordinary nontechnical language. For a word means an object only under a mode, so if a word is used for the limiting property, the mode under which the object meant by the word is cognized would be a higher order abstract. For example, when I perceive something as a cat, I say only "a cat," but if instead, I say "something possessing cat-ness," then "cat-ness" is being used to mean cat-ness under the mode cat-ness-ness, and so on, ad infinitum. If, on the other hand, the word "cat-ness" is not used but only cognized as a mode under which a cat is perceived, then cat-ness is cognized in and through itself, not requiring any mode of cognition. So "cat-ness" can only be used in the *technical* language, which makes explicit the structure cognized.

Thus it appears that the ordinary Sanskrit, which is used to state what is cognized, is like the object language. And the technical language is like the metalanguage. But there is a fundamental difference between the Western concepts of object language and metalanguage and Navya-Nyāya concepts of ordinary Sanskrit and technical language. The difference is due to the Navya-Nyāya theory that all expressions in ordinary Sanskrit stating cognized structures have to be given a *de re* interpretation. So the technical language, too, talks about objects, not about names of objects. The technical language does not mention any names, but uses names of objects, which are used in ordinary Sanskrit. But, as is obvious, the technical language is essentially richer than ordinary Sanskrit as it uses words for all concepts (objects) introduced into the technical language. These words, too, have to be given a *de re* interpretation. According to the Navya-Nyāya philosophers, *being limited* by a limiting property or a limiting relation is itself as objective as a cat, cat-ness, being a qualificand, or a qualifier or a qualification. But whether the properties like being a qualificand (qualificand-ness), qualifierness, limiter-ness, and so on belong to separate categories of reals is debated in Navya-Nyāya. For all Navya-Nyāya philosophers do not hold the same view on this point.

b. Navya-Nyāya philosophers use the term "limit" very often in the second sense to state the quantity of cognized structure. According to Western logicians, every judgment has both a quantity and a quality, which determine its form. Navya-Nyāya logicians make a radical distinction between quality and quantity of judgments. The quality of judgment is always expressed by means of the negative particles, used once, twice, or thrice. Whatever is expressed by words becomes a part of the content of the judgment; the words cannot determine the forms. The quantity of judgments is always understood and is therefore to be expressed only in the technical language by means of the concept of limiter. I give here an example of negation. Thus, in ordinary Sanskrit when one says absence of (a) cow it is not clear what he has cognized. Is it the absence of a particular cow or of all cows? This distinction is made in the technical language by using the concept of a limiter. Thus if the limiter of the property of being negated (counterpositive-ness of the negation) is cow-ness in general, then absence of all cows is what is cognized. If, on the other hand the limiter of the counterpositiveness is this cow-ness (or this-ness and cow-ness) or that-cow-ness (or that-ness and cow-ness), then absence of this or that is cognized. Thus the concept of limiter is used here to indicate the quantity of what is cognized.

5.3. Some Logical Words

5.3.1. Negation

Navya-Nyāya logicians treat negation not as belonging to the form of cognitions but as belonging to their matter. This means that a negative cognition cannot be true without a negative object corresponding to the negative cognition. So Navya-Nyāya logicians have posited the reality of negative objects.

From any term "t," its negative can be formed by the suffix "—— *abhāva*" (-negation), which we symbolize by "(t) —— neg." In the ontology there is a negative object as the referent of "(t) —— neg," where the referent of "t" is the *counterpositive of the referent of* "(t) —— neg." Just as we have positive objects like tables, chairs, and so on, so also we have negative objects like absence of a chair, difference of a table from a chair, and so on. Thus Navya-Nyāya theory of negation is a theory of negative terms and their referents. There is no concept of propositional negation in Navya-Nyāya. Negative objects are of two types—absence and difference. Absence, again, is of three types, (i) absence of an object before it is produced; (ii) absence of an object when it is destroyed; (iii) the absence which is eternal like absence of color from air, since air is never colored. The first two types of absence, the not-yet type and the no-more type, are limited in time, the former is beginningless but ceases to be with the object; the latter is endless but begins with the destruction of the objects. The never-type absence is both beginningless and endless.

Navya-Nyāya logicians put some restrictions on objects, which can be said to be absent: that is, objects, which can be counterpositive of absence. (i) The first restriction is, if the object be a global property, or a ubiquitous object, it cannot be said to be absent, for there can be no locus of its absence. (ii) The second restriction is that a purely imaginary or fictitious object cannot be said not to exist in a certain locus, for it does not exist at all. To say significantly that an object does not exist in a certain locus, it is necessary that it exists elsewhere. Gaṅgeśa (circa thirteenth century A.D.) has given the example of "hare's horns do not exist" and has interpreted it to mean that there are no horns on a hare. As horns do exist somewhere, they can be significantly said not to exist in a certain locus. Now to exist elsewhere an object has to be related to its locus by a certain relation. To say that the object does not exist in this room is to say that the object does not exist in this room by that relation in that it exists in some other locus. This relation is technically called the limiting relation of the counterpositiveness of the absence. If this relation is not understood as the limiting relation of counterpositiveness of absence, then there will be no opposition between presence and absence of the same object. Thus we say correctly a pot is present in its parts by the relation of inherence, but is not present in them by the relation of contact. A pot is on the ground by a relation of contact, but it is not present in its parts by contact; it inheres in its parts. Thus the counterpositive and the negation are opposed only if the limiting relation is understood.

The difference between the never-type of absence and difference is explained by reference to the limiting relations. We symbolize difference of t as "(t)-diff" and absence of t as "(t)-ab." Thus when we say that the pot is not the cloth, the limiting relation of the counterpositiveness resident in the cloth is identity. This limiting relation of identity is always present in cognition when we say that this is not that, or that this is different from that. But it will not do to say that the pot is not identical with the cloth, for in this cognition it is the identity of the cloth that has become the counterpositive. Because negations differ if their counterpositives differ, the two negations cannot be the same. This means that identity, which is the limiting relation of counterpositiveness, has always to be understood, but never said.

In the case of absence of the third type, the limiting relations of the counterpositiveness can never be identity, but will always be other relations. The not-type and the no-more types of absence do not have limiting relations of their counterpositiveness. That is because, in the not-yet type of absence, the counterpositive has not come into being; so it does not make sense to say in which relation it is present somewhere, so also with no-more type of absence. Here the counterpositive is destroyed and cannot be said to be present somewhere else. It is only general absence, which has limiting relations of counterpositiveness.

The limiting relations of counterpositiveness are relations in which the counterpositive is located elsewhere. These relations are different from the relations by which negations are located in their locus. For example, if there is no pot on the ground, there is absence of pots located in the ground; this relation is very different from the limiting relation of the counterpositiveness of this absence.

Although I have explained the concept of limiter (limiting property) as stating generality of cognitions, and the concept of limiting relation only in case of negations, Navya-Nyāya philosophers use these concepts extensively in various contexts.

5.3.2. Double Negation

As negation has two forms, absence and difference, double negation has four forms.

i. a. Absence of absence of $t = t$,

where "t" is any term whatever.

This identity has been found defective by Raghunātha (circa fifteenth century A.D.) His argument is that the lefthand side of the identity is a negative object (being absence), but the righthand side may be a positive object. It is not possible to identify a negative object with a positive object. He therefore revises the identity thus:

b. Absence of absence of absence of t = absence of t.

Now both sides of the identity sign are negative objects. Raghunātha does not find any difficulty in identifying the third absence with the first absence.

ii. a. Absence of difference of $t = t$-ness.

The argument for this law is that difference of t is present everywhere except t. So it's absence is present in t. What is present in all t's and only t's is t-ness. So the absence is t-ness.

But this argument has been challenged by many Navya-Nyāya logicians. Their argument is that absence of difference is identity. So they have the following law.

c. Absence of difference of $t = t$ (i.e., identical with t). (For a different form of the argument, see Ingalls.)[80]

iii. Difference from absence of A: This difference is located in everything except absence of A. For example, if B is different from A, then absence of B is different from absence of A. Moreover, absence of A is different from B, which is a positive object. This positive object is the locus of the negative object–(difference).

iv. Difference of difference of A from B: It is also a negative object-difference. It is present everywhere except difference of A from B, which is different from difference of C from D. So difference of difference of A from B, is located in the difference of C from D, also from both C and D, and even A and B, for all of them are different from difference of A from B.

Bimal Matilal has, however, explained Raghunātha's theory in his *Pada-ārthatattva-nirũpaṇa* in the following way. "Fl. 'Difference from whatever is different from anything that has x or is delimited by $x = x$'." Matilal, however, rejects the usual formulation: "The difference from difference from $x = x$."[81]

5.3.3. Conjunction

As in the case of negation, so also with conjunction, the conjunctive particle ("and") occurs only between two terms. Terms may denote either simple or complex objects. The meanings of sentences are complex objects, not propositions.

The realistic metaphysics of Navya-Nyāya, explains conjunction as obtaining metaphysically in cognitions. The reason is that although we may say "both a horse and a cow," still in reality there is no collection of objects. The objects remain metaphysically separate. So a conjunction is really a conjunctive cognition that has two principal qualificands, in the example, a horse is the qualificand of horseness, which is its qualifier, and a cow is also a qualificand of cowness, which is its qualifier. But neither a horse nor a cow is a qualifier of each other. So both of them are principal qualificands.

5.3.4. Disjunction

The disjunctive particle has two forms: "either-or," and "either one of the two." "Either-or" is used to express doubt. For example, when we are not sure whether the thing before us is a post or a man, we say that it is either a post or a man.

Ingalls[82] explained the sense of "either one of the two" as explaining De Morgan's law: $p \vee q = {\sim}({\sim}p \,.\, {\sim}q)$ (p. 64), and ${\sim}(p \vee q) = ({\sim}p \,.\, {\sim}q)$ (p. 65) (he uses "$\sim$" to symbolize "absence"). Thus "either one of the two" is interpreted as inclusive "or."

5.4. Theory of Inference

5.4.1. General Introduction

Inference is a form of mediate, informative, cognition. It is cognition of something by cognizing something else. The process of inference is of two types—(i) inference for one's sake, and (ii) inference for the sake of others. After performing an act of inference for one's sake, one may express this inference in language to communicate to others. I shall, first, explain inference for one's sake by a stock example.

Inference for One's Sake We shall have to cognize certain objects to draw the conclusion from them. In Nyāya there is no immediate inference. For example, we first see that (i) the hill possesses smoke. This is perceptual cognition. Then seeing the smoke, we are reminded that (ii) wherever there is smoke there is fire. This second step is memory of pervasion of smoke by fire. Then (i) and (ii) are combined to make a complex sentence called "Consideration" as exemplified in

iii. The hill possesses smoke pervaded by fire.

And then the conclusion (iv) follows.

iv. The hill possesses fire.

There are several points that have to be explained here.

a. Although inference for one's sake does not involve any language, yet for the discussion of this type inference language has to be used. This language has a canonical form. The verb "possess" functions like the copula in traditional Western logic. In the sentence "the hill possesses fire" the hill is the subject and fire is the predicate. But the sentence "this is fire" is not in its canonical form, so it has to be transformed into "this possesses fireness." Then the predicate is fireness, not fire (and never "fire").

b. The predicate of the conclusion is called “the probandum” (*sādhya* abbrev. “*s*”) (roughly the major term) and the subject of the conclusion is called the “locus of the inference” (*pakṣa*, abbrev. “*p*”) (roughly the minor term).

c. The most important cognition is the memory cognition of pervasion recalled by the perception of smoke. Pervasion is the relation between the pervader (fire in the example) and the pervaded (smoke). This relation can be stated in language thus. Whatever possesses the pervaded possesses the pervader. In symbols, in “x pervades y,” x is the pervader and y is the pervaded.

It is necessary to indicate the relation the pervader has to the locus of the inference, and also the relation the pervaded has to it. In the locus of the inference the pervaded occurs in a certain relation; then one can infer that in the locus of inference the pervader occurs in a certain relation. The relation in which the pervaded occurs in the locus of the inference (the hill, in the example) is the limiting relation of pervadedness; and the relation in which the pervader occurs in the locus of inference is the limiting relation of pervaderness.

d. In pervasion, of course, wherever the pervaded occurs in the limiting relation of pervadedness, the pervader occurs in the limiting relation of pervaderness. “Pervasion” may be defined in symbol thus: x pervades $y = \text{Df.}(z)$ $[Hy, Ryz \supset Sx \supset Txz]$, where R stands for the limiting relation of being pervaded, $Hx = x$ is the probans (the *hetu*) of the inference, and $Sx = x$ is the probandum (the *sādhya*), and T stands for the limiting relation of pervaderness.

R and T are not necessarily different for in some cases they may be the same relation. As in the stock example, wherever smoke occurs by the relation of contact, fire occurs there by the relation of contact. So in this case contact is the limiting relation of pervadedness and also the limiting relation of pervaderness.

Although in Navya-Nyāya inference the sentence expressing pervasion is a universal sentence, still in many inferences, the pervasion is between two particulars, hence there is no scope for generalization. For example, if one eats a ripe mango in a dark room, one may infer thus. This possesses this color, the reason being this taste. The pervasion will take the form: Whatever possesses this taste possesses this color. Even though the sentence is universal there is only one instance of this taste and this color, namely, this mango. So there is no universalization here.

e. The most important difference of Nyāya theory of inference for one’s sake from the usual traditional syllogism is to be found in step (ii) in the example. In traditional syllogism, it is said that the conclusion follows from the conjunction of the two premises. This conjunction requires that the middle term be specified. Without this specification of the middle term, it is not possible to have conjunction of the premises. For example, in

1. All smoky things are fiery
2. The hill is smoky thing
3. $\therefore$ The hill is fiery

(1) and (2) can be conjoined. But Nyāya philosophers point out that it is not necessary to specify the middle term at all. For example, if we want to have the conclusion,

4. The hill is fiery,

the minor term and the major term are given; but what about the middle term? It may be one of many things, smoke, light, and so on. So we may have two different Western syllogisms thus:

A. 1. All smoky things are fiery
   2. The hill is a smoky thing.

B. 5. All things which emit light are fiery.
   6. The hill has a thing, which emits light.

From this we can draw the same conclusion, (3) the hill is fiery.

So even if the minor and the major terms are fixed, the middle term is not always fixed. So Nyāya has the more complicated qualificative cognition, which shows how the conclusion follows from the premises taken together thus:

Consideration (x) (the hill possesses x, and x is pervaded by fire).

From this we can conclude that the hill possesses fire.

What is necessary is that the pervaded (roughly the middle term) should be related to the minor term and the major term in the required manner. It is not necessary to further determine the specific nature of the middle term.

f. There is a fundamental difference between traditional Western logic and Navya-Nyāya on the interpretation of universal sentences, especially those expressing pervasion. On traditional Western logic, (i) "Socrates is mortal," has a subject and a predicate; but so also (ii) "all men are mortal" has one subject and one predicate. But on Navya-Nyāya theory, (ii) has two predicates. In its canonical form it will be "whatever possesses humanity possesses mortality," where humanity and mortality are the two predicates, and humanity is pervaded by mortality. This Navya-Nyāya interpretation of universal sentences has, therefore, some affinity with the interpretation of modern symbolic logic.

Inference for the Sake of Others In this type of inference, language has to be used for communication. Navya-Nyāya has a fixed number and order of the sentences used to convey the inference to others. There are five different sentences, which make one long sentence, thus:

Proposition: (1) The hill possesses fire.

This sentence states what is to be proved.

Reason: (2) The reason is smoke.

This sentence states the reason for proving (1).

> Pervasion with example: (3) Wherever there is smoke, there is fire as in the kitchen.
>
> Application: (4) The hill is similar (possesses smoke).
>
> Conclusion: (5) $\therefore$ The hill possesses fire.

The conclusion (5) is the same as (1); that is, what is to be proved has been proved. The part sentences (1)–(4) are supported by the four different sources of knowledge admitted in Nyāya. Thus step (1) produces only verbal cognition. Step (2) is established by inference; step (3), especially the example, is established by perception. Step (4) is due to analogy based on a cognition of similarity. As the steps (1)–(4) are supported by all the four sources of knowledge admitted in Nyāya, the whole sentence from (1)–(5) is called "supreme demonstration" or "demonstration par excellence" (*parama-nyāya*).

Here the presence of the example in step (3) obtained by perception shows that the whole inference has material truth. For the pervasion is based on this example. All other steps are supported by means of knowledge admitted in this system. All of them, therefore, are materially true.

5.4.2. Definition of Pervasion

Many defective definitions are first explained and examined and at last the accepted definition is stated and explained. I shall explain here only one defective definition of pervasion, and then the accepted definition.

Defective definition: A definition of pervasion is needed, for in the ontology of Nyāya, relations are not admitted as a separate category; only one particular relation, inherence, is given an independence status as real. Pervasion is not inherence, so it has to be reduced to one of the seven categories of real admitted in the system. The defective definition that I state and explain reduces pervasion to absence, which is a negative object. The definition is the following.

> Pervasion is the absence of the occurrence of the probans (*hetu*) in every locus of absence of the probandum (*sādhya*).

Because pervasion is not a necessary relation between the pervader and the pervaded, I have used "$\supset$" in the definition just given. Very roughly this definition of pervasion is the definition of "$\supset$" in terms of negation and conjunction, thus,

> $p \supset q \ = \ \text{Df.} \sim(p \ . \sim q)$ where "p" stands for the occurrence of the probans in a locus and "q" stands for the occurrence of the probandum in the same locus.

The whole definition of "$\supset$" is a negation just as the defective definition stated in Nyāya.

This definition of pervasion is defective because it involves the absence of a probandum in a locus. But it may be that the probandum is a global property, which does not have any absence anywhere. The Nyāya example of such a property is object-hood; everything real is an object of true cognition. Even though ordinary persons cannot know every thing real, a yogin can. Surely God, who is omniscient, knows all objects, past, present, and future, in one act of direct knowledge. So everything real is an object, at least of divine omniscience. So if we have an inference like the following:

> (1) The jar possesses nameability; (2) nameability is pervaded by object-hood; (3) the jar possesses nameability pervaded by object-hood; (4) hence, the jar possesses object-hood.

In this inference, the probandum is object-hood, which pervades nameability. Yet the definition cannot apply here for we cannot say that nameability is absent in every locus of absence of object-hood. But as nameability as well as object-hood is global property, neither can have any locus of absence. So the definition becomes inadequate (too narrow).

The accepted definition of pervasion does not involve an absence of the probandum anywhere. The definition is:

> Pervasion is co-presence of the probans with the probandum such that the probandum is not a counter-positive of any absence, which may occur in any locus of the probans.

In every locus of the probans, something or the other would not occur; that is, that something will be the counterpositive of its absence. But the probandum will not be a counterpositive of any absence in any locus of the probans. This definition does not involve any absence of the probandum; it is simply different from being a counterpositive of absence of whatever in a locus of the probans. So this definition does not suffer from the inadequacy of the defective definition.

5.4.3. Defective Probans (Fallacy)

The term *hetvābhāsa* is interpreted to mean either "defective probans" or "a defect of the probans." Here we shall use the first interpretation.

Definition: A defective probans is the object of a true cognition, which blocks the conclusion of the inference.

We may explain this by an example. If someone infers from (i) $q \supset p$. (ii) p to (iii) q, then he does not know the difference between $q \supset p$ and $p \supset q$. If he comes to know this distinction, then he will not make the inference. His knowledge will prevent him from making the inference. We may explain this by an example from Nyāya thus:

1. Fire is pervaded by smoke.
2. The hill possesses smoke.
3. $\therefore$ The hill possesses fire.

Now this inference is fallacious. Anyone who makes this inference does not know the difference between "fire pervades smoke" and "smoke pervades fire." When he realizes this distinction and realizes that smoke is not a probans here, he will not make the inference. It is interesting to note that many Western philosophers have used the notion of blocking without explaining it. For example, Gilbert Herman writes: "We might suggest that an inferable conclusion is essential to an inference only if the assumption that *t* was false would block the inference."[83]

Navya-Nyāya philosophers explain the concept of blocking in terms of the causal law. For any effect there are many causal conditions, some of which may be positive, and some of which may be negative. The effect will be prevented from occurring if the absence of something, which is a causal condition is replaced by its counterpositive. Thus the counterpositive of a negation will prevent or block the production of the effect if the negation were necessary for it. In the case of cognition, Navya-Nyāya philosophers formulate the following conditions. In the case of one cognition blocking another, there will be two cognitions—the preventer cognition and the prevented cognition. There are conditions for these two types of cognitions.

1. Conditions for the preventer cognition:
   - i. The cognition must be attended with firm belief, it must not be a supposition, assumption, or doubt.
   - ii. It may be true or false, in either case it must not be cognized, in a second-order introspective cognition, to be false or even doubted to be false.
   - iii. It must have the negation of the object cognized by the prevented cognition, as its own object.
2. Conditions for the prevented cognition:
   - iv. The cognition may be either true or false.
   - v. It may not be attended with belief. For even a doubt may be prevented from occurring.
   - vi. It must not be a supposition or assumption.
   - vii. It must not be an ordinary perception, or an illusory perception due to psychophysical defect. No perception can be prevented by any cognition. It can be prevented only by the presence of noncognitive impediments, like absence of light for visual perception, and so on.
   - viii. The cognition must be a qualificative cognition.

There is a detailed discussion of these conditions, which we do not give here. But there is one point, which should be noted. As (i) shows, a mere assumption cannot prevent or block any cognition. This is, therefore, contrary to what Herman says. According to Navya-Nyāya philosophers, it has to be a firm belief, not a mere assumption, that *t* was false, which will then block the inference. For even when we know something to be a fact, we can still assume its contrary; in such a case, this will be a contrary-to-fact assumption, without any attendant belief.

There are other cases of invalid inferences, for example, if one infers that there is fire in the lake, for there is smoke there. Now this conclusion that there is fire in the lake will be contradicted by perception. As a matter of fact, if the person making the inference observed carefully that there is no fire there, he would not have made the inference. Here the probans, smoke, is defective because it does not exist in the locus of the inference, and the conclusion is contradicted by knowledge derived from perception.

Notes

1. Gautama composed 528 aphorisms of Nyāya philosophy. These *Nyāya-sūtra-s* are included in five parts, each consisting of two chapters. The standard notation for referring to a *Nyāya-sūtra* is to use three digits each separated from the other by a vertical line. The first digit stands for the part, the second for the chapter and the third for the *sūtra* itself. Vātsyāyana has written a commentary (*Bhāṣya*) on the *Nyāya-sūtra*-s on which Uddyotakara has given a gloss known as *Nyāyabhāṣyavārttika*. Vācaspati has written a commentarial work called *Tātparyaṭīkā* on which again Udayana has written his commentary called *Pariśuddhi*. References to these authors have been given in the body of the text by mentioning the number of the *Nyāya-sūtra*. These can very easily be checked by consulting any of the available editions of these works. On our part we have used the Metropolitan edition (Gautama's *Nyāya-sūtra*, edited along with *Bhāṣya*, *Vārttika*, *Tātparyaṭīkā*, and Viśvanatha's *Vṛtti* by Amarendramohan Tarkatirtha and Taranatha Tarkatirtha, Metropolitan Printing and Publishing House, Calcutta 1936–1944; reprinted by Munshiram Manoharlal, New Delhi, 1985). For Udayana we have used the Mithila edition of *Nyāya-sūtra*, edited with *Bhāṣya*, *Vārttika*, *Tātparyaṭīkā* and *Pariśuddhi* by Anantalal Thakur, Volume I, Mithila Institute of Post-graduate Studies, 1967.

2. For a detailed account of these topics see chapters 2 and 3 of Matilal (1998) and chapter 6 of Saha (1987).

3. In addition to the two cognitions mentioned by Vātsyāyana, Navya-Nyāya admits *parāmarśa* as causally necessary for inference. It is taken to be an instance of intuitive knowledge (*mānasapratyakṣa*) of a composite proposition about the presence of reason in the subject of inference and also about the concomitance between the probans and the probandum. Philosophers belonging to this school were probably influenced by the views of Uddyotakara and Vācaspati. They hold that (NS 1/1/5) a cognition by testimony results from the fourth step in argument called *upanaya* that involves reference to earlier steps. Such a testimonial knowledge is also about a composite content.

4. Philosophers before Gautama, for example, Caraka and Kaṇāda have given many examples. For Caraka's examples see Matilal (1985), and for Kaṇāda see Tarkavagisa (1981) 179.

5. See ND, vol. 1, p. 169 for these examples.

6. *Tadhi sāmānyatodṛṣṭam yathā vahnyanumānādikam*. Quoted in ND, vol. 1, p. 179.

7. ND, vol. 1, p. 178.

8. LLR, pp. 30–31.

9. This concept of *pramāṇa-samavāya* is certainly different from that of *pramāṇa-saṃplava*, which Vātsyāyana introduces in his commentary on NS 1/1/3 inasmuch as the former hints at some kind of adequate or sufficient justification while the latter refers to the idea of corroboration of the same fact which can be ascertained by different *pramāṇa*-s or independent lines of justification.

10. See PNLE, chapter 6.

11. See CLI, pp. 6–11 for views of other schools.

12. CLI, p. 5.

13. For a fairly elaborate discussion of the Buddhist science of debates covering rules of debate and rejoinders, tricks, and defeat situations mainly reconstructed from the Chinese-Tibetan texts, one may consult Matilal (1998).

14. Vide *Pramāṇa-samuccaya* 2.5, translated in Hayes (1998). Diṅnāga writes, "[the sign] must be present in object and what is similar to it and absent in their absence."

15. We must remember that according to most Buddhists, even atoms are impermanent. They are actually borrowing this example from their opponents.

16. *Pakṣadharmatā* means the property of the sign's residing in the subject-location.

17. *Vyāpti* is the relation of universal concomitance that obtains between the sign and the signified.

18. For a detailed defense of Diṅnāga's stance, see Matilal (1998).

19. *Tattvasaṃgraha* and its commentary, *Pañjika*, by Kamalaśīla, were written in the eight century A.D. For Pātrasvāmī's criticism and Śāntarakṣita's defense see verses 1364–1380.

20. Ibid., p. 99.

21. *Śiṃśapā* is a kind of tree. In cases like "whatever is existent, is momentary," we have *vyāpti* of the *tādātmya* type.

22. He writes in *Svārthānumānapariccheda*, "invariable concomitance between two items cannot be known merely from simple observation of things or failing to have the required properties."

23. For a detailed discussion of this point of view please see Shah (1967).

24. Nyāyabindu III.31.

25. See Viśālāmalavatī commentary on *Pramāṇa-samuccaya* by Jinendrabuddhi.

26. Nāgārjuna asks in *Mādhyamika-kārikā*, a. Does a thing come out of itself? No. b. Does it come out of the other? No. c. Does it come out of both, itself and the other? No. d. Does it come out of neither? No.

27. See note 30.

28. JTB: 16, §50 §51; PKM: 3/37, PNT: 3/28.

29. Dharmakīrti, *Nyāyabindu*; 81, ch. 3. Also supported by Devasūri in PNT, 3/23.

30. See Anantavīryācārya's (PRM) note on sūtras 36, 37, of Mānikyanandī's PMS.

31. JTB: 12; PMS: 2/52,53; PNT: 3/9,10; PM: 1/2/8,9.

32. Translation of *Tarkasaṃgraha*, §52, §53, in Bhattacharya (1983).

33. See Phanibhusana Tarkavagisa (PBT), vol. 1, 288; Bhattacharya (1983).

34. PBT, vol. 1, 288.

35. JTB: 12 §34; Buddhist logician Śaṇkarasvāmī in his *Nyāyapraveśa* used two distinct names—*sādhanā* and *anumāna* for *parārtha* and *svārtha anumāna*,

respectively. This suggests that an awareness about the difference in logical status of the two types of inference was vaguely present in some logicians.

36. See PBT, 288–289, for detail.

37. Dharmottara, *Nyāyabindutīkā*, chapter 3.

38. Śaṇkarasvāmī, *Nyāyapraveśa.*

39. *Syādvādasiddhi*, 4/82, 83.

40. Mānikyanandī, PMS, 3/15.

41. Vidyāratna, P-Par.

42. Siddhasena N Av. Ka-22.

43. Akalaṇka AGT, 102 and P-Sam ka 21.

44. *Dhavalatīkā* on *Ṣatkhaṇḍāgama* by Vīrasena, p. 287.

45. Akalaṇka, *Nyāyaviniścaya.*

46. Dharmabhūṣaṇa, ND, 19 from TSV; Mānikyanandī, PMS, 3/2.

47. Udayana, *Nyāya-vārttika-tātparya-parisśuddhi*, 1/1/5. Also Yaśovijiaya, JTB, 11.

48. Akalaṇka, AGT, 100.

49. *Anyathāsambhava siddheranavasthānumānataḥ.*

50. See JTB, 10; JTB, 11 (44b); JTB, 10–11; JTB, 12; JTB, 12; JTB, 16; JTB, 11. Vidyānanda, TSV, 1/13/84–119; (a) *uhaṃ vinā jñātena sāmānyenāpi sakalavyaktyanupasthiteśca* (JTB), (b) *tarkāt tanniścaya* (PMS, 3/21), (c) *sa ca hetu svarūpam tat hi antarvyāptisca viddhiḥ naḥ* (SS, 78–97), (d) *antarvyāpterataḥ saiva gamakatva prasādhani.* Compare (a) (b) with (c) (d) and the similarity of the conceptual roles of and the close link between *tarka* and *antarvyāpti* in establishing inseparable concomitance (*sādhya-sādhna avinābhāva*) becomes obvious.

51. Vādidevasūri, PNTL, 3/38–39; Siddhasena (Nyav.: 20) concurs, but Yaśovijaya (JTB, 12) considers this classification of *vyāpti* untenable as *vahirvyāpti*, by itself, is totally ineffective.

52. See PBT, vol. 1, 339–341 for further detail.

53. *Antarvyāpteḥ sādhya saṃsiddhi-śaktau bāhyavyāptervarṇanam vandhyāmeva antarvyāpteḥ sādhyasaṃsiddi aśaktau bāhyavyāptervarṇanam bandhyāmeva.*

54. S. C. Nyāyacārya, JDD, 39–40.

55. See note 27 and Yaśovijaya, JTB, 12.

56. See Arcaṭa, HBT.

57. Diṅnāga; Dharmakīrti, *Nyāyabindu.*

58. Yaśovijaya, JTB, 12–13.

59. Akalaṇka, N. Vini, 127–130 and Yaśovijaya, JTB.

60. Vādirāja, *Nyāyaviniścayavivṛti*, 2/196.

61. Akalaṇka, N. Vini, 2/167.

62. Akalaṇka, N. Vini, 2/120.

63. Dharmabhūṣaṇa, ND, 102–103.

64. Yaśovijaya, JTB, 18.

65. Yaśovijaya, JTB, 14.

66. Yaśovijaya, JTB, 14.

67. Haribhadra, SDS, 123–125.

68. Haribhadra, SDS, 126.

69. Haribhadra, SDS, 126.

70. Haribhadra, SDS, 126.

71. Yaśovijaya, JTB, 15.

72. Prajaha Sūtra, Bhāṣāpāda, 15–19.

73. Ibid.

74. Haribhadra, SDS, 215.
75. Haribhadra, SDS, 212.
76. Mahalanobis (1990).
77. Kothari (1985), Ghosh (1991), Srinivas (1988).
78. Kothari (1985).
79. Haribhadra, SDS, 212–216.
80. Ingalls (1951), 71.
81. Matilal (1985), 199ff.
82. Ingalls (1951).
83. Herman (1973), 123.

Primary Sources

Akalaṇka. *Akalaṇka-grantha-traya*, ed. M. K. Shastri. Singhi Jain Series 12, Ahmedabad, 1939.

Akalaṇkadeva

i. Nyāya Viniścaya (N. Vini). 1954.

ii. Akalaṇka *Grantha Traya* (AGT, 1953), ed. M. K. Jain, Sindhi Jain Granthamālā, Ahmedabad [AGT includes PS and LS].

Akṣapāda Gautama. *Nyāyadarśana*, with the *Bhāṣya* of Vātsyāyana, the *Vārttika* of Uddyotakara, the *Tatparyatīkā* of Vācaspati and Parśiuddhi of Udayana, ed. Anantalal Thakur. Mithila Institute Series Ancient Text No. 20, Muzaffarpur, 1967.

Anantavīrya (Jr.). *Prameya Ratna Mala* (PRM), V.S. 2020 Chowkhamba, Varanasi.

Arcaṭa. *Hetubindu Tīkā* (HBT). Oriental Institute, Baroda, Asiatic Society, Calcutta, 1941.

Bhadrabāhu. *Daśavaikālika Niryukti* (DVN). Agamodaya Samiti, Surat.

Devasūri (Vādidevasūri)

i. *Pramāṇanaya Tattvāloka* (PNT).

ii. *Pramāṇanaya Tattvālokālaṇkāra* (PNTL).

Dharmabhūṣaṇa. *Nyāyadīpikā* (ND). Vir Seva Mandir, Delhi, 1945.

Dharmakīrti. *Nyāyabindu* with Dharmottara's *Nyāyabindutīkā* and Durveka Miśra's *Pradīpa*, ed. V. Malvania. K. P. Jayasawal Research Institute, Patna, 1955.

Dharmakīrti. *Hetubinduprakaraṇa*, with Arcaṭa's commentary and Durvaka Mirśa's sub-commentary. Gaekwad Oriental Series No. 113, Oriental Institute, Baroda, 1949.

Dharmakīrti. *Pramāṇavārttika*, ed. Swami Dwarikadas Shastri. Buddha Bharati, Varanasi, 1968.

Dharmakīrti. *Sopajñavṛtti* on *Svārthānumāna* of *Pramāṇavārttika*. Kitab Mahal, Allahabad, 1943.

Diṅnāga. *Pramāṇasamuccaya* (PS) with Sanskrit retranslation, ed. Muni Jambuvijayaji. Gaekwad's Oriental Series No. 136, Baroda, 1961.

Diṅnāga. *The Nyāyamûkha of Diṅnāga*, ed. G. Tucchi. Heidelberg, 1930.

Gaṇgeśa. *Tattvacintāmāṇi* (TCM). Syādvāda Mahavidyalaya, Varanasi, 1961.

Haribhadra. *Ṣaḍdarsanasamuccaya* with Guṇratna's Commentary (308), ed. L. Suali. 1986.

Hemacandra. *Pramāṇamīmāṃsā* (PM). Sindhi Jain Granthamala, Ahmedabad, 1939.

Jinendrabuddhi. *Viśālāmalavatī* on *Pramāṇasamuccaya*. Oriental Institute, Baroda, 1961.

Mānikyanandī. *Parikṣāmukha Sūtra* (PMS). Pt. Ghanasyamdas Jain Syadvada Mandala, Varanasi.

Mokṣākaragupta. *Tarkabhāṣā*, ed. E. Krishnamacharya. Oriental Institute, Baroda, 1942.

Nāgārjuna. *Mūlamādhyamaka-kārikā*, ed. P. L. Vaidya. Bauddha Samskrita Granthavali No. 10, Mithila Vidyapith, Darbhanga, 1960.

Prabhācanadra. *Prameyakamala Martaṇḍa* (PKM). J. Nirṇaya Sāgara Press, Bombay, 1941.

Pujyapāda. *Sarvārthasiddhi* (SS). Bharatiya Jñānapitha, Varanasi, 1955.

Ratnakīrti. *Ratnakīrtinibandhāvalī*, ed. Anantalal Thakur. K. P. Jayasawal Research Institute, Patna, 1957.

Sāmantabhadra. *Āptamīmāṃsā* (AM). 1967.

Śaṇkarasvāmin. *Nyāyapraveśa* (N. Prv.), Part I, ed. B. Dhruva. Gaekwad Oriental Series No. 38, Oriental Institute, Baroda, 1930.

Śāntarakṣita. *Tattvasaṃgraha* with Kamalaśīla's *Pañjikā*, ed. Swami Dwarikadas Shastri. Buddha Bharati, Varanasi, 1968.

Siddhasena. *Nyāyāvatāra* (Nyav.). V.S. 2004, Bharatiya Vidyabhavan, Bombay.

Udayana. *Nyāya-vārttika-tātparya-pariśuddhi* (NVTP). Govt. Sanskrit College, Calcutta, 1911.

Vādibhasiṃha. *Syadvādasiddhi* (Sy.S). Bharatiya Digambara Jaina Granthamala, Bombay, 1950.

Vādirāja. *Nyāyaviniścaya Vivṛti* (N. Vin. V). Bharatiya Jñānapitha, Varanasi, 1954.

Vidyānanda. *Tattvārthaśloka-vārttika* (TSV). Seth Rama Candra Natharang Trust, Bombay, 1918.

Yaśovijaya. *Jaina Tarkabhāṣā* (JTB), ed. D. N. Bhargava. MLBD, Delhi, 1973.

Secondary Sources

Bhattacharya, H. M. 1993. *Jaina Logic and Epistemology*. Calcutta: K. P. Bagchi.

Bhattacharya, Kamaleswar. 1986. Some Thoughts on *Antarvyāpti*, *Bahirvyāpti* and *Trairūpya*. In *Buddhist Logic and Epistemology*, ed. B. K. Matilal. Dordrecht: D. Reidel.

Brahma, S. 1999. Jaina Logic and Ontology: A Formal-Analytical Approach. Ph.D. dissertation, Jadavpur University.

Ghosh, P. 1991. *Syādvada*, Relativity and Complimentarity. Occasional Paper No. 15. Delhi: Project on History of Indian Science, Philosophy and Culture.

Hayes, R. P. 1988. *Dignāga on Interpretation of Signs*. Dordrecht: Kluwer.

Herman, G. 1973. *Thought*. Princeton, N.J.: Princeton University Press.

Ingalls, D. H. H. 1951. *Materials for the Study of Navya-Nyāya Logic*. Cambridge, Mass.: Harvard University Press.

Kothari, D. S. 1985. The Complimentarity Principle and Eastern Philosophy. In *Niels Bohr: A Centenary Volume*, ed. A. P. French and P. J. Kennedy. Cambridge, Mass.

Mahalanobis, P. C. 1990. The Foundations of Statistics. In *Studies in the History of Indian Philosophy*, vol. 3, ed. D. P. Chattopadhyaya. Calcutta: K. P. Bagchi.

Matilal, B. K. 1970. Reference and Existence in Nyāya and Buddhist Logic. *Journal of Indian Philosophy*, 1.

Matilal, B. K. 1985. *Language, Meaning and Reality*. Delhi: Motilal Banarasidass.

Matilal, B. K. 1998. *The Character of Logic in India*, eds. J. Ganeri and H. Tiwari. Albany: State University of New York Press.

Nyayacharya, S. C. *Jaina Darśaner Dig-darśana* (JDD) [in Bengali]. Calcutta: Govt. Sanskrit College.

Saha, S. R. 1987. *Perspectives on Nyāya Logic and Epistemology* (PNLE). Calcutta: K. P. Bagchi.

Sarkar, T. K. 1992. Some Reflections on *Syādvāda* and *Anekāntavāda*. *Jadavpur Journal of Philosophy*.

Sarkar, T. K. 1997. Jaina Logic in Perspective. In *Essays in Indian Philosophy*, ed. S. Saha. Calcutta: Allied Publishers.

Sarkar, T. K. 1998. Logics beyond Consistency. In *Logic, Identity and Consistency*, ed. P. K. Sen. Calcutta: Allied Publishers.

Shah, N. J. 1967. *Akalaṇka's Criticism of Dharmakīrti's Philosophy*. Ahmedabad: L.D. Institute of Indology.

Srinivas, M. D. 1988. *The Indian Approach to Formal Logic and Methodology of Theory Construction*.

Stcherbasky, T. 1930. *Buddhist Logic*. New York: Dover. Originally published by the Academy of Sciences of the U.S.S.R., Leningrad (1930).

Tarkavagisa, Phanibhusana. 1981. *Nyāyadarśana* (ND), vol. 1. Calcutta: West Bengal State Book Board.

Vidyabhusana, S. C. 1921. *A History of Indian Logic*. Calcutta: University of Calcutta.

This page intentionally left blank

Index