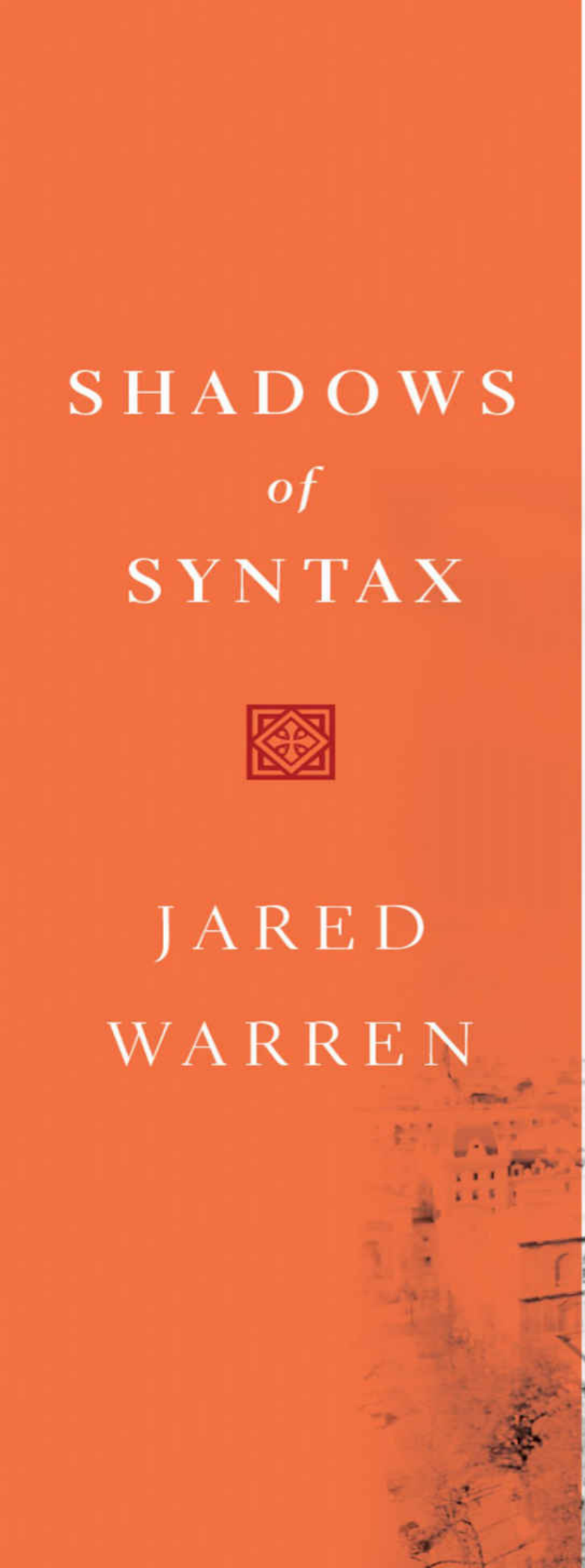
SHADOWS *of* SYNTAX





JARED WARREN

Shadows of Syntax

Shadows of Syntax

Revitalizing Logical and Mathematical Conventionalism

JARED WARREN

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford. It furthers the University's objective of excellence in research, scholarship, and education by publishing worldwide. Oxford is a registered trade mark of Oxford University Press in the UK and in certain other countries.

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America.
© Oxford University Press 2020

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without the prior permission in writing of Oxford University Press, or as expressly permitted by law, by license, or under terms agreed with the appropriate reproduction rights organization. Inquiries concerning reproduction outside the scope of the above should be sent to the Rights Department, Oxford University Press, at the address above.

You must not circulate this work in any other form and you must impose this same condition on any acquirer.

Library of Congress Cataloging-in-Publication Data
Names: Warren, Jared, author.
Title: Shadows of syntax : revitalizing logical and mathematical conventionalism / Jared Warren.
Description: New York, NY, United States of America : Oxford University Press, [2020] | Includes bibliographical references and index.
Identifiers: LCCN 2019035617 (print) | LCCN 2019035618 (ebook) | ISBN 9780190086152 (hardback) | ISBN 9780190086169 (updf)| ISBN 9780190086176 (epub) | ISBN 9780190086183 (Digital OnLine)
Subjects: LCSH: Convention (Philosophy) | Logic. | Mathematics—Philosophy. | Pluralism.
Classification: LCC B809.15 .W37 2020 (print) | LCC B809.15 (ebook) | DDC 160—dc23
LC record available at https://lccn.loc.gov/2019035617
LC ebook record available at https://lccn.loc.gov/2019035618

To the ruin of Eliphaz the Temanite, Bildad the Shuhite, and Zophar the
Naamathite

It gradually became clear that a great part of philosophy can be reduced to something that may be called “syntax,” though the word has to be used in a somewhat wider sense than has hitherto been customary.

—Bertrand Russell (1945)

The irresistible metaphor is that pure abstract objects … and the states of affairs to which … merely minimally true sentences correspond, are no more than *shadows* cast by the syntax of our discourse. And the aptness of the metaphor is merely enhanced by the reflection that shadows are, after their own fashion, real.

—Crispin Wright (1992)

Contents

Preface
Acknowledgments
Introduction

PART I THE PHILOSOPHICAL BACKGROUND

1 What Is Conventionalism?

2 Linguistic Conventions

PART II REVITALIZING LOGICAL CONVENTIONALISM

3 Unrestricted Logical Inferentialism

4 Logical Conventionalism

5 Alternative Conventions, Alternative Logics

6 The Epistemology of Logic

7 The Very Idea of Truth by Convention

PART III REVITALIZING MATHEMATICAL CONVENTIONALISM

8 From Logic to Mathematics

9 Metaontology, Existence, and Reference

10 Mathematical Determinacy

11 Truth, Paradoxes, Freedom, Applications, and Knowledge

12 Metamathematics versus Conventionalism

PART IV REFLECTIONS

13 Old Slogans, New Dogmas

14 The Facts of the Matter

Bibliography
Name Index
Subject Index

Detailed Contents

Preface
Acknowledgments
Introduction

PART I THE PHILOSOPHICAL BACKGROUND

1 What Is Conventionalism?
- I Rise and Fall
- II Three Versions of Conventionalism
- III Explanation
- IV Scientific Naturalism

2 Linguistic Conventions
- I Post-Positivist Accounts of Conventions
- II Conventions as Stipulations?
- III The Case for Inference Rules
- IV Beyond Standard Inference Rules
- V Acceptance and Rejection
- VI Inference
- VII Rule-Following in the Time of Kripkenstein
- VIII Inference Rule-Following
- IX Things We Do with Sentences

PART II REVITALIZING LOGICAL CONVENTIONALISM

3 Unrestricted Logical Inferentialism

I Logical Inferentialism
II Unrestricted Inferentialism
III Which Expressions Are “Logical”?
IV The Meanings of the Logical Constants
V Inferentialism and Understanding
VI Inferentialism and the Metaphysics of Logic
VII Unrestricted versus Restricted Inferentialism

4 Logical Conventionalism
I From Inferentialism to Conventionalism
II Are These Really *Conventionalist* Explanations?
III The Role of Semantic Completeness
IV Conventionalism versus Deflationism
V The Metaphysics of Logical Facts
VI Logically Necessary Connections
VII The Normativity of Logic
VIII Rules and Roles
Appendix: Completeness for Bilateralist Validity

5 Alternative Conventions, Alternative Logics
I The Translation Mistake
II Tonk, Translation, and Holism
III Pluralism and Semantic Counterparts
IV Equivalence between Logics
V Logical Disagreements and Disagreements about Logic

6 The Epistemology of Logic
I The General Strategy
II Arguments and Knowledge
III Inferentialism and Epistemic Entitlement
IV Clearing Out the Bad Company

7 The Very Idea of Truth by Convention
I The Contingency of Conventions
II The Master Argument
III Quine, 1936
IV Variations on a Theme by Quine
V Williamson and Expert Disagreement

PART III REVITALIZING MATHEMATICAL CONVENTIONALISM

8 From Logic to Mathematics
I Mathematical Conventionalism in History
II Mathematical Inferentialism
III Major Problem #1: Ontology from the Armchair
IV Major Problem #2: Determinate Truth
Appendix: Bilateralism and Arithmetical Incompleteness

9 Metaontology, Existence, and Reference
I Trivial Ontological Realism
II The Metatheory of Existence and Reference
III Quantifier Pluralisms
IV Antirealism, Collapse, and Other Antagonisms
V Answering Kant's Challenge
VI The Metaphysical Nature of Mathematical Objects

10 Mathematical Determinacy
I Determinacy, Pluralism, and Categoricity
II Open-Endedness Examined
III Strategy #1: Collapse
IV Strategy #2: Anti-Overspill
V Strategy #3: Communicative Generality
VI From Categoricity to Determinacy?

VII The Omega Rule
VIII Supertasks and Mental Magic
IX Is Open-Endedness Still Needed?
Appendix: The Omega Rule and True Arithmetic

11 Truth, Paradoxes, Freedom, Applications, and Knowledge
I From Inferentialism to Conventionalism (Again)
II As Always, Carnap versus Quine
III A Conventionalist Look at the Paradoxes
IV A Conventionalist Look at Mathematical Freedom
V A Conventionalist Look at Applications
VI A Conventionalist Epistemology for Mathematics
VII A Conventionalist Look at Knowledge of Consistency

12 Metamathematics versus Conventionalism
I The Arithmetization of Syntax
II Pluralism, Consistency, and Consistency Sentences
III Circularity and Syntax
IV Pollock, Provability, and Paradox
V G ödel Against Conventionalism

PART IV REFLECTIONS

13 Old Slogans, New Dogmas
I Conventionalist Slogans
II Carnap's *The Logical Syntax of Language*
III Giannoni's *Conventionalism in Logic*
IV New Dogmas

14 The Facts of the Matter
I Factualism in Metaphilosophy
II How Conventionalism Could Be Factually True

III Philosophical Progress
IV One Against Many
V Doubts
Bibliography
Name Index
Subject Index

Preface

I think that the preface should tell a book's story. So here, in brief, is this book's story.

I decided to write this book in early 2014; the first complete draft was finished near the end of July of that year. But the story of the book starts a little bit before that. I have long been puzzled by logic and mathematics. After learning about Gödel's incompleteness theorems, some of the puzzles seemed more like unfathomable mysteries. I entered graduate school at New York University (NYU) in 2008 with a wavering commitment to logical monism along with developing metaontological views that I hoped would demystify the ontology of mathematics. The pieces wouldn't come together. Some were wrong, others were missing entirely. But slowly, bit by bit, as I puzzled and read and discussed and wrote, a unified picture of logic and mathematics began to take shape.

That unified picture is presented in this book. A few crucial steps along the winding path from there to here are worth mentioning.

In the fall of 2009 I attended Graham Priest's dialetheism course at the City University of New York. I came in antecedently committed to logical monism and classical logic, but during the class I thought more and more about the metasemantics of logic. Throughout I argued against Priest's monist dialetheist view and had wide-ranging discussions with Jack Woods, who was also in attendance. During an exchange of letters with him in late November 2009, I challenged the coherence of Priest's logical metasemantics, and began to explore the possibility of the more charitable, pluralistic approach toward which I was becoming more and more sympathetic.

Things remained in flux for a time. Then my views on logic coalesced in early 2011. I was the teaching assistant for an advanced logic class at NYU taught by R.A. Briggs. Over many conversations with Briggs that spring, I explored my developing

approach. By early April, I had finally broken completely with logical monism in favor of a radical version of pluralism. In assorted discussions I began defending an embryonic version of what I eventually came to call unrestricted inferentialism.

In early August of that year I laid out my newly developed unrestricted inferentialist position in a chat with Kit Fine. Kit pressed me on tonk, arguing that the tonk rules aren't truth-preserving. I countered that his argument illicitly assumed a homophonic translation between the tonk language and our own. Seeing that this position hadn't previously been defended in the literature, I decided to write up my account of tonk. The initial draft of "Talking with Tonkers", my first paper expounding and defending unrestricted inferentialism, was finished less than a week later. That fall I finished a draft of "Change of Logic, Change of Meaning", another important step forward in the development of my logical views.

Soon after this I began to realize that combining my unrestricted inferentialist views with my metaontological views could lead to a form of conventionalism about mathematics. In early November, during my first time attending NYU's thesis preparation seminar, I wrote and presented a paper—"Revitalizing Mathematical Conventionalism"—that represented, in effect, the earliest version of part III of this book. My views developed throughout 2012, and in early May of 2013 I presented another paper—"Revitalizing Logical Conventionalism"—that represented, in effect, the earliest version of part II of this book. Previously, in early February, I had presented a paper that became my "Conventionalism, Consistency, and Consistency Sentences", which answered a technical challenge for mathematical conventionalism and developed a general framework for conceptual pluralism.

By this time I had learned to stop worrying and love the bomb. I was an unrepentant conventionalist about both logic and math. Maybe the first since Carnap, from what I could tell. Throughout the remainder of 2013 and the first half of 2014 I developed my approach in various directions, working on a number of loosely related projects. During this time I wrote the first versions of several relevant publications: "The Possibility of Truth by Convention", "Revisiting Quine on Truth by Convention", "Quantifier Variance and

the Collapse Argument", and "Quantifier Variance and Indefinite Extensibility". I also completed drafts of several relevant but still unpublished papers: "Defending Understanding-Assent Links", "The Epistemology of Validity: Rule-Circularity, Epistemic Rights, and Unrestricted Inferentialism", and "The Determinacy of Arithmetic: The Puzzle and Its Solution" (later to be reworked into "Open-endedness and Arithmetic" before being reworked again, more recently, into two separate papers).

For a number of reasons, I gradually came to believe that my conventionalist theory would be best presented in a book. A book would afford me the space to develop things in detail and to answer the many anti-conventionalist arguments that had piled up over the years. This had became a firm plan by May 2014. During an emotional low point, I feverishly put together the first full draft of the book in July 2014. In addition to writing new material, I also cannibalized parts of many of the unpublished conventionalist papers I had already written. I then revised this draft from scratch, finishing the second complete draft of the book in September 2014. The book—at that time titled *Syntactic Shadows: A Linguistic Theory of Logic and Mathematics*—was then put aside so that I could commence work on a dissertation exploring a new approach to the liar and other paradoxes.

Eventually I realized that it wouldn't have been wise to begin another major research project from scratch, and so, with the approval of my committee, I presented a heavily revised and rewritten version of the logic portion of the book as my dissertation. This was finished in the spring of 2015 and defended in early May 2015 before a panel chaired by Hartry Field and consisting of David Chalmers and Crispin Wright, with Max Barkhausen, Cian Dorr, and Tim Maudlin in attendance.

After graduating I continued to work on the book. A major step forward came in the early summer of 2015, when I wrote the first draft of my "Following the Omega Rule" after an epiphany in a cafe. This added the final major piece to my account of mathematics. After moving to Florida late in the summer of 2015 I devoted all of my efforts to producing another draft of the book, heavily reworking both my thesis and the mathematics portion of the book (now titled

Shadows of Syntax). This was finished at the start of October, but at the time it only led to some extremely unsatisfactory preliminary dealings with a university press. On the basis of that experience, I gave up all hope that the book would ever be published.

After this disappointment, I put the book aside to deal with life and other distractions. But I still intended to finish it eventually, for my own sake. In the summer of 2016 I began another heavy revision, completing parts I and II by December, at which point I again ran out of hope and put the book aside. Then in mid-May 2017 a health scare lit a fire under me. I finished the revision started the previous year, completing parts III and IV at a lightning pace, writing two entirely new chapters and adding several sections to existing chapters. This was finished in early June. Toward the end of June, when my future health looked better, I polished the entire draft, finishing at the end of August. Then, while adjuncting three classes, I undertook another full revision from late October through mid-November before submitting it to OUP for review.

In February 2018 I was offered a tenure-track position at Stanford; in March the book was accepted by OUP with two sets of very positive and helpful comments. I moved to California in the summer; in the fall I taught a graduate seminar going over the book in its entirety, revising it heavily in the process and adding several new sections. I then spent the winter break revising and rewriting and reworking the book several times over. Christmas and New Year's Day were spent alone with the book, writing and researching, aiming for clarity and truth.

I realized that I was going to miss my initial January 1 deadline, so asked for and received an extension to February 1. I then missed that deadline too, as I continued to rework, revise, and polish. I wanted the book to be philosophically and aesthetically perfect. Unfortunately I have fallen far, far short of these goals in ways both major and minor. But the book would never get submitted if I waited until I was completely happy with it. I submitted a "final" version on February 10, 2019, and then made some further, minor revisions in early June, when I submitted the truly final version for typesetting. I corrected proofs through November and December of 2019, then again in February of 2020, then *again* in April and May of 2020.

Conventionalism is a hated view, but my hope is that people will read and enjoy this book despite that. Some of my favorite philosophy books—books I think that every philosopher should read—develop and defend views that I disagree with in the strongest terms (Lewis's *On the Plurality of Worlds*, Williamson's *Vagueness*, Chalmers's *The Conscious Mind*, for but three examples). I hope that *Shadows of Syntax* will be enjoyed even by readers who think conventionalism is crazy. But I can't worry about that anymore. I have to let it go. So I am. Here it is.

There is a much longer, much more dramatic version of this story. It isn't without interest or lessons about academia, but this isn't the place for it. Throughout a long period in the proverbial wilderness I imagined this book as a medieval cathedral, built not for the world but to honor the metaphorical glory of God. The work mattered, to me and for its own sake. But during writing I found myself filled with a hundred different kinds of apprehension. These bubbled forth unexpectedly at the end, where the book glances at its own shadow.

Which feels appropriate.

Palo Alto, California

Acknowledgments

My thanks to:

David Chalmers, Hartry Field, and Crispin Wright for comments on and discussion of early versions of parts I and II, and also for many helpful discussions of related topics when my ideas were developing.

Hartry Field (again) for chairing the dissertation that became parts I and II. And for asking, in 2014, to see the book that he heard I had written.

Crispin Wright (again) for giving me his blessing to use one of the epigraph quotes—unfortunate chronological considerations prevented me from obtaining Lord Russell's blessing for the other.

Eli Hirsch and Daniel Waxman for comments on and discussion of multiple chapters. Aside from referees for Oxford University Press (OUP), they read more of the book than anyone else. Both pushed me to discuss things that I wouldn't have otherwise (and Waxman in particular pushed me to simplify my presentation of several proofs).

Rosa Cao, Joshua Reagan, Lauren Sperber, and Nick Zangwill for comments on a chapter or chapters.

Mark Balaguer, Sharon Berry, Jonathan Matheson, and Michael Strevens for comments on a section.

Matti Eklund and Jack Woods for comments on a full draft manuscript in their capacity as anonymous referees for OUP, and for kindly agreeing to be unmasked after the fact so that I could thank them properly.

Jack Woods (again) for follow-up discussion of his extensive and helpful comments as I finalized the manuscript.

Audiences at Stanford, Toronto, and Yale.

Everyone who attended PHIL 353 during the 2018 Autumn quarter at Stanford, but especially to Nadeem Hussain, Penn Lawrence, Scott Mutchnik, and Declan Thompson.

Jennifer Fisher, Yu Guo, and Nicolas Michaud for many discussions and debates about the philosophy of logic and mathematics, long before this book was even a dream.

Peter Unger for encouraging me when I really needed it.

Peter Ohlin and OUP for giving me a chance.

Everyone else mentioned in the preface or thanked in the main text.

My parents for love, support, and rent-free accommodations during the long dark times when I could afford nothing else.

Introduction

The necessary *a priori* truths of logic and mathematics have long been a stumbling block to a unified picture of the world and our place in it. They drove Plato to Platonic metaphysics and Kant to mysterious epistemological faculties in search of a satisfying account. Empiricists and other scientifically minded philosophers rejected such maneuvers, striving instead for a philosophy of logic and mathematics that stayed fully within the bounds of the scientific picture of the mind and the world.

Broadly speaking, there are only two promising ideas for this: radical empiricism and conventionalism. Radical empiricism rejects the *a priority* of logic and mathematics. For radical empiricists, logic and mathematics differ from physics, chemistry, and biology in degree but not in kind. The conventionalist approach is less revisionary and has been, historically at least, much more popular. Unlike radical empiricism, conventionalism recognizes the special nature of logic and mathematics but attempts to tame it. It does so by seeing the truths of logic and mathematics as utterly trivial. They are akin to truths by definition, stipulation, or fiat. Mere byproducts of our linguistic and conceptual practices rather than substantive features of the world. Though anticipated by Leibniz and Hume, this view reached the apex of its influence in the twentieth century, when it was championed by the Vienna Circle and its followers.

Both of these approaches are widely rejected in contemporary philosophy, but nothing solid or lasting has been put in their place. Worse still, the standard-issue philosopher seems to endorse an embarrassing combination of epistemological pragmatism and metaphysical realism about both logic and mathematics. This is not a coherent combination, so it should be no surprise that philosophical progress has stalled. A new approach is needed, or perhaps an old approach, newly renovated.

I believe that conventionalism can be revitalized. The view that scientists and scientifically minded philosophers once saw as the long-needed breakthrough has more recently been misunderstood and rejected for bad reasons. The way forward is to first take a step backward, by re-examining the case against conventionalism in light of the machinery of contemporary philosophy. When this is done, a stronger, better conventionalism can be developed, carrying forth all of the virtues of the old versions but none of their vices. The logical positivists were right: logical and mathematical truths are a byproduct of our linguistic conventions. They are, so to speak, shadows of the syntax of our language. At least, that is what I shall argue here.

In PART I: THE PHILOSOPHICAL BACKGROUND, I explain exactly what conventionalism is, contrasting my approach with previous approaches and explaining what linguistic conventions are. In PART II: REVITALIZING LOGICAL CONVENTIONALISM, I develop a theory of the meanings of the logical constants—unrestricted inferentialism—and show that this theory is a version of logical conventionalism in the sense explained in part I. I then defend and develop conventionalism through extended discussions of pluralism, epistemology, and the influential objections that led to the historical demise of conventionalism.

In PART III: REVITALIZING MATHEMATICAL CONVENTIONALISM, I apply unrestricted inferentialism to mathematics. But the move from inferentialism to mathematical conventionalism is not as straightforward as in the case of logic. Unlike logic proper, mathematics is chock-full of existential assumptions and involves a computationally rich notion of truth. To overcome these worries I develop conventionalist-friendly but independently plausible accounts of both metaontology and arithmetical determinacy.

The final part—PART IV: REFLECTIONS—steps back to address big picture worries and meta-worries about my conventionalist theory. These include worries about whether conventionalism is objectively correct, what it means to think that it is, and how I can rationally believe it objectively correct in the face of the anti-conventionalist consensus in contemporary philosophy.

I have endeavored to make the book clear and enjoyable to read. Though it develops a single overarching theory, many particular sections and even entire chapters can be pulled out and studied independently. Many aspects of my discussion will be of interest to metaphysicians, epistemologists, and philosophers of mind and language apart from any special interest in logic or mathematics. And even philosophers who reject my overall conventionalist theory might find satisfying the solutions I offer to specific problems, such as those of existence and determinacy. But while I hope that many will find something here to interest them, that is not enough.

I don't merely want to interest you.

I want to convince you.

Shadows of Syntax

PART I

THE PHILOSOPHICAL BACKGROUND

1

What Is Conventionalism?

In the mid-twentieth century, conventionalism was extremely popular among philosophers, mathematicians, and scientists. More recently the view has fallen out of favor. In fact, that's a drastic understatement: at present, conventionalism is a *maligned* view in the philosophy of logic and mathematics. This book is the first major defense of conventionalism in nearly fifty years.[1] It is perhaps not quite literally true to say that if I get hit by a bus logical and mathematical conventionalism will die with me, but it is close.[2]

For the most part I won't be concerned with philosophical history, but it is worth briefly sketching the history of conventionalism before developing my own view. Doing so will illustrate that historical conventionalists didn't have a single, unified theory in mind. In broad strokes it was clear enough what conventionalism amounted to—conventionalists accounted for logic and mathematics in terms of human linguistic conventions—but a detailed understanding of what this involves is needed before the view's merits and demerits can be assessed. I'll briefly recount the rising and falling fortunes of conventionalism before trying to get clear on exactly what conventionalism is and how it should be understood.

I Rise and Fall

Conventionalism was the favored theory of logic and mathematics, as well as necessity and the *a priori*, of the logical positivists and other scientifically minded mid-twentieth century philosophers. It's easy to find pithy statements of the conventionalist position; here's Hans Hahn writing in 1933:

> … logic does not by any means treat of the totality of things, it does not treat of objects at all but *only our way of speaking about objects*; logic is first generated by language. The certainty and universal validity, or better, the irrefutability of a proposition of logic derives just from the fact that it says nothing about objects of any kind. … The propositions of mathematics are of exactly the same kind as the propositions of logic: they are tautologous, they say nothing at all about the objects we want to talk about, but concern only the manner in which we want to speak of them.[3]

Here is A.J. Ayer in *Language, Truth, and Logic,* his popular distillation of logical empiricism, first published in 1936:

> The principles of logic and mathematics are true universally simply because we never allow them to be anything else. And the reason for this is that we cannot abandon them without contradicting ourselves, without sinning against the rules which govern the use of language, and so making our utterances self-stultifying. In other words, the truths of logic and mathematics are analytic propositions or tautologies.[4]

Here is Ernest Nagel writing in 1944:

> … although their subject matter is not the language of which they are parts, they occur in that language because of the habits of usage or the tacit or explicit rules which govern that language. For example, if the characterizations “true” and “false” are employed in their customary manner, no statement can properly (that is, without contravening that usage) be characterized as both true and false; and if the word “not” is so used in connection with acts of affirming and denying statements that a false statement is rejected as not true, the principle of noncontradiction is instituted as a necessary truth. … the laws which are regarded as necessary in a given language may be viewed as implicit definitions of the ways in which certain recurrent expressions are to be used or as consequences of other postulates for such usages.[5]

Here is Carl Hempel writing in 1945:

> … the propositions of the system of mathematics as here delimited are true by virtue of the definitions of the mathematical concepts involved, or that they make explicit certain characteristics with which we have endowed our mathematical concepts by definition. The propositions of mathematics have, therefore, the same unquestionable certainty which is typical of such propositions as "All bachelors are unmarried," but they also share the complete lack of empirical content which is associated with that certainty: The propositions of mathematics are devoid of all factual content; they convey no information whatever on any empirical subject matter.[6]

Here is Hans Reichenbach, in his own popular distillation of logical empiricism, *The Rise of Scientific Philosophy*, from 1951:

> Logic formulates rules of language—that is why logic is analytic and empty. … Logical necessity and emptiness go together and make up the analytic, or tautological nature of logic. All purely logical statements are tautologies … they say nothing and thus inform us just as much…[7]

And finally here is Rudolf Carnap, in a retrospective from the early 1960s:

> Logical statements are true under all conceivable circumstances; thus their truth is independent of the contingent facts of the world. On the other hand, it follows that these statements do not say anything about the world and thus have no factual content. … the truths in logic and mathematics are not in need of confirmation by observations, because they do not state anything about the world of facts, they hold for any possible combination of facts.[8]

Examples could be multiplied, but I will stop.[9] In addition to the Vienna Circle and those influenced by them, conventionalism was

(seemingly) endorsed by Wittgenstein and his many followers, at least during his middle period.[10]

These quotations partially illustrate a point that is implicit in philosophical folklore—statements of the conventionalist doctrine took a variety of different forms: "logical and mathematical truths are *true by convention*", "logical and mathematical truths are *true in virtue of meaning*", "logic and mathematics are *analytic*", "logical and mathematical truths are *tautologies*", "logic and mathematics are *true by definition*", "logical and mathematical claims are *instructions* for the use of language", "logical and mathematical axioms are *linguistic rules*", "logic and mathematics are *factually empty*", *et cetera*. These slogans aren't all equivalent or even compatible with each other—how can claims be true while also being rules or instructions? But there is at least a family resemblance among all of these formulations. They all make human linguistic conventions central to our philosophical account of logic and mathematics. In the mid-twentieth century it was very widely believed that conventionalism successfully accounted for the metaphysics and epistemology of logic and mathematics without needing to posit spooky facts or mysterious mental abilities, thus solving, eliminating, or bypassing most puzzles in the philosophy of logic and mathematics. But by century's end, conventionalism was all but a dead theory.

So what went wrong? One problem was that conventionalists very rarely went beyond the slogan phase in developing their views. Maybe this was because they thought the view was obviously right; or maybe it was a sociological byproduct of the conventionalist consensus; or maybe the philosophical tools for properly formulating and defending the view didn't yet exist. Whatever the reason, the lack of detailed development became a major problem once serious criticisms of both logical and mathematical conventionalism started to pop up. Shortly after mid-century, it was already clear that the prominent criticisms of conventionalism by W.V. Quine, Arthur Pap, and others far surpassed most conventionalist writings in philosophical sophistication.[11] This situation gradually eroded conventionalism's support. The last vestiges of support for the general conventionalist approach to necessity and the *a priori* were destroyed in the 1970s when Saul Kripke, David Kaplan, and others

came up with convincing examples of necessary *a posteriori* ("water is H_2O") and contingent *a priori* ("I am here") truths.[12] These examples seemed to show that, contrary to the standard conventionalist line, necessary and *a priori* truth come apart.

Armchair intellectual history is a dubious business if pursued with too much seriousness, but I think my rough, big-picture account of conventionalism's woes is substantially correct, even if painted with broad brushstrokes. Yet while points were certainly scored against conventionalism, the nearly universal rejection of the view and the stigma it carries in contemporary philosophy is an extreme overcorrection.[13] I'll argue for this claim by developing and defending a conventionalist theory of logic and mathematics.[14] Along the way, I will attempt to conclusively answer all of the serious objections that have ever been raised against conventionalism. By this I mean not just that I will write some broadly coherent blather in response, but that I will identify decisive failings of these criticisms without any ad hoc maneuvering or philosophical razzle dazzle. I announce this here, baldly and boldly, but the proof, as they say, is in the pudding. The pudding starts with a clear understanding of what conventionalism *is*.

II Three Versions of Conventionalism

The quotes in the previous section demonstrate that conventionalist slogans came in a variety of forms. But careful attempts to state the central doctrine are extremely rare in the conventionalist corpus. Here I'll consider three different ways that conventionalism might be understood: two implausible, and one that is both plausible and, I will go on to argue, true.

(*i*) DESCRIPTIVE CONVENTIONALISM. The first implausible version of conventionalism says that logical claims somehow state facts *about* our use of logical words like "not", "and", and "if", and that mathematical claims somehow state facts about our use of mathematical words like "plus" and "zero" and "set". Ayer's *Language, Truth, and Logic* gives a slogan statement of

conventionalism that suggests this reading when he says of logical and mathematical truths:

> They simply record our determination to use words in a certain fashion.[15]

One reading of this, certainly not the only one possible, sees Ayer as claiming that logical and mathematical truths are equivalent to descriptive truths about how we use certain words. On this view, the logical truth "either it is raining or it is not raining" is equivalent to something like "we use the words 'or' and 'not' in such a way that the sentence 'either it is raining or it is not raining' is always true". But this is silly, since the former sentence expresses an *a priori* necessary truth, while the latter expresses an *a posteriori* contingent truth. A necessary truth can be "equivalent" to a contingent truth only in the extremely weak sense of sharing a truth value. So there is no way that these two sentences are equivalent in any deep sense, much less share the same meaning.

Ayer himself disavowed the descriptive reading in the introduction to the second edition of his book:

> It has, indeed, been suggested that my treatment of a priori propositions makes them into a sub-class of empirical propositions. For I sometimes seem to imply that they describe the way in which certain symbols are used, and it is undoubtedly an empirical fact that people use symbols in the ways that they do. This is not, however, the position that I wish to hold…[16]

I think that we can all agree that the relationship between logical/mathematical claims and linguistic conventions is not one of *description*; no credible version of conventionalism can take logic or mathematics to directly describe linguistic rules or regularities. This brings us to a second and perhaps more plausible (but still implausible) version of conventionalism.

(*ii*) NON-COGNITIVIST CONVENTIONALISM. The second implausible version of conventionalism simply identifies logical/mathematical claims with linguistic rules. Once again, Ayer provides a nice

statement, this time in his contribution to a symposium held shortly after publication of the first edition of his book:

> I think that our view must be that what are called a priori propositions do not describe how words are actually used but merely prescribe how words are to be used. They make no statement whose truth can be accepted or denied. They merely lay down a rule which can be followed or disobeyed.[17]

On one reading of this, logical/mathematical claims simply *are* linguistic rules. But this is strange: rules aren't the kind of thing that can be true or false, while logical/mathematical claims and, in particular, logical/mathematical *truths* most certainly are. In other words, non-cognitivist conventionalism faces the famous Frege-Geach problem of harmonizing a non-assertoric account of a branch of discourse with the apparent truth-aptness of the sentences in said branch.[18] This problem recurs even if Ayer's claim isn't that logical/mathematical claims are themselves linguistic rules, but rather that they are prescriptions that some particular linguistic rule be adopted.

One response to this problem is to deny that logical/mathematical claims are truth-apt, appearances be damned. I'm not sure if this would have been attractive to Ayer, but others have found it appealing. Yemina Ben-Menahem's book on the history of conventionalism argues, by drawing on logical conventionalism's roots in the geometric conventionalism of Henri Poincare and David Hilbert, that historical conventionalists shouldn't be seen as embracing *truth* by convention.[19] Whatever the status of this claim as history, I think it is bad philosophy: instances of the law of excluded middle and simple arithmetical claims are true, whatever else we may want to say about them. Of course, there are moves and strategies that the non-cognitivist could appeal to in an effort to save the appearances.[20] I'm not optimistic about the success of such moves and strategies, but rather than diving into them here, I will let appearances stand. The relationship between logical/mathematical claims and linguistic conventions is not identity.

(*iii*) THE EXPLANATORY READING. Conventionalism clearly posits a tight relationship between logical/mathematical claims and linguistic conventions. But I have rejected versions of the view where this relationship is a descriptive one; I have also rejected versions of the view where the relationship is identity. I think the correct relationship is *explanatory*: the conventions of a language fully explain the truth and necessity of the logical and mathematical claims in that language. Though conventionalists have struggled to formulate this clearly—some of the quotes in section I illustrate this struggle in action—I believe that historical conventionalists were questing for this formulation.[21] From our present perspective it is the natural way to interpret historical talk of truth "by" convention or "in virtue of" meaning.[22]

But my interests here aren't exegetical. Whatever its historical pedigree, the explanatory reading of conventionalism is the most philosophically interesting conventionalist thesis and is the one that I'll be endorsing and arguing for here. For the case of logic, it goes as follows:

> **Logical Conventionalism**: Facts about logical truth, logical falsity, logical necessity, and logical validity in any language are fully explained by the linguistic conventions of that language.[23]

For mathematics the statement is similar:

> **Mathematical Conventionalism**: Facts about mathematical truth and falsity in any language are fully explained by the linguistic conventions of that language.

I say that these truths are "fully" explained by linguistic conventions because everyone must admit that the truth of any sentence is at least *partially* explained by the conventions of the language in which it is stated.[24] The truth of a paradigmatic empirical sentence like "snow is white" is partially explained both by the fact that snow is the color white and also by the fact that according to our linguistic conventions, the sentence "snow is white" expresses that fact. But its

truth is *fully explained* by neither of these two facts separately, only by both of them together.

What is interesting and distinctive about the conventionalist thesis is that it sees logical truths and validities in a language *L* as *wholly and fully explained* by the linguistic conventions of *L*, likewise for mathematical truths.[25] Unlike "snow is white", for logical/mathematical claims there is absolutely no explanatory contribution from the world. In a sense (see 7.II), the worldly facts play no role at all in explaining why "either it is raining or it is not raining" is true in our language. As section I's parade of quotations illustrated, the disentanglement of logic and mathematics from the empirical world was a constant and important theme stressed by historical conventionalists and it is one that I will uphold. The relationship of my view to other conventionalist slogans will be discussed in part IV (in 13.I), after I present and defend conventionalist theories of both logic (part II) and mathematics (part III).

Something that has been implicit above is worth making fully explicit here. As I understand it, logical conventionalism is a thesis about logical truth and validity in natural languages. And mathematical conventionalism is likewise a thesis about mathematical truth in our mother tongue. This is how most historical conventionalists seemed to think of things, as the quotations illustrate. This point would hardly be worth mentioning but for the fact that Carnap, the most famous and influential conventionalist, usually presented his philosophical views using artificial languages. Because of this some have thought that Carnap's supposed conventionalism extended only to formally constructed languages.[26] I don't think this is the correct interpretation of Carnap (but the matter isn't fully clear).[27] In any case, this move threatens to drain conventionalism of its philosophical interest—who cares if logical truth in some carefully constructed formal language is conventional if logical truth in natural language has some other source? Like most if not all historical conventionalists, my conventionalism is a philosophical theory about logic and mathematics *in the wild*.

Before trying to construct a conventionalist theory, let me say a bit more, at a general level, about how the "explanatory" conventionalist

thesis should itself be understood.

III Explanation

Consider the inference rules modus ponens and affirming the consequent:

$$(MP)\quad \frac{\phi \quad \text{if } \phi \text{ then } \psi}{\psi} \qquad (ATC)\quad \frac{\psi \quad \text{if } \phi \text{ then } \psi}{\phi}$$

Where "ϕ" and "ψ" are schematic letters for which sentences in the language may be uniformly substituted (much more on schematic letters to come, starting in 2.IV). In our language modus ponens is valid (necessarily truth-preserving, see 3.I) but affirming the consequent is not.[28] Similarly, assuming classical logic, all instances of the law of excluded middle are true:

$$(LEM)\quad \text{either } \phi \text{ or not } \phi$$

These facts about validity and truth call out for an explanation—they aren't the kind of thing that can appropriately be taken as brute and unexplained. Logical conventionalism supplies an explanation of these facts that makes substantive appeal only to the linguistic conventions of our language. And mathematical conventionalism explains facts about mathematical truth and falsity in our language along the same lines. These explanations will be provided in parts II and III, respectively, but here it's worth establishing that conventionalist explanations, as such, don't require any special or nonstandard assumptions about the nature of *explanation* itself.

In ordinary talk, "explanation" sometimes refers to particular acts of communication, and at other times to something about the world.[29] Philosophers usually focus on the worldly or ontological sense of "explanation", but not always.[30] Sometimes it is not entirely clear which sense a given philosopher has in mind, but most of this

ambiguity is harmless. Here I'll be primarily focused on the worldly sense, but some of my remarks will bear on the communicative sense as well. The conventionalist thesis is not a merely surface-level claim about explanations considered as particular, presumably highly context-sensitive, acts of communication.

The general structure of explanation is basically as follows: there is some fact we want explained, which I'll call "the fact to be explained" (often called the *explanandum* in the literature) and this fact is somehow guaranteed by/illuminated by/accounted for by some other facts, which I'll call "the explaining facts" or "the explanation" (often called the *explanans* in the literature). Of course, all of this "fact"-talk shouldn't be taken too seriously. We could talk of truths instead of facts, though it's less natural in this context.[31] We could probably even talk in terms of *things* explaining other *things*, or *this* explaining *that*, but I don't think anything important hangs on this terminological choice. To talk naturally about explanations we need some way to both generalize and talk about particulars; "fact"-talk is quite well suited for this. The facts that conventionalists aim to explain include the validity of (*MP*), the invalidity of (*ATC*), and the truth, logical truth, and logical necessity of (*LEM*) in a given language. The particular facts that they use in their explanations concern only the linguistic conventions of the language in question.

Moving on from mere structure, the point of giving an explanation is, in essence, to resolve the mysteries and puzzles concerning the fact to be explained. Often this is done by showing where and how the fact to be explained fits into our total picture of the world.[32] An explanation of chemistry (the chemical facts), for example, integrates chemistry into our total picture of the world by accounting for chemical facts using our best overall theories of the physical world.[33] Logical/mathematical conventionalism aims to do the same thing for logic and mathematics by fully integrating logic and mathematics into a standard, scientific picture of the world.

It is often assumed that explanation is transitive.[34] This means that there are likely many different ways to fully explain high-level facts. If the psychological facts are fully explained in terms of the biological facts, and the biological facts in terms of the chemical facts, which are explained in terms of the physical facts, then the psychological

facts too can be explained in terms of the physical facts (or the chemical facts). There are many complete explanations, because the middlemen can be cut out of the picture. Still, if our conventions fully explain logical and mathematical facts, then it is often assumed that any other complete explanation of these facts must either be a "middleman", or facts that themselves explain the explaining facts about linguistic conventions. In some situations or contexts, this assumption can be relaxed or denied entirely (see 1.IV and 7.II).[35]

Some may find it problematic that putative theoretical explanations of this kind aren't *causal* in a standard sense, but I don't think this is a pressing concern. It is certainly true that when asked to think of an explanation, the first thing that comes to mind is typically an event that requires a causal explanation. But on reflection it seems clear that not all explanations are straightforwardly causal.[36] We can and do explain mathematical facts, and mathematical facts are causally inert and causally inefficacious, so mathematical explanation, at least, cannot be causal.[37] I don't think this can reasonably be rejected—surely Galois theory explains why the quintic can't be solved by radicals, and the transcendence of π partially explains why the circle cannot be squared, as the Greeks hoped, and so on and so forth. Beyond mathematics, even some scientific explanations aren't obviously causal in any direct sense, such as Newton's derivation of Kepler's laws of planetary motion from more general laws of motion. Given all of these points, we shouldn't be troubled if it sounds odd to say that our linguistic conventions "caused" the logical and mathematical truths in our language to be true, since the same oddity would attach to many perfectly standard explanations in science and everyday life were they expressed in directly causal terms.[38]

These are general philosophical points about the structure and purpose of explanations, but what about the *logical form* of explanations? I don't intend this as a precise formal question. Some may think that developing a "logic" of explanation is interesting in its own right. I'm not so sure. But I am sure that, at this point, presenting a formal account wouldn't be *philosophically* illuminating. Formal accounts or explications of informal notions often end up imposing precision in places where the natural language notion is crucially

imprecise. I am not interested in legislating away any imprecision in our notion of an explanation, but rather in understanding the broadly logical features of actual explanatory contexts, with all of their naked imprecision.

A few of these broadly logical features are relatively uncontroversial. Nothing explains itself. So explanation is not mere logical entailment. More generally: if some sentences logically entail another sentence, it doesn't follow that the facts expressed by the former sentences explain the fact expressed by the latter. Nor is explanation mere necessitation. Explanatory contexts are hyperintensional—swapping out necessarily equivalent sentences or terms in an explanation claim can sometimes change the claim's truth value.[39] All of these points distance explanation from mere entailment or necessitation.

Despite this distance, it seems clear that linguistic conventions can only hope to explain the facts about logical truth and validity in a given language by taking for granted background bridge principles that link the facts about truth or validity in a language with the usage of sentences. Linguistic usage alone can explain very little without metasemantic bridge principles that link usage to meaning. I admit this, and will discuss and develop the relevant metasemantic bridge principles in detail starting in the first chapter of part II. However, I will also argue (in 4.II) that allowing for bridge principles of this kind neither undermines the conventionalist credentials of the offered explanations nor makes my theory appreciably less controversial.

Conventionalists don't need to make any controversial or ad hoc assumptions about explanation in order to defend their views. Conventionalism aims to explain logic and mathematics in a different way than, but in the same general manner as, competing philosophical theories of these subjects. The explanations won't be merely pragmatic, nor merely causal, but this is standard in philosophy and doesn't involve any tendentious assumptions about explanations. In particular, I think conventionalists can admit all of the above points about explanations while remaining fairly neutral on the exact role of laws and general principles in explanations.

Against this, some may worry that the just-made point about metasemantic bridge principles is at odds with with influential

philosophical theories of explanation, most notably the so-called *deductive-nomological* or *covering law model of explanation*.[40] In such accounts, laws are a crucial part of explanations, since to have an explanation we must be able to deduce the *explanandum* from laws and initial conditions in some suitable language. But I don't think we need to worry too much about this. It's plausible that in ordinary contexts explanation takes place against general background assumptions about the way the world works.[41] Which bits of information are to be included in this background will likely vary somewhat from context to context. Conventionalists can usually remain neutral about whether laws are part of the explanation itself or part of the background against which explanation takes place. And even if it is required that "full" explanations have a completely empty background, so that nothing at all is taken for granted, conventionalism isn't undermined or trivialized. It is almost universally agreed in contemporary philosophy that linguistic conventions, even together with law-like bridge principles, cannot fully explain logical or mathematical truth.

The context sensitivity and hyperintensionality of explanation can be maintained without thinking that explanation is completely subjective or dependent upon the whims of some audience to which an explanation is propounded. To think this risks confusing the worldly and communicative senses of "explanation". But even focusing on the communicative sense, context sensitivity doesn't lead to subjectivism. From the other direction, a large helping of objectivity is fully compatible with there being some context sensitivity in "explanation"-talk. When asking the types of questions about logic and mathematics that we do in the philosophy room, it is often a matter of objective fact whether or not a given explanation is correct or not.[42] Decidedly though, this does not amount to denying that there may be a great deal of indeterminacy in our explanatory practices.

Nor does it require thinking that explanation is a *metaphysically* robust notion.[43] This needs to be stated outright, since distinctively metaphysical explanatory relationships have been widely endorsed recently. These notions come under different names—grounding, metaphysical definition, metaphysical dependence, *et cetera*—what

is common to all of them is that the relation is meant to be an objective, factual, hyperintensional relationship that is somehow metaphysical in character.[44] Personally I find these notions of heavyweight metaphysics puzzling, and have elsewhere argued that they violate basic scientific naturalism and leave us with irresolvable epistemological mysteries.[45] Arguing for this here would be a distraction. Instead I simply state loudly that the perfectly reasonable everyday notion of explanation is not a heavyweight metaphysical notion of the kind envisioned in recent philosophy (a foot stomp can be added upon request). Of course, if adherents of an internalized, lightweight, grounding-esque notion want to use their approach to translate my explanatory claims, then I have no great objection. My concern is only that, whatever we do, we remain within the bounds of basic scientific naturalism. So it is worth saying a bit more about exactly what this involves.

IV Scientific Naturalism

Conventionalism was historically popular because it promised to integrate logic and mathematics into the standard scientific picture of reality. Talk of "the standard scientific picture of reality" is a distortion, but a helpful one. There is really no such thing as "the" standard scientific picture of reality, at least not in any detail. New scientific theories are offered all the time, and old theories are continuously refined. Keeping this in mind, say that integrating a subject into "the" scientific picture of reality requires explaining the objects and laws of the subject in terms of more fundamental principles, ultimately bottoming out in fundamental physics.

There is controversy about how such explanations should go—do they require some type of *reduction* of chemical laws to physical laws, and of biological laws to chemical laws, and so on and so forth?[46] We can leave issues of strict reductionism aside here. What I have in mind is more general and more vague. And it applies well beyond the special sciences: facts about tables and chairs and roller coasters and ferris wheels must ultimately be accounted for in terms of fundamental science. This isn't to say that we have such

explanations now, but they must exist. This, in a very imprecise, amorphous nutshell, is the basic thesis of scientific naturalism. With something like mathematics, this is best seen as applying to our mathematical practice itself. If a practice involves some kind of causal access to particular objects, then those objects must be fit into the scientific world-picture. If not, then not. Our practices themselves can't resist scientific explanation. But there may well be freestanding practice-internal "mathematical" explanations (for example), provided that the overall practice itself is accounted for in naturalistic terms.

We have few successful explanations of high-level facts in terms of bottom-level facts. Maybe we don't have any, at least not in detail, at least not yet. But in most cases it is easy enough to see how such explanations will go, at least if we squint. Puzzles arise not when we currently lack a full explanation or reduction, but *when it is hard to see how one could possibly exist*. In other words, a subject matter only poses a prima facie problem for naturalism if it is difficult to see how it could possibly be integrated it into the overall scientific worldview. This isn't very precise, but the worries themselves aren't very precise. Yet imprecise worries can still be pressing. They can also be ubiquitous.

To illustrate, many philosophers and scientists have argued that it is impossible to integrate consciousness into our overall scientific worldview. So since we can't plausibly reject the existence of consciousness, we must instead broaden our conception of reality as a whole.[47] And some have argued that even our more workaday, non-conscious cognitive abilities cannot be accounted for in naturalistic terms.[48]

Worries arise about some facts when either the facts themselves seem *metaphysically weird* or our knowledge of the facts seems *cognitively spooky*. It goes without saying that "metaphysical weirdness" and "cognitive spookiness" are not precise notions. But the consciousness facts seem metaphysically weird, or would if they weren't so familiar. In a different way, moral facts seem metaphysically weird.[49] And both logic and mathematics are paradigm cases of subjects that seem both metaphysically weird and cognitively spooky. The facts of logic and mathematics are utterly

unlike any other facts about the world, and our knowledge in these areas comes in atypical ways. For this reason, logic and math have —more than anything except consciousness—driven philosophers from Plato to Kant to the present day to posit *both* metaphysical extravagances *and* implausible mental abilities.

The challenge for naturalists is to explain logic and mathematics without succumbing to either of these tendencies. Doing so will show that logic and mathematics can, in principle, be integrated into our standard scientific worldview. It is difficult to say in advance which philosophical theories avoid the twin pitfalls of metaphysical extravagance and cognitive implausibility. But we know it when we see it. As a rough guide, we can think of putatively *naturalistic* explanations as satisfying two key constraints.[50]

The first constraint is broadly metaphysical; it bars appeal to platonic heaven or other extra-physical realms of fact:

> **Metaphysical Naturalism**: A naturalistically acceptable explanation of some facts must not appeal to an independent realm of non-causal facts.

Despite how it sounds, this constraint doesn't force us to reject abstract objects out of hand. As long as an appeal to abstract objects can itself be explained in acceptable terms, it is fine. Nominalism is not the only naturalist view. Many philosophers have thought that our uses of numbers, propositions, properties, and other abstract objects can ultimately be explained in a scientifically acceptable fashion.[51] This is relevant to practices far beyond mathematics, including fiction and storytelling. Practice-internal explanations can appeal to abstract objects provided the practice as a whole can be fit into our scientific worldview. We can hold on to this point even if we forswear cross-level "explanations" in these areas (this will be relevant in 7.II).

The metaphysical constraint only rules out extreme views that see truths about abstract objects as concerning some extra-physical realm of objective, independent facts. Positing any domain of facts that are causally isolated from the natural world violates the constraint. This is as true of the Platonic realm of numbers as it is

the magical realm of moral facts or the logical heaven of crystalline purity. Theorists in mathematics, morality, and logic are still sometimes tempted, usually by metaphors, into adopting something like these pictures.[52] But views that succumb to these temptations don't count as "scientific" or "naturalistic" by my lights. Nor would they by the lights of the empiricists or the positivists or the Quineans or any other scientifically minded philosophers.

The second constraint is broadly psychological; it bars attributing to humans mysterious, unexplained mental powers:

> **Cognitive Naturalism**: A naturalistically acceptable explanation of some facts must not attribute to humans non-causal cognitive powers.

This constraint codifies the scientific understanding of the human mind as a byproduct of the causal operation of the human brain. One immediate application of this is to rule out mysterious faculties of "pure reason" or "intuition" that are still sometimes posited to explain the epistemology of logic and mathematics.[53] More generally, the cognitive constraint demands that we explain human cognition in terms of causal processes, principally those issuing from the operations of the brain. How this goes in detail is something that we have, as of yet, been unable to figure out. But somehow, someway, cognition arises from the embodied brain.[54]

This sounds trivial, but it has an important corollary: views that attribute non-computational mental powers to humans are not naturalistically acceptable. The basic reasoning behind this entailment is neatly expressed by Vann McGee:

> Human beings are products of nature. They are finite systems whose behavioral responses to environmental stimuli are produced by the mechanical operation of natural forces. Thus, according to Church's Thesis, human behavior ought to be simulable by a Turing machine. This will hold even for idealized humans who never make mistakes and who are allowed unlimited time, patience, and memory.[55]

Church's thesis ties informal computability to formal analyses of computability, such as recursion and Turing computability.[56] The thesis can't be *formally* proved, since it partly concerns the informal notion of a computation, but no truly serious threat to the truth of the thesis has ever been described.[57] The bounds of the cognitive constraint are not completely precise, but at the very least the constraint rules out supposed "explanations" that attribute to humans the ability to enumerate a non-recursively enumerable sequence, or to decide a formally undecidable notion.

Both of these constraints used the notion of "causation", but no controversial philosophical theory of causation is being assumed. I mean instead for "causation" to cover any physical difference-making relation, not just some specially circumscribed theorist's notion. I think this matches our natural language concept of causation quite closely. Philosophers who officially reject a narrow notion of causation should still be able to find a way to paraphrase these constraints in their own favored vocabulary. So understood the constraints are—I think—quite general and uncontroversial.

Unfortunately, even so it is extremely difficult to come up with a theory of logic and mathematics that meets both of these constraints. The many who have defied them did so not willy-nilly, but because they felt they had no other option. A conventionalist explanation of logic and mathematics would satisfy both of these constraints, or so I will attempt to show. This is all I mean when I refer to the conventionalist account as "scientific" or "naturalistic".

I'll present my explanation of logical facts in part II and of mathematical facts in part III. For now the crucial point is only that the conventionalist explanations I will go on to offer, though rejected by contemporary philosophers, don't require implausible assumptions about the notion of an explanation itself, nor do they require any violations of scientific naturalism. No tricks are being played. What is being offered is an honest-to-goodness account of why some sentences in some languages are logically [mathematically] true, others logically [mathematically] false, and some rules logically [mathematically] valid. Of course, whether or not my offered "explanation" is the actual explanation remains to be seen.

THUS FAR I'VE been talking as if we know and understand exactly what linguistic conventions are. And at some points I've even implicitly identified them with linguistic rules without saying anything about what those rules are or how they are to be understood. But before building a conventionalist theory, we need to be absolutely clear on the nature of conventions and linguistic rules. Only with this groundwork laid can we begin to build conventionalist theories of logic and mathematics.

---

1 Gianonni (1971), defending logical conventionalism, is the last extended defense in the published literature—I discuss Gianonni's views in 13.III—chapter 13, section III.

2 Post-positivist defenses of *logical* conventionalism are in the aforementioned Giannoni (1971), in Hellman (1986), and, at least in name, in Syverson (2002). Sidelle (1989) and (2009) defends conventionalism about the necessary *a posteriori*. Einheuser laid some groundwork for a defense of conventionalism in her unpublished doctoral dissertation (2003) and in her (2006). And Glock (2003) and (2008) tentatively defends a Wittgensteinian conventionalist-style approach to necessity. Coffa (1991) also seems to endorse a conventionalistst style account, but since his main focus is historical, I am uncertain about his own views. As far as I'm aware, no unified version of logical and mathematical conventionalism has been defended at length since the heyday of logical positivism.

3 Hahn (1933), translated into English in chapter 7 of Ayer (1959).

4 Ayer (1946), page 77.

5 Nagel (1944), quoted from page 319 of Benacerraf and Putnam (1964).

6 Hempel (1945), quoted from page 378 of Benacerraf and Putnam (1964).

7 Reichenbach (1951), pages 222–223.

8 Carnap (1963*a*), pages 25 and 64.

9 See Britton (1947) for another early endorsement of conventionalism.

10 See Wittgenstein (1953) and (1974); Malcolm (1940) is an important early work on conventionalism by a philosopher in the Wittgensteinian tradition. It's also worth noting that Wittgenstein (1922) endorsed a conventionalist-adjacent theory that was a major formative influence on the positivists of the Vienna Circle. The end of Baker and Hacker (2009) discusses the many ways in which the positivists misunderstood the *Tractatus*, but in fairness to them, as Coffa (1991) discusses, the positivists were also influenced by personal contact with Wittgenstein as he was transitioning away from his early views.

See also Carnap's (1963*a*) discussion of the Circle's contact with Wittgenstein.

11 See Quine (1936), (1951*a*), and (1960*a*), and Pap (1958).

12 See Kripke (1980) and Kaplan (1989).

13 There have been a few prominent efforts to rehabilitate the philosophical reputation of logical positivism—see Coffa (1991) and Friedman (1999).

14 Sidelle (1989) develops a conventionalist account of necessary *a posteriori* truths. I am sympathetic, but my views on logic and mathematics are largely independent of these sympathies.

15 Ayer (1946), page 84. Ayer's conventionalism is applied to all necessary and *a priori* truths, not just those of logic and mathematics.

16 Ayer (1946), page 16. I have eliminated a footnote in the original when reproducing this passage.

17 Ayer (1936), page 20.

18 See Geach (1965) for the problem and the attribution to Frege; see Woods (forthcoming) for a recent general overview. Apparently, Waismann (1968) objects to rule-based approaches to necessity along these lines.

19 See Ben-Menahem (2006); see also Coffa (1986) and (1991).

20 An alternative strategy for non-cognitivist conventionalists is to mimic the quasi-realist strategy used to defend non-cognitivism in metaethics, growing out of Blackburn (1984*a*) and Gibbard (1990). This is not unprecedented, as Blackburn (1987) already applies something like quasi-realism to necessity.

21 This isn't true of every version of conventionalism. Reichenbach's (1920) "axioms of coordination" are conventions that function as a framework used to ensure that questions are empirically well defined. In this, Reichenbach's view is somewhat similar to Kant's (1787); see chapter 3 of Friedman (1999) for discussion. I think that this type of approach would probably count as a partial version of explanatory conventionalism, but it isn't fully clear.

22 This is apparent for Giannoni's (1971) discussion, where he uses epistemic terminology rather than explanatory terminology while at the same time decrying the epistemic terminology's inadequacy (see my discussion in 13.III). Hellman (1986) also passes from directly epistemic terminology to explanatory sounding talk of "determination" in his development of logical conventionalism (though Hellman's (1985) official cashing out of "determination" is in model-theoretic terms). Criticism of the first two versions of conventionalism, coming from someone with conventionalist sympathies, occurs early in Gasking (1940). See also Ayer's discussion in the introduction to his (1946) where he rejects both of his earlier formulations in favor of something like the explanatory reading. Glock (2003) thinks that an explanatory-style formulation, which he associates with Carnap, can be combined with a rule-style approach, which he associates with Wittgenstein. Ebbs (2001) and (2011) argue that Carnap never endorsed an explanatory conventionalist thesis and that Quine knew that Carnap didn't. Ebbs's central argument for this is based

on denying that Carnap saw contradictory rules as issuing in truth. I don't think that Carnap thought this, but if he did, he shouldn't have—see chapter 5. I'll discuss these issues again in 13.II and 14.I.

23 Some logics, like First-Degree Entailment (introduced in Anderson and Belnap (1975)), have no logical truths, but for any language that uses FDE, this too will be explained by the conventions of the language.

24 I will usually use sentences as truthbearers, but this is a matter of convenience, not a substantive philosophical stand against approaches that use propositions, utterances, or mental tokens instead—see 2.V, 3.I, and 6.II. Since my focus will be on contextinsensitive notions, I don't think that any confusion will arise based on this choice.

25 A full explanation need not be an explanation traced all the way back to whatever it is that is explanatorily fundamental— conventionalists are not committed to thinking linguistic conventions are themselves explanatorily brute!

26 See Ricketts (1994) and Ebbs (2001) and (2011).

27 See Carnap (1955) for a rare bit of attention to natural language. Hellman (1986) is a modern approach that seems to apply conventionalism only to formally regimented artificial languages. Sometimes Carnap is explicit that an approach being advocated is not applicable to natural languages, as in his (1952). Bar-Hillel (1954) argued that linguistics needs to adopt Carnap's "logical syntax" and Chomsky (1955) replies; in many ways though, this debate was more focused on the general role of logic and model theory in linguistics, rather than anything particular to Carnap's philosophy. See 13.II for discussion of Carnap's views in *Logical Syntax.*

28 The validity of modus ponens in English is very widely accepted but not entirely uncontroversial, see McGee (1985).

29 This distinction is made in many treatments of explanation—see Strevens (2008), pages 6–7, for one example.

30 Achinstein (1983) gives precedence to the communicative sense.

31 This entails that explanations, or at least full explanations, must be true. This can be denied—see Van Fraassen (1980)—but that a correct explanation is (at least approximately) true is independently plausible and very widely accepted. Of course, there must also be a fact there to be explained: if something isn't true, there is nothing to explain (though we could still explain the fac that there is nothing to be explained).

32 Friedman (1974) and Kitcher (1989) provide accounts of explanation and understanding of roughly this kind. My remarks here shouldn't be read as a full endorsement of a unification-based account of explanation.

33 Arguably this was done by Pauling (1939).

34 Sometimes the transitivity of explanation is denied, but even if so, we can usually work with the transitive closure of our non-transitive relation of explanation.

35 This is because there are likely other senses in which logic and math can be explained—see 7.II.

36 Non-causal explanations are rejected in Quine and Ullian (1978), chapter IX. However most philosophers, even those sympathetic to broadly causal accounts of explanation, such as Lipton (2004), admit that a purely causal account isn't the entire story of explanation.

37 Steiner (1978) is a classic early paper on mathematical explanation.

38 Though I'll take no direct stand on it here, it's possible that the conventionalist explanation I will offer is broadly causal in a much more mediated, far less direct sense.

39 Witness the recent philosophy of language textbook, Daly (2013), page 282: "explanatory contexts … are hyperintensional contexts."

40 These terms are sometimes distinguished when, for example, we are concerned with statistical laws, but I'll ignore this here. For a history of covering law accounts of explanation, see Salmon (1989); see Hempel and Oppenheim (1948) for the advent of this approach and Hempel (1965) for its most influential development.

41 This is related to a distinction that is sometimes made in discussions of causal explanation between "explainers" and "enablers"—see Bermudez (1995) and Woodward (2011), for example.

42 This differentiates what I'm saying here from the pragmatic theory of explanation offered in Van Fraassen (1980), despite some similarities.

43 My own (developing) view of hyperintensional relations, of which explanation is merely one among many, holds that there is a sense in which explanation claims are sensitive to the concepts involved in a statement of the fact to be explained. I won't attempt to develop this view here, but I think that it is entirely compatible with the anti-subjectivist view of explanations I've been espousing.

44 This type of approach was inaugurated by Fine (2001); variants have been developed and explored by Fine (2012), Rosen (2010), and Schaffer (2009) among others. Sometimes grounding is not taken to be an explanatory relation itself, but is instead seen as "backing" explanations, in roughly the way that causation backs causal explanations.

45 These arguments are aimed directly at the views of Lewis (1984) and Sider (2011), but I think the case generalizes—see my (2016*a*), (2016*b*), and (2016*c*). For direct criticisms of grounding and related notions, see Hofweber (2009) and Wilson (2014).

46 For classic readings here, see Oppenheim and Putnam (1958), Nagel (1961), and Fodor (1974). For more recent discussions see Block (1997) and Loewer (2009).

47 For a small sample of such views, see Nagel (1974), Levine (1983), McGinn (1989), and Chalmers (1995).

48 See Lucas (1961), Penrose (1989), (1994), and Boghossian (2014).

[49] Mackie (1977).

[50] The constraints in this section are more general versions of the constraints on acceptable metasemantic theories from Warren and Waxman (forthcoming *a*).

[51] This is somewhat related to Field's (1984*c*) distinction between "moderate" and "heavy-duty" Platonism.

[52] See also Rosen (1993) for a discussion of the picture thinking that motivates this move.

[53] In logic see Bonjour (1998), in mathematics see Gödel (1964/1983).

[54] Note that the constraint concerns only *cognition*, not consciousness.

[55] Quoted from McGee (1991), page 117.

[56] See Odifreddi (1989) or other textbooks in recursion theory for details.

[57] This doesn't mean that there aren't nonstandard approaches to computability that don't affirm the thesis. Some intuitionists and constructivists are committed to rejecting the thesis, though not all—see Vandoulakis (2015) for some relevant discussion.

2

Linguistic Conventions

What are linguistic conventions? Most historical conventionalists took linguistic conventions to be syntactic rules of language use. I think this is the correct approach; here I'll clarify it, defend it on general grounds, and extend it along several dimensions. I will also argue that conventions, so understood, are fully compatible with a naturalistic view of the world.

Some may be wondering why this chapter is even necessary. Haven't post-positivist philosophers successfully analyzed the notion of a *convention*, thereby clearing away all of the lingering philosophical mysteries? Can't contemporary conventionalists simply stand on the shoulders of these giants when giving an account of linguistic conventions? Not exactly. Whatever the merits of these analyses, for the purposes of logical and mathematical conventionalism, something more and something different needs to be said. I'll begin by explaining why.

I Post-Positivist Accounts of Conventions

In 1969, David Lewis published *Convention: A Philosophical Study*. In the book, Lewis was explicitly concerned with analyzing our linguistic conventions in a way that would answer the anti-conventionalist arguments of W.V. Quine. Interestingly, Lewis's book is based on a dissertation written under Quine's supervision.

Roughly put, Lewis's basic idea is that conventions are arbitrary solutions to social coordination problems. Arbitrary because social coordination problems are recurrent problems that have two or more coordinating equilibria, and it is effectively arbitrary which equilibrium strategy is chosen. A classic example concerns choosing whether to

drive on the left versus the right side of the road. Choosing one of these options over the other is arbitrary for society at large.

After various refinements and complications Lewis's final characterization of a convention is:

> A regularity *R* in the behavior of members of a population *P* when they are agents in a recurrent situation *S* is a *convention* if and only if it is true that, and it is common knowledge in *P* that, in almost any instance of *S* among members of *P*,
>
> (1) almost everyone conforms to *R*;
> (2) almost everyone expects almost everyone else to conform to *R*;
> (3) almost everyone has approximately the same preferences regarding all possible combinations of actions;
> (4) almost everyone prefers that any one more conform to *R*, on condition that almost everyone conform to *R*;
> (5) almost everyone would prefer that any one more conform to *R′*, on condition that almost everyone conform to *R′*,
>
> where *R′* is some possible regularity in the behavior of members of *P* in *S*, such that almost no one in almost any instance of *S* among members of *P* could conform both to *R′* and *R*.[1]

This is an analysis of conventions, in general. But Lewis's central goal was to account for conventions of language, in particular.

Accordingly, after analyzing conventions, he introduces the idea of a signaling game and discusses the use and role of conventions of truthfulness in a natural language. A great deal of later philosophical work is based on the signaling games that Lewis introduced.

Brian Skyrms and a number of others have shown that it is possible for rudimentary signaling systems to evolve via simple learning rules and other evolutionarily plausible processes.[2] Sometimes Skyrms suggests that these results show how meaning evolved and even lend support to a conventionalist-friendly view of logic.[3] But while it is important to have models of how basic reasoning processes could have evolved, Skyrms's results by

themselves have virtually no bearing upon philosophical debates about the nature of logic.[4]

After introducing signaling games, Lewis spends some time sketching a formal treatment of languages. Ultimately, he defines the old positivist notion of "analyticity"—or truth in virtue of meaning—as truth in all possible worlds.[5] These were pre-*Naming and Necessity* days, so Lewis wasn't yet distinguishing between the analytic and the necessarily true. Largely for this reason, as interesting as all of this material is, it doesn't have much to do with conventions of language in the sense that concerned Carnap and Quine.

Quine himself made this point clear in his foreword to Lewis's book:

> The problem of distinguishing between analytic and synthetic truths was apparently one motive of the study. In the end, Lewis concludes that the notion of convention is not the crux of this distinction. He does not for this reason find the analyticity notion unacceptable, however. He ends up rather where some began, resting the notion of analyticity on the notion of possible worlds.[6]

This is not a fault with Lewis's analysis of social conventions. Any plausible alternative approach to general social conventions will likewise fail to capture the particular notion of a linguistic convention that Quine argued against. This is because Quine and his positivist opponents were concerned with the linguistic regularities (or rules) that will sit at the heart of a general definition of a social convention, in Lewis's sense.

That these linguistic regularities in natural languages are "conventional" in something like Lewis's sense might account for their social status in a community.[7] But this point has nothing much to do with whether the regularities themselves can do useful and interesting philosophical work. What is needed for logical and mathematical conventionalism is an account of the conventional linguistic regularities that become socially conventional in a linguistic community.

II Conventions as Stipulations?

One folkloric approach sees the conventions of language as explicit *stipulations*. This view is enticing, in part, because abbreviations and other newly minted linguistic conventions are often introduced as explicit stipulations via proclamation. Along these lines, imagine that I stipulate that "it is raining or it is not raining" is, henceforth, to express a necessary truth.

The first thing to point out is that stipulations like this take place *within language.* So the very ability to stipulate requires operating within a previously understood language. The same also goes for understanding such stipulations. This alone shows that such stipulations cannot be the fundamental starting point for language, as they require a preexisting language for their very formulation. Such stipulations are occasionally made, but they weren't made for *every* meaningful linguistic item. This point is widely recognized and agreed upon. Here is Bertrand Russell, writing nearly a century ago:

> We can hardly suppose a parliament of hitherto speechless elders meeting together and agreeing to call a cow a cow and a wolf a wolf. The association of words with their meanings must have grown up by some natural process, though at present the nature of the process is unknown.[8]

Natural language does not consist of a series of explicit stipulations. Perhaps some conventionalist writings have fostered an over-reliance on stipulation by focusing on artificial languages, rather than natural languages.[9] For completely artificial languages we are free to stipulate any facts we like about them, since they are not being used by anyone. These stipulations take place—where else?—in natural language. But my concern here is with the linguistic conventions of actual languages. Not with toy cases.

Even when stipulations succeed in the wild, they don't seem able to secure the kind of determinacy needed. If I stipulate that "it is raining or it is not" is to express a truth, it is natural to ask: which truth? Similarly for necessity. One response might be to then stipulate a synonymy claim. For instance, I could stipulate that "it is

raining or it is not" is to express the same content as "2 + 2 = 4". This might give content to a simple logical truth by stipulation, but only by piggybacking upon a priorly understood mathematical truth. In any case, as Quine famously argued, this kind of explicit stipulation picture cannot possibly serve the logical conventionalist's needs (these arguments will be discussed in detail in 7.III).

All of these issues point to a deeper worry: explicit stipulations are *metasemantically idle*. If I stipulate or declare in a loud voice, with the utmost gravitas, that I will now be speaking French, but then continue to speak in my normal way, my stipulation will have failed. It would be mad to interpret me as speaking French incomprehensibly rather than speaking English passably.[10] Similarly, if I stipulate that "dog" is to now mean what "cat" previously meant, and vice versa, but I continue to call dogs "dogs" and cats "cats", and to otherwise behave as I always did, I will have failed to change the meanings of the relevant terms. Stipulation alone has no power to found—or even alter—meanings. This is illustrated by these examples and supported by our intuitions. Meaning is determined by language use, not direct stipulation. A pattern of usage always trumps a stipulation. Of course, this is not to deny that a verbal stipulation can be causally relevant in changing the usage of some speech community. But even in such cases it is the underlying pattern of use, and not the stipulation, that is doing the work. Perhaps removing the stipulation would also remove the pattern, but *if* the pattern of use remained without the stipulation, meaning would be unaffected.

Once a language is up and running, stipulation can play a role. But that role is mainly to communicate an intended pattern of use, or an intention to coin a new use. All of this is compatible with the metasemantic impotence of explicit linguistic stipulations. Usage founds meaning, if anything does.[11] Not stipulations. For this reason above all the central linguistic conventions are not explicit announcements, stipulations, or declarations about linguistic items.

Instead, as early twentieth-century conventionalists already realized, the "conventions" that are potentially relevant in the philosophy of logic are the linguistic rules that govern the use of logical expressions, and similarly for mathematics. Early conventionalists didn't say very much about these rules, and they

never argued for their existence. They didn't seem to feel the need to. But a fully satisfying conventionalist theory must do both of these things and, additionally, must show that linguistic conventions, so understood, are scientifically respectable posits.

III The Case for Inference Rules

Our linguistic conventions are linguistic rules that govern, in various different ways, our usage of sentences, words, clauses and—in spoken language— language-relevant sounds. This is how historical conventionalists seemed to understand linguistic conventions and this understanding dovetails with the importance of rules in contemporary linguistics. But the types of rules that are relevant in logic and mathematics are inference rules, and inference rules aren't usually posited by contemporary linguists.

Yet clearly, linguists should posit inference rules. Our practice of performing inferences is not entirely haphazard; instead, it is rule-governed. This isn't necessarily to say that *every* inference we make involves the application of some inference rule or other.[12] It is just to say that we, as a matter of fact, follow inference rules. Consider again an informal version of the inference rule modus ponens (*MP*):

$$(MP)\quad \frac{\phi \quad \text{if } \phi \text{ then } \psi}{\psi}$$

It is widely agreed that English speakers follow (*MP*) in reasoning, arguing, and going about their day-to-day lives.[13] In the second half of this chapter I'll outline a naturalistic account of inference rule-following. But first, leaving the exact nature of rule-following aside, I want to make a case for positing syntactic inference rules like (*MP*) as part of our overall linguistic competence.

Though inference rules aren't much discussed by linguists, syntactic rules of other sorts are the central posit of contemporary linguistics. There is very wide agreement among linguists and cognitive scientists that we language users follow syntactic rules of

both morphology (rules for forming complex words) and grammar (rules for forming words into phrases and sentences).[14] Some have also argued that we follow rules of phonology (rules for pronouncing words), but this isn't as widely accepted.[15] While controversy surrounds the exact nature and implementation of these rules, the scientific case for positing rules of morphology and grammar, in some fashion or other, is nearly conclusive.

One of the keystone experiments in the case for linguistic rules was the famous wug-test, administered in 1958 by psychologist Jean Berko Gleasen. Gleasen showed four-to-seven-year-old children an illustration of an indeterminate bird-like creature accompanied with the following text:

This is a wug.
Now there is another one.
There are two of them.
There are two _____

The children were then asked to fill in the blank. About 75% of the pre-schoolers and 99% of the first-graders replied with "wugs".[16] Since "wug" is just a made-up term, in forming its plural the children couldn't possibly have been appealing to a plural form that they had previously heard from adults. Instead they were following a syntactic rule for forming the plural form of a regular noun: for regular noun *α*, the plural form is *α*⁀"*s*".[17]

The reasons for positing morphological rules carry over to inference rules like modus ponens. I have already made this point in print, but let me review the case here.[18] Consider the following nonsense argument, based on Berko's wug test:

1. Wugs kill philosophers
2. If wugs kill philosophers then lugs kill philosophers
3. *So*: Lugs kill philosophers (1,2)

Nearly all English speakers, including relatively young children, would accept this reasoning. Moreover, if we presented English speakers with (1) and (2) and asked them what, if anything, they

could “infer” or “conclude” from these two claims, a large majority of them would come up with something very much like or even identical to (3). But since the sentences “wugs kill philosophers” and “lugs kill philosophers” are nonsense sentences, language users who perform this bit of modus ponens reasoning cannot be performing any semantic operation on meanings or contents or propositions.

In fact, in answering in this way, they aren’t even literally performing an inference. They are instead applying a syntactic rule to a degenerate case. This is exactly what children were doing in the original wug test. Human language is messy and complicated, and there need be no foolproof syntactic criterion of degeneracy. Recognizing this doesn’t undermine a syntactic approach to linguistic rules. It just means recognizing that, as with grammatical rules, our inference rules can overshoot. This is another thing the inferential wug-test illustrates.

One might doubt that containing meaningless terms like “lugs” and “wugs” is sufficient to make an entire sentence meaningless—maybe it just makes the sentence false? One may even doubt that “lugs” and “wugs” are themselves truly meaningless—maybe they have highly indeterminate meanings? I am not convinced, but even if it’s right none of this undermines the central point. Speakers apply modus ponens to novel sentences. They do so based on the syntactic form of the sentences, not because of familiarity with or a habit of operating with the sentence.[19] I also think that the modus ponens intuition will continue to hold for clearer cases of nonsense strings, though perhaps less strongly.

What happens in the modus ponens case is directly analogous to what happens in the original wug-test: English speakers extend patterns of linguistic behavior to novel cases in a uniform way. In other words, they follow linguistic rules that extend to and cover such cases. Just as English speakers extend their patterns of sentence formation to meaningless cases, as in Chomsky’s famous and shopworn example of a meaningless but grammatical sentence, “colorless green ideas sleep furiously”, so too do they extend their patterns of sentence transformation to meaningless cases.[20] The extended wug-test is just one example. But it, along with similar data, provides very strong evidence that language users follow

syntactic rules of inference just as surely as they follow syntactic rules of morphology and grammar.

On this basis, in order to account for human linguistic *performance*, we should posit wholly syntactic rules of inference as part of human linguistic *competence*.[21] Some psychologists have implicitly denied this in discussions of reasoning, typically citing the dispiritingly poor performance of the average human on abstract reasoning tests, such as Wason's famous selection task.[22] It is of course true that humans are generally very bad at abstract logical reasoning.[23] But the competence/performance distinction was designed to handle precisely this kind of situation. Even poor reasoners correct their behavior in the right manner, if given extra information (and world enough, and time). An observation by Johan van Bentham makes this clear:

> A psychologist, not very well disposed toward logic, once confessed to me that despite all problems in short-term inferences like the Wason Card Task, there was also the undeniable fact that he had never met an experimental subject who did not understand the logical solution when it was explained to him, and then agreed that it was correct.[24]

To say that speakers follow inference rules is not to say that they always do so perfectly, immediately, or without difficulty. Nor is it to say that there can't be various factors that make it easier or harder for speakers to act in accordance with the rules that they follow. Occasionally task-specific mental modules are proposed as a psychological alternative to inference rules, but for the reasons just given, I think these are better thought of as aids to inferential performance in particular contexts.[25]

In claiming that speakers follow syntactic inference rules I don't mean to be endorsing any of the controversial psychological claims that some linguists have made for grammatical rules. For example, that these rules are innate, universal to all possible human languages, or the output of an adaptive psychological mechanism. Perhaps most importantly, I am not assuming that linguistic rules are explicitly represented in the brains of language users (exactly what is

required for inference rule-following will be discussed in section VIII). Whatever the exact nature of their character and psychological implementation, inference rules govern aspects of our usage, broadly construed, of sounds and marks. In other words, they are, in a standard sense, both *conventional* and *syntactic*.

This is true of all of our linguistic conventions, I believe. But standard inference rules don't exhaust our linguistic conventions. While rules like (*MP*) are familiar from the logic classroom, it's extremely plausible that in natural languages we also follow other kinds of inference rules. This will be important in parts II and III and is worth explaining more carefully.

IV Beyond Standard Inference Rules

Inference rules are often formally modeled using a natural deduction system.[26] In standard natural deduction systems for formal languages (with the conditional written as "→"), modus ponens can be written as:

$$(MP): \quad \frac{\phi \quad \phi \rightarrow \psi}{\psi}$$

(*MP*) sanctions introducing ψ into any proof that already contains ϕ and $\ulcorner \phi \rightarrow \psi \urcorner$. Formally, this makes modus ponens a rule that governs the construction of proofs. But the formal methods involving proof construction model our actual processes of inference and argument construction.

In terms of inference, the best way to understand modus ponens is as a rule of acceptance, broadly construed (the next section will explain this "broad" notion of acceptance). So understood, the rule gives rise to norms of inference, both positively and negatively. Positively, (*MP*) provides justification for accepting the conclusion upon acceptance of all of the premises (though perhaps not all things considered justification). Negatively, (*MP*) bars rejection of the conclusion upon consideration and sustained acceptance of all of the

premises. Thus (*MP*) "allows" certain inferential transitions, and bars certain patterns of attitudes. This is a simplified picture of the normativity of logic; a fuller picture is given in 4.VII (see also chapter 6's discussion of epistemology).

Formal proofs model these rule-sanctioned chains of natural language inferential transitions. In addition to rules of this kind, standard natural deduction systems also include rules that formally simulate the natural language practice of reasoning under and discharging assumptions. A central example is the rule of *conditional proof* :

$$(CP)\quad \begin{array}{c} (\phi)^n \\ \vdots \\ \psi \end{array} \Big/ \; \phi \rightarrow \psi \; n$$

Assumptions occur within parentheses "(...)"; the assumption with superscript "*n*" is discharged at the line indexed by "*n*". In terms of inference, these are also rules of acceptance, broadly construed, but they differ a bit from rules like (*MP*). (*CP*) gives rise to norms in a slightly more complicated fashion. If ψ can be accepted on the basis of ϕ, then it is permissible to accept $\ulcorner \phi \rightarrow \psi \urcorner$ full-stop. By including both standard rules like (*MP*) as well as meta-rules like (*CP*), natural deduction proof systems allow for the perspicuous simulation of a significant chunk of human reasoning. Even still, there are aspects of our rule-governed inferential practices that go beyond what can be represented in such systems. Two of these aspects will be of crucial importance in parts II and III, so let me briefly introduce them here.

(*i*) REJECTION. We must allow for rules that sanction rejection as well as acceptance, since our natural language inferential practices include inferences leading to and from the rejection of various claims. These are rule-governed inferences just as surely as anything is. Rules of this kind will directly involve the notion of rejection. A rejection rule might, for example, tell us that we can reject some sentence whenever we accept some other sentences.

Or it might lead from rejection to rejection, or from rejection to acceptance. Once the need for such rules is pointed out it seems obvious, but historically rejection rules have been neglected in formal treatments of inference. The reason for this historical neglect is simple: in the formative years of mathematical logic, Frege influentially analyzed rejecting ϕ as accepting its negation, $\ulcorner \neg\phi \urcorner$, and most subsequent logicians and philosophers followed him in this.[27]

More recently Frege's analysis has rightly fallen into widespread disrepute.[28] This shouldn't be that surprising. *Accepting a sentence* involves adding something to your beliefs; *rejecting a sentence* involves excluding something from your beliefs. It would be bizarre to think that rejecting ϕ is a matter of accepting some *other* sentence, even ϕ's negation. No matter what other sentence you may add to your beliefs, if you don't also exclude ϕ from your beliefs, then you have not rejected ϕ. This is only to say that rejecting a claim and accepting its negation are *conceptually* distinct. It remains an open possibility that whenever someone rejects a claim they also accept its negation. I don't think this is the case in general, but simply treating rejection as primitive (with respect to acceptance) doesn't settle or bias the matter.

I happen to think that rejecting a sentence and accepting its negation can come apart in reality, too. It is easy enough to imagine simple language communities that reject claims without even having a negation symbol in their language. And several non-classical logics can only be understood if rejecting is a different mental state than accepting a negation.[29] We must include rejection rules in our natural deduction systems if we want to perspicuously model the full range of natural language inferences, in our language and beyond.

It is a simple matter to expand our standard logical formalism to handle rejection. We can do so by using the force indicators "+" and "-": the plus is an acceptance indicator, while the minus is a rejection indicator.[30] By using *signed sentences*—sentences prefixed with a force indicator—we can formulate natural deduction rules for both acceptance and rejection, as well as mixed rules, leading from acceptance to rejection or vice versa. It is important to stress that these indicators are not operators or connectives and as such they

do not embed in complex sentences. So while we can consider ⌜ $+\phi$ ⌝ in our formalism, ⌜ $\phi \rightarrow +\psi$ ⌝ is ill-formed nonsense. Proofs in this kind of system can be used to model a wider range of our inferential practices than can be handled by standard systems. A system of this kind is called a *bilateralist* system.

Let me briefly illustrate some of the philosophical power of bilateralist systems. A very natural form of the law of non-contradiction tells us to reject *any* contradiction. If the standard Fregean account of rejection as acceptance of negation is used, this version of non-contradiction would be formulated in rule-form as follows:

$$(LNC\ unsigned)\quad \frac{}{\neg(\phi \wedge \neg\phi)}$$

In effect, this rule allows us to accept the negation of any contradiction at any point in time. But this fails to capture the intuitive force of the injunction to reject all contradictions. Dialetheists think that some contradictions are true and so, intuitively, they reject the law of non-contradiction. But in the logic favored by Graham Priest, the most influential dialetheist, all sentences of the form ⌜ $\neg(\phi \wedge \neg\phi)$ ⌝ are logical truths.[31] In other words, dialetheists like Priest seem to follow the unsigned non-contradiction rule despite intuitively rejecting the law of non-contradiction. Accordingly, a better formulation of non-contradiction as understood in natural language uses force indicators:

$$(LNC\ signed)\quad \frac{}{-(\phi \wedge \neg\phi)}$$

This rule sanctions rejection of any contradiction at any point in time. Unlike the unsigned rule, this non-contradiction rule is not followed by dialetheists like Priest. In accepting some contradictions, dialetheists are not also permitted to reject all contradictions.[32] By allowing for both acceptance and rejection we are able to formulate a version of the law of non-contradiction that nicely captures what is at issue between dialetheists and non-dialetheists. This is just one

small illustration of the philosophical advantages gained by working in a bilateralist system. This generalization will be philosophically crucial in part II, where it will be discussed further.

(*ii*) OPEN-ENDEDNESS. As the wug-test illustrates, inference rules like (*MP*) apply not just in our current language, but also continue to apply as our language changes via the addition of new vocabulary. Say that a schematic rule of inference *R* in a formal language *L* is *language closed* just in case all appropriate substitution instances of *R*'s schematic letters are sentences of *L*. This is the way that schematic principles and rules are standardly understood in logic textbooks. But as a model of natural language reasoning, language-closed rules are inadequate. When the word "google" was added to English, we didn't have to reflect upon whether or not we could use modus ponens reasoning on sentences containing "google"! If someone did react in that way, we would question whether they even understood modus ponens.

Some of the inference rules we follow, including both logical and mathematical rules, hold not just in the current timeslice of our language, but in any coherent alteration of our language as well. That is to say that these rules are *open-ended*. Over the last twenty years this has been very widely accepted by logicians and philosophers. What is more: several philosophers have argued that open-endedness is useful in explaining mathematical determinacy and in answering Skolem's paradox.[33] I will be making important uses of open-endedness in part III, but even in part II, open-endedness will play a crucial background role.

By countenancing rejection rules as well as open-ended rules, we get a more realistic model of the full range of our actual linguistic rules than is provided by textbook accounts of inference rules. Many further extensions would likely be needed to cover the full range of human linguistic practices.[34] But I think that rejection rules and open-endedness are enough for successful theories of logic and mathematics. At least, this is what I will argue in parts II and III, respectively. The point here is not that inference rules, including rejection rules and open-ended rules, fully exhaust the types of linguistic rules found in natural language, but rather that these rules

exhaust the types of linguistic rules that govern our logical and mathematical terminology.

According to conventionalists, our linguistic conventions—in the form of syntactic rules of inference—can be used to explain logic and mathematics. This explanation is supposed to be acceptable to scientific naturalists. Some may be willing to grant that linguistic rules don't violate naturalism in either of the ways discussed in 1.IV, but others might be skeptical. To answer such skepticism, I need to do more to justify describing these rules as "syntactic" and to explain how inference rule-following itself fits into the natural world. The rest of this chapter aims to do this.

V Acceptance and Rejection

By basing philosophical theories of logic and mathematics on linguistic rules of inference, conventionalists hope to fully integrate logic and mathematics into a naturalistic picture of the world. But for this to work as intended, linguistic rules, themselves, must be accounted for in wholly naturalistic terms. Of course, we could simply stipulate that linguistic rules are primitive notions that count as naturalistically acceptable, but doing so wouldn't be very plausible or satisfying. High-level notions of psychology and sociology aren't the kinds of things that can appropriately be taken as *naturalistic* primitive notions, they must instead be explained in more fundamental terms.[35] If conventionalism is to be scientifically plausible, conventions themselves must be naturalized.

Though they rarely discussed the matter in any detail, I think that historical conventionalists simply assumed that inference rule-following could be understood in terms of actual and possible language use.[36] If this is true then conventions aren't at all mysterious, but are themselves explained in terms of general facts about language use. According to this approach, to follow a linguistic rule *R* is simply to have appropriate behavioral dispositions. Obviously we need an account of what the "appropriate" behavioral dispositions are, but I think that this simple dispositionalist approach

is on the right track. And if dispositionalism of some form is true, then linguistic rules are scientifically respectable notions.

This might be questioned. Dispositions are themselves often analyzed in conditional terms.[37] And these counterfactual conditionals may be thought to rely on a primitive notion of metaphysical modality that it is itself scientifically questionable.[38] Long gone are the days where any and all modality is viewed with suspicion.[39] Yet surely somebody following in the footsteps of the logical empiricists should be wary of modality.[40] And I am wary, at least of *metaphysical* modality. But I don't think metaphysical modality is needed to understand dispositions. Even if dispositions must be understood counterfactually, it is just as likely that it is the modality that is explained in terms of the counterfactuals, rather than vice versa.[41] More importantly, the modality needed here is *physical* or *nomological*, rather than metaphysical. Dispositions, capacities, and the counterfactuals associated with them are of a piece with our standard causal notions. These causal notions are an important and familiar part of standard scientific explanations. So no background metaphysical or logical or conceptual modality need be assumed here.[42]

But it can't simply be taken for granted that rule-following, and thus linguistic conventions, can be naturalized using dispositions. Many philosophers have argued that thinking in general, and inference rule-following in particular, cannot be naturalized.[43] These challenges need to be met head-on. Over the next few sections, I aim to do this by building up to a full dispositionalist account of what it is to follow a linguistic rule of inference. I'll start with the more basic ingredient notions and work up to rule-following, step-by-step. Inference rule-following involves rule-governed inferences; and inferences themselves are transitions between our attitudes toward sentences, so naturalizing inference rule-following will proceed in three discrete steps. In this section, I'll discuss the basic attitudes. In the next, inference itself, and after that, inference rule-following.

In theoretical inference, inferences are transitions between attitudes of *acceptance* and *rejection*. Practical inferences might also involve transitions between or involving *intentions to act*.[44] But the

linguistic rules of inference I have argued for above are of the theoretical kind, so practical inferences will be set to one side. Acceptance and rejection are paradigm cases of the so-called *propositional attitudes*.[45] These attitudes are so-named because they seem to involve a relation between an agent and a proposition. I have no great objection to this manner of speaking, but I don't think it should be taken too seriously. Nobody thinks that the mind grasps propositions in a direct, unmediated fashion (I hope).

For this reason, I usually prefer to talk about accepting and rejecting *sentences* rather than propositions.[46] Full acceptance of a sentence on a particular occasion amounts to believing the content the sentence has on that particular occasion. The relativity to occasions of use is important because of familiar issues of context-sensitivity and indexicality. For this reason, it is usually a bit more accurate to talk of "utterances" or "mental tokens" of some kind, but for the purposes of this book, no confusion will result from simplifying slightly and focusing on sentences.

Similar remarks go for rejecting. Importantly though, as discussed in section IV above, rejection is a matter of exclusion. It isn't equivalent to accepting some other sentence, even the rejected sentence's negation. Even more generally, accepting and rejecting a sentence amounts to having certain behavioral dispositions, including but not limited to dispositions toward the sentence itself (or some related sentence—for instance, a translation of the sentence in question).

The best account of mental states like these is some form of *functionalist dispositionalism.*[47] An agent accepts a sentence *p* just in case the agent is in a mental state that plays the *accepting p role*. The role itself is a matter both of causal relations to other mental states and also of behavioral dispositions. So accepting *p* might involve the disposition to act as if *p* is correct, to use *p* in certain types of reasoning, to express surprise if *p* is shown to be false, *et cetera*. And rejecting *p* might involve acting as if *p* is incorrect, refusing to accept it in reasoning, showing surprise if *p* is shown to be true, *et cetera*.

While "using *p* in certain types of reasoning" includes inferences that involve *p*, there is no deep circularity here. Functional definitions

often involve this kind of benign circularity, as when a functionalist about pain characterizes the pain role in terms of other mental states that are also functionally defined in terms of their connection to pain. There is even a standard and familiar method—the so-called Ramsey/Carnap/Lewis method—for allowing various notions to be jointly characterized in relation to each other.[48]

The important point is that, understood in this way, acceptance and rejection are a matter of having dispositions to use sentences in certain ways and not others. This is less tendentious than it might seem. Even representationalists about belief-like attitudes, who require that some register of each belief be stored in memory, also admit derivative dispositional notions of *implicit* or *tacit* acceptance and rejection. They do this to account for the evident fact that, in a perfectly intuitive sense, our beliefs and disbeliefs at any given moment far outstrip what can possibly be stored in our finite brains.[49]

I have said that I prefer a *functionalist* brand of dispositionalism. One reason for this is easy integration into a general functionalist picture of mental states.[50] Another is easy inter-definition of these states, in both a positive and negative sense. Part of the functional role of rejecting *p* is not at the same time being in the mental state of accepting *p*, in the same manner. This can be built in explicitly, in functionalist terms. But what really matters is that the dispositions that constitute accepting *p* and those that constitute rejecting *p* can't be had by the same person at the same time.[51] Of course, context-sensitive sentences like "I am hungry" may unproblematically be rejected at one time and accepted at another. But even these cannot be both accepted and rejected at any given time.

This incompatibility is not a *logical* matter. It is more like missing a bus and also catching it at the same time. Given what acceptance and rejection *are*, you can do at most one toward a given sentence at a given time. This is not a matter of human limitations. It isn't as if much smarter Martians would be able to both accept and reject that 2 and 2 added up to 4. Because of the incompatible demands of each role, no matter how bright or exceptional someone may be, they cannot both accept and reject the very same thing at the very same time. Obviously, people can make momentary mistakes, verbal slip-ups, and suffer from many other transitory lapses. But

misspeaking and adding a “not” when you don’t mean it is not tantamount to rejecting a claim.

Thus far I’ve been talking about acceptance and rejection as if they each picked out a unique attitude toward a sentence, corresponding to believing and disbelieving the content of a sentence on an occasion, respectively. In reality, each picks out a family of related attitudes. Membership in the acceptance family comes from sharing a broadly similar functional role while remaining distinct. To illustrate: accepting a sentence is different than accepting a sentence as a supposition or hypothesis to reason from. As was already noted, reasoning under a hypothesis is a familiar part of our ordinary inferential practices. The loveliness of formal natural deduction comes largely from capturing this aspect of our practices. Similar points go for rejection as well.

All of the acceptance-like attitudes and all of the rejection-like attitudes are characterized functionally. And the incompatibility claims made for full acceptance and rejection also go for accepting and rejecting under an assumption, for the same reason. Within the context of a single assumption or hypothesis it is impossible to both accept and reject the same thing at the same time. Perhaps different degrees of partial belief also fall into the acceptance and rejection families, but I won’t assume that here. Full belief and belief under a hypothesis are two members of the acceptance family, and full disbelief and disbelief under a hypothesis are two members of the rejection family. All of these attitudes, at least, can figure in *inference*, to which I now turn.

VI Inference

Inference involves patterned changes in our attitudes of acceptance and rejection.[52] The problem of characterizing inference resides in the borderlands of cognitive psychology, epistemology, and logic. Surprisingly, until very recently, the problem had been relatively little studied by philosophers.[53]

When I come to accept “this book will not be read” on the basis of accepting both “if the author is unknown, then this book will not be

read" and "the author is unknown", I have performed an inference. But the notion of *accepting/rejecting on the basis of* is essentially what we are trying to understand. It cannot be taken as primitive if inference itself cannot. At some level of complexity, it is a causal notion. But it isn't a merely causal notion. Not every single causal path from acceptance of these premises to acceptance of this conclusion counts as accepting the conclusion on the basis of acceptance of the premises. If acceptance of these premises causes me, for some reason, to slip and hit my head on the floor, and that causes me, via jostling around my neurons, to accept the conclusion, we can all agree that I have not performed an inference. The lesson is that not every causal path from premise attitudes to a conclusion attitude counts as an inference.[54]

The failure of a merely causal account of inference is neither surprising nor distressing. In fact, this situation is typical with causally realized psychological notions. Humans can feel pain, but so can Martians, even though Martian's don't have any neurons but instead have inflatable cavities in their feet that realize their mental states.[55] Indeed, the *multiple realizability* of mental states has long motivated functionalism in the philosophy of mind. Applying functionalism to inference requires a slight generalization, since inference is not a mental state, it is a *state transition*. To infer is to transition from having attitudes toward some premise(s) to also having an attitude toward some conclusion.

Earlier in this chapter I gave a functionalist-dispositionalist account of accepting and rejecting. To accept some sentence *p* is to be in a mental state that plays the acceptance role toward *p*, likewise for rejection and the rejection role. Here I am adding to this a functionalist-dispositionalist account of inference. An inference is performed when someone's premise attitudes cause them to form a conclusion attitude in a way that plays the inference role.

I won't attempt a full characterization of the inference role here.[56] But I think it likely involves a number of different elements. When *S*'s premise attitudes cause *S*'s conclusion attitude in a way that plays the inference role: (*i*) *S* is not also in the state of judging that the inference in question isn't a good one; (*ii*) *S* is disposed to cite the premises (or their negations) when asked to justify their attitude

toward the conclusion; and (*iii*) the formation of the conclusion attitude isn't wholly automatic, but can be delayed or called into question, at least up to a point. Inferring isn't an action in the same way that raising your arm is, but it is partially under our conscious control, much like breathing. Also, in standard inferences, the premise attitudes must be retained throughout the process of forming the conclusion attitude. These are some of the things that constitute the intuitive inference role.

This is analogous to functionalist accounts of pain. *S* is in pain just in case *S* is in a mental state that plays the pain role, where the pain role, like the inference role, is specified in functional terms. Just as no functionalist about pain ever tried to fully and perfectly spell out the pain role, it would be silly to demand that functionalists about inference fully and perfectly spell out the inference role. Rough and intuitive characterizations convey the general idea. The further development of cognitive science is likely to expand our understanding of the psychology of inference. But nothing will undercut the basic functionalist framework itself.

Just as I took the acceptance and rejection roles to be inter-defined in some ways, I have also appealed to other mental states in characterizing the inference role. Once again, this kind of circularity is benign within a functionalist framework. We even saw above that the acceptance and rejection roles involved inference. I think it is likely that these roles are partly inter-defined, where the inference role and the acceptance and rejection roles together pick out a broad pattern of actual and possible interrelated behaviors.

The benign circularity here is simply a reflection of the obvious fact that states like accepting and rejecting and processes like inferring are deeply and intimately intertwined elements of our psychology. One may question how beings with brain states that realize these complicated and deeply intertwined roles could have come to be, but this is really no more mysterious than any of the other physiological or psychological complexities we find in nature. Incremental, step-by-step processes, whether naturally selected for or arising socially, explain complexity. Acceptance, rejection, and inference did not come into the world fully formed, but instead developed gradually, with creatures who accept, reject, and infer being preceded by

creatures who only proto-accepted, proto-rejected, and proto-inferred. The situation here, whatever its ultimate details, presents no special difficulties for developmental psychology.

A somewhat related point concerns indeterminacy in the notion of inference itself. There are many cases where we make quick judgments that aren't merely simple and direct exercises of conceptual knowledge in response to some stimulus, without clearly being inferences either. Some of these cases are inference-like in that they involve some causal state transitions in our attitudes—consider the judgment we make in response to Gettier cases, for example.[57] Are these inferences? In some cases the matter is unclear. The kinds of inferences I will be concerned with here are, for the most part, clear cases of inference. Borderline cases of inference can be left to one side, though much of what I say will apply to them as well (or, at least *some* of them).[58]

As I've said, a full account of the inference role must await the further development of cognitive psychology. But however it is ultimately spelled out, the functional role of inference bottoms out in connections to our behavior with sentences, both direct and mediated. For instance, when agent S infers sentence ϕ from the sentences in set Γ, S is disposed to cite sentences of Γ in response to challenges of various sorts. The point is not that this and this alone is constitutive of inference, and of course this particular disposition can be masked, blocked, or otherwise interfered with.[59] But it is quite plausible that performing an inference constitutively involves dispositions to behave with the involved sentences in some but not other ways.

Inference doesn't only involve transitions from full belief to full belief, or full disbelief to full disbelief. It also includes transitioning from accepting something under a hypothesis to accepting something else under the same hypothesis. Likewise for rejection. We can even count more complex transitions as inferences, such as fully accepting something on the basis of a transition made wholly under a hypothesis (this is often what happens when we discharge a hypothesis in natural deduction systems). In fact, many distinct types of complex state transitions plausibly count as inferences. In line with how we typically use the word "inference", we should also allow that

inferences can be *vacuous* (involving no premises) or *trivial* (involving inferring a sentence from itself). We can take this as stipulative, just to make things easier, though these cases are very naturally seen as limiting cases of our standard inferential practices.

Inference is a functionally defined type of state transition between functionally defined mental states. Combining these two functionalist views, in the manner urged here, leads to an account of accepting, rejecting, and inferring that bottoms out in our behavior—principally our behavior with sentences, syntactically characterized. I think I have said enough to indicate *exactly why* and *generally how* I think these notions can be naturalized. But this doesn't yet show that inference rule-following can be similarly understood. In fact, many have doubted that rule-following *of any kind* can be accounted for in dispositional terms.

VII Rule-Following in the Time of Kripkenstein

Despite such doubts, I think that a functionalist–dispositionalist approach is also the best approach to rule-following in general and inference rule-following in particular. Before turning to inference rule-following, I need to answer some general worries about rule-following dispositionalism.

Dispositionalism is the main target of Saul Kripke's infamous "Kripkenstein" arguments, developed in his *Wittgenstein on Rules and Private Language*, one of the most widely discussed works in contemporary philosophy.[60] Given the prominence, influence, and importance of Kripke's arguments against general rule-following dispositionalism, some response must be given before turning to inference rules. I have already addressed Kripkenstein's anti-dispositionalist arguments at length in print, aiming at a canonical response.[61] So here I will only give a general overview of how I think Kripke's central arguments can be answered.

Following Wittgenstein, Kripke challenges us to say what it is that makes us mean addition by "plus", as opposed to some deviant quaddition function that agrees with addition for small numbers but swiftly diverges. He considers a simple dispositionalist reply,

according to which someone means addition by “plus” just in case, when they are asked “what is *n* plus *k*?” for numerals “*n*” and “*k*” they are disposed to answer with a numeral “*m*” that denotes the sum of *n* and *k* (as opposed to the quum). Kripke raises three challenges for this view, and suggests that versions of his challenges plague *any* dispositionalist approach to rule-following, meaning, or thought.

(*i*) FINITUDE. Kripke notes that we have no dispositions to respond to “plus” queries when the associated numerals are astronomical. If they are large enough, we would die before being able to reply. If we even attempted to reply we’d never be able to produce a token (whether written or spoken) of the numeral referring to the correct answer. So our finitude seems to make it impossible for us to follow the addition rules and mean addition by “plus”, if dispositionalism is correct.

But Kripke’s charge is too coarse-grained. While we don’t have simple dispositions to respond to large arithmetical queries, we may have complex dispositions to reply. A simple disposition is akin to a brute response, while a complex disposition is a response that is the outcome of some intermediate actions that we are disposed to perform when given the query.[62] We are exercising just such a complex disposition when we execute the elementary school addition algorithm in response to a long arithmetical query.

This extends the range of our dispositions, but not far enough to cover all possible cases. We would still die before completing the process of adding two very large numbers.[63] Yet even in such cases we have the disposition to undertake step 1 of the process, the disposition to undertake step 2 upon finishing step 1, and the disposition to undertake step 3, and so on, for each step in the process. When we have linked simple or complex dispositions (singular dispositions) of this kind I say we have composite dispositions.

I believe that we mean addition by “plus” because we have composite dispositions that independently specify the answer to any given arithmetical query. This is true even if we can’t actually string these dispositions together to finish the process because of our finite lifespans. In noting this, it is important not to confuse a complex disposition, like:

> Ludwig has the disposition to execute step 1 of process *A*, and then step 2 of process *A*, and … and then step 1,000,000 of process *A*

With the related composite disposition:

> Ludwig has the disposition to execute step 1 of process *A* and the disposition to, with step 1 completed, execute step 2 of process *A*, ... and the disposition to, with step 999,999 completed, execute step 1,000,000 of process *A*

Complex dispositions are dispositions to undertake the entire process, while composite dispositions are singular dispositions (whether simple or complex), had in a moment, that together wholly constitute the entire process. It is because we have composite dispositions of this kind that we can follow rules with unbounded cases, whether in addition, in grammar, or anything else.

(*ii*) ERROR. Kripke also claims that dispositionalists cannot account for error. Suppose that Ludwig is disposed to make a mistake and reply to "what is 5 plus 7?" with "13". In this situation, by the dispositionalist's own lights, Ludwig does not mean addition by "plus". But this is absurd. Surely all of us have dispositions to make mistakes, perhaps even to systematically make mistakes. Kripke charges the dispositionalist with equating performance with correctness, and in so doing destroying the very possibility of error.

This problem has generated a large number of responses.[64] However, I don't think that any of the ideas in the literature manage to solve the problem all on their own. One idea is to limit the meaning-constituting dispositions to normal conditions or situations, for Ludwig. If these normal conditions can be spelled out non-circularly, then if mistakes are impossible in them, the problem is avoided.[65] But it seems unlikely that we can spell out conditions in which mistakes are completely impossible. Another idea is that we can check and recheck our work, focusing only on those answers that are stable—that tend to survive checking and rechecking. But in cases where Ludwig is tired, or drunk, or the like, either no stable

answer will be forthcoming or the intuitively "wrong" answer will be stable.[66]

While these responses don't work singly, together they are enough to ward off error. Let's use normal conditions, but instead of requiring perfection, let's require only that the same answer be given in the overwhelming majority of normal situations—call these *M*-general dispositions. And let's require that those answers be stable in normal conditions, since normal conditions are exactly designed to ward off the problems with stability. All together, we can say that the "right" answer to "what is *n* plus *k*" is "*m*" just in case Ludwig has a stable disposition to reply to the query with "*m*" in the overwhelming majority of normal situations. Several broken parts lead to an unbroken whole, using a method of multiple screenings to ward off error.[67]

(*iii*) NORMATIVITY. Kripke's last problem concerns normativity. Following the addition rules and meaning addition by "plus" is supposed to provide us with *justification* for replying with the sum rather than the quum. But, he claims, a disposition to reply with the sum is not at all the right sort of thing to provide us with justification, so dispositionalism must be wrong. Many have seen this as a deep and powerful problem, but others have worried that there is really no problem here at all, over and above the problem of error.[68] I fall closer to the second camp, myself.

Whatever the nature of semantic normativity, there is nothing to stop dispositionalists from appealing to whatever theory of normativity they favor to explain their justification for giving the *semantically correct* answer to addition queries. And they can obviously explain whatever *practical* justification they have for replying with the right answer. The "right" answer itself was determined above when dealing with error. So there just doesn't seem to be a special problem here for dispositionalists. There are some interesting puzzles concerning the link between norms and rules, but they are more general. I will discuss them in chapters 4, 6, and 11, where they most naturally arise. Here, as with the previous two problems, the central issue with Kripkenstein's case against dispositionalism can be put simply: it successfully targets only a simplistic, strawman version of rule-following dispositionalism.

A general dispositionalist approach to rule-following is still in the cards, post-Kripkenstein. With these general worries sidestepped, we still need to show that a particular dispositionalist approach to inference rule-following can be given.

VIII Inference Rule-Following

Consider again an inference rule like modus ponens:

$$(MP): \quad \frac{\phi \quad \text{if } \phi \text{ then } \psi}{\psi}$$

Say that someone *violates* a standard inference rule like (*MP*) when they accept all of the premises but reject the conclusion. Can we say that someone follows an inference rule when they are disposed not to violate the rule? No, for intuitively following a rule requires more than merely not violating the rule, even in a dispositional sense. Somebody without the concept of the conditional ("if") never violates (*MP*), since they never accept both premises. Yet, since they never form any "if" attitudes, they surely don't follow the rule of modus ponens.

Slightly more demanding than the notion of failing to violate an inference rule is that of complying with the rule. Someone complies with a rule like (*MP*) just in case they either reject one of the premises or accept the conclusion. This is more demanding because someone who never formed "if" attitudes would never violate (*MP*), but they wouldn't automatically comply with it. If you comply with a rule then you don't violate it, but not always vice versa. Can we say that someone follows an inference rule if they are disposed to comply with it? Obviously not, since they could trivially comply merely by always accepting the conclusion, without forming any "if" attitudes at all. What if we require that they both comply and have attitudes toward both premises and the conclusion; is that enough?[69]

No: intuitively following a rule requires more than this kind of compliance. Somebody disposed to reject all conditional sentences

would have a disposition to comply with (*MP*), but they don't intuitively follow it on that basis. For if they did they would also follow the rule:

$$(MP\bot): \quad \frac{\phi \quad \text{if } \phi \text{ then } \psi}{\bot}$$

Where "⊥" is some arbitrary absurdity.

Let's say that someone infers according to a rule like (*MP*) when their acceptance of the premises of an instance of (*MP*) causes them to also accept the conclusion of that same instance in a way that plays the inference role (this might require, in this type of case, that the premise attitudes are retained throughout the process). Can we say that someone follows an inference rule if they are disposed to infer according to it? Presumably we can. But can we always require this of rule-followers? No, for it is too demanding. Beings like us aren't disposed to accept all conclusions of things we accept, according to rules we follow. That process would never end.

It is important that this demand not be misunderstood. It is possible to infer according to a rule like (*MP*⊥), even though the conclusion is never accepted. This is better illustrated by leaving the made-up rule of (*MP*⊥) behind to focus on a more realistic case. Consider the rule of explosion:

$$(Explosion) \quad \frac{\phi \quad \text{not } \phi}{\psi}$$

This rule, valid in classical logic and many other systems, has inconsistent premises. Given this, how can anyone ever infer according to this rule? Must we think that inferring according to (*Explosion*) requires the ability to fully believe a sentence and its negation at the same time? I don't think so. As discussed previously, acceptance is a family term. And the acceptance family includes not only full, heartfelt belief, but also supposing to be the case. We can, and often do, suppose two incompatible or inconsistent things to be the case. Inferring an absurdity under a supposition is just to accept

the supposition as itself absurd. This too, is inference. Mock acceptance of the premises causes, in a way that plays the inference role, mock acceptance of the conclusion. At no point need an actual absurdity be genuinely accepted.

The problem with taking a general but simplistic disposition to infer as part of rule-following is that it requires too much of us. Consider a rule like "or"-introduction:

$$(\text{or } I) \quad \frac{\phi}{\phi \text{ or } \psi}$$

This rule is valid in classical logic and in many other systems. But infinitely many distinct instances of this schematic rule can be formulated in our language. So even for a fixed premise, there will be infinitely many inferences in accord with this rule. Would a disposition to infer according to (or *I*) have to involve a disposition to make all of these infinitely many inferences? It seems not, since inference rule-following can't possibly be so demanding if we are ever to follow inference rules.

A similar problem also afflicts weaker dispositionalist demands. For example, Julien Murzi and Florian Steinberger presented a dispositional account of inference rule-following where following a rule like (*MP*) requires only that "entertaining" or "supposing" the premises of a (*MP*) argument causes the conclusion to be "considered", where consideration doesn't imply any inclination toward acceptance, even within the scope of the relevant supposition.[70] Murzi and Steinberger admit that their approach faces roughly the problem I've been considering, which they call "the problem of alternatives". This will be a problem for any brand of dispositionalism about inference rule-following that requires a disposition to enter into an occurrent mental state with respect to the conclusion, upon accepting the premises. I think the best response to this type of problem is to adopt a more thorough-going dispositionalism which doesn't require any particular occurrent mental state at all, nor even a disposition toward one.[71]

Doing so requires that we look at even broader patterns of behavior. Let's say that Ludwig performs a modus ponens inference but then immediately claims he was wrong to have done so, and resolves never to do so in the future. In this kind of situation, it would be hard to describe Ludwig as following (*MP*). This is true even if in the moment he was disposed to infer according to modus ponens. Whatever else we may do, when we follow an inference rule we also enforce the rule, in two directions. In the first direction, we sanction both compliance with the rule and inference according to the rule. This need not always be social, as it might merely involve altering our own behavior so as to comply with the rule, and resolving to do so in the future. In the second direction, we refuse to sanction and sometimes even condemn noncompliance with the rule, whether in ourselves or others, upon reflection.

Sophisticated enforcement behavior might depend upon having some explicit representation of the rule. But that isn't essential for rule-following, as such. More primitive, direct, and intuitive enforcement behaviors can exist without any representation of the rule itself. Children are often able to enforce the rules of a game they are playing long before they are able to explicitly codify the rules involved. This is one of the upshots of Piaget's famous work on child development.[72] Likewise, a bit of reasoning can be rejected without any explicit representation of the rules used in the reasoning, on the part of the rejecter. A two-pronged disposition to enforce the rule, at least in some limited class of circumstances, seems required for true rule-following.

Let's bring all of these strands together. Those who follow an inference rule like modus ponens are disposed to avoid violating the rule and to enforce the rule in two directions. These dispositions are over a very general range of circumstances (though enforcement might be limited to yourself, in the limit). But while necessary for following (*MP*), neither of these dispositions alone is sufficient. Nor are they jointly sufficient. This might be doubted, but I think that truly following a rule also requires, at least in some circumstances, compliance with the rule. It might also require inferring according to the rule, in some even more limited range of circumstances. These seem needed because you only truly *follow* a rule, especially an

inference rule, when it is a rule that you yourself *might possibly use in the intended way*.

So let us also require a disposition to comply with the rule whenever there is any disposition at all to form attitudes toward the rule's premises and conclusion. And also a more limited disposition, not to infer according to the rule, but to infer according the rule when given the chance and when other conditions are in place. What is it to be given the chance? It is when someone or something brings the conclusion to your direct attention, perhaps by querying you on the matter (I also want to allow that your own sub-personal cognitive systems can bring the conclusion to your person-level, direct attention). What other conditions need to be in place? For a standard rule like (*MP*), the extra conditions include a disposition to continue to accept the premises throughout the process. If a modus ponens conclusion is antecedently unlikely, you would probably reject a premise instead rather than accepting the conclusion, thereby complying with the rule without inferring according to it.

Though they are fairly vague and general, I think that these four dispositions—for non-violation, two-way enforcement, compliance in relevant cases, and inference in relevant cases when given the option—are jointly necessary and sufficient for inference rule-following. To ward off Kripkenstein's problem of error, all of these dispositions should be understood as *M*-general and stable over normal conditions. But the dispositions toward compliance and inference can be either simple, complex, or composite, for all I've said here (in the next chapter, I'll return to this issue). I also haven't built it in to this approach that we can follow only syntactic rules of inference. As section III illustrated, a rule like modus ponens is syntactic, and our behavior with the rule outside of strict inference illustrates this. But the argument for that is, for the most part, independent of rule-following dispositionalism. The discussion here shows that rule-following itself is a broadly syntactic notion, even if some rules are not syntactic in the direct way that modus ponens is.

This account of inference rule-following allows that we are highly imperfect epistemic engines. I may utterly fail to come to believe a highly relevant, important conclusion that follows from some things I believe, via an inference rule that I follow, not because I am not

following the rule, but because I simply don't think of it. This case can't be warded off by an appeal to normal or ideal conditions, since even when in an epistemically ideal situation I may not draw the right conclusions. I could simply fail to think of it. Perhaps this means I'm not very good at reasoning, but it isn't because I don't follow the right rules, it is because I don't exploit them to my epistemic advantage. Yet I still follow the rules and I would draw the right inferences if given the option, all else being equal. Both beginning students and world-class mathematicians can fail to extract important conclusions that can be derived from things they accept via rules that they follow. Alas.

None of the dispositions that underwrite inference rule-following require any particular occurrent mental states. Nor do they require any explicit representation of the rule on the part of the rule-follower. Rule-following is compatible with a great many patterns of overall inferential behavior. Yet it is not compatible with *all* patterns of inferential behavior. There may well be a healthy amount of indeterminacy involved in rule-following, but there are some rules that we clearly follow, and some that we clearly do not.

IX Things We Do with Sentences

Let's bring the discussions of attitudes, inference, rule-following, and inference rule-following together. What we have is a three-tiered functionalist theory, one that is dispositional *thrice over*.

Tier one: for *S* to be in the state of accepting *p* is for *S* to be in a mental state *M* that plays the acceptance role with respect to *p*, and similarly for rejection. Where *M* plays the acceptance role for *S* and *p* just in case *S* has various behavioral dispositions regarding *p* (perhaps indirectly, since we want different languages to be able to express some of the same contents). Tier two: for *S* to infer *q* from $p_1,\ldots ,p_n$ is for (in the standard case) *S*'s acceptance of each of $p_1,\ldots ,p_n$ to jointly cause *S*'s acceptance of *q* in a way that plays the inference role, where playing the inference role is a matter of *S* having certain behavioral dispositions regarding $p_1,\ldots ,p_n$ and *q*. Tier

three: for *S* to follow an inference rule *R* with premises $p_1,\ldots,p_n$ and conclusion *q* is for *S* to be disposed to not violate *R*, to enforce *R* in two directions, to comply with *R* when having a disposition to form attitudes toward all of *R*'s component sentences, and to infer according to *R* when given the chance and having the disposition to continue to accept the premises.

Some have thought that inference itself must be understood in terms of inference rules.[73] By contrast, I think that the attitudes and inference are explained in terms of each other, while inference rule-following involves overall patterns in our practices of inferring. Nothing much hangs on this here. If someone prefers to understand my approach as requiring that all three tiers are inter-defined, I have no great objection. In any case, taken together the three-tiered theory is an account of inference rule-following that bottoms out in the things we are disposed to do with sentences. This fully vindicates the idea that *following rules of inference* is a naturalistically acceptable notion.

The first tier of my theory concerns attitudes of acceptance and rejection rather than speech acts like assertion and denial. This is because I think working with attitudes rather than speech acts results in a more natural theory, largely because inferences are more plausibly seen as state transitions between attitudes rather than speech acts. Maybe being in the state of accepting that *p* is simply being in the state of being disposed to assert that *p*, and similarly for rejection and denial. But I doubt it. Our dispositions to perform speech acts are influenced by all kinds of things, including context and pragmatic and practical factors. So I prefer instead to see the speech acts and the attitudes as closely related, each ultimately being constituted by behavioral dispositions, principally involving syntactic items like sentences.

Accepting this account doesn't amount to saying that non-linguistic animals couldn't be in related mental states, such as proto-acceptance or proto-rejection. But the kind of complexity involved in at least some inference rule-following that we engage in is so demanding that it is hard to imagine non-linguistic creatures doing it. Even without language, when inferences are performed, some physical (syntactic) representational states are entering into complex

causal interactions, and that is still what matters and what constitutes acceptance, rejection, inference, and whatever rule-following is involved.

Moving from non-linguistic creatures back to us—the triply functional-dispositional account of linguistic rule-following allows us to see inference rule-following as implicit and broadly syntactic. When the inference rules themselves are implicitly followed and broadly syntactic, as was the case with linguistic rules of inference like modus ponens, then we have a fully naturalistic, syntactic account of linguistic conventions and our applications of them.

Against this, worries have been raised about the explanatory worth of the standard linguistic rules posited by linguists. W.V. Quine, arch-critic of conventionalism, raised a general worry of this kind when discussing Chomsky-style linguistics. Quine's concern is based on a distinction between a rule *R fitting* a speaker's behavior, where fit is simply a matter of true description, and a rule *R guiding* a speaker's behavior, where this is a matter of cause and effect at some level of abstraction. Quine claimed that:

> …behavior is not *guided* by the rule unless the behaver knows the rule and can state it.[74]

From this it seems to follow that implicit (dispositional) linguistic rules of inference don't truly guide our behavior, and this seems to undercut much of the motivation we had for positing them in the first place, at least in the eyes of Quine.

Whatever plausibility this has at first glance, the distinction is based on a false dichotomy. Quine himself recognized that the morphological and grammatical rules of contemporary linguistics didn't neatly fit into his dichotomy between guiding and mere fitting. This wasn't a problem for him, since he already rejected those rules because he thought they went beyond our (behavioral) evidence. For a behaviorist like Quine, this evidence is all that the scientist has to go on. I am not a behaviorist in Quine's sense. I don't think anybody else is either, at least in the contemporary debate. But even on quasi-behaviorist grounds, we have the resources to distinguish between different collections of inference rules that lead to the very

same set of valid arguments. This is because they can still sanction different behaviors along the way. One collection of rules may allow a direct inference from some premise *q* to *p*, while the other only an indirect one, going by way of some third sentence *r*. This means that the enforcement behavior, as well as some of the involved inferences, may well be different, at least in their fine-structure. So even if Quine's criticisms worked against grammatical rules—something that is by no means certain—they don't work against inference rules, which can be distinguished behaviorally in ways that even Quine himself should accept. And as we have already seen in the case of modus ponens, some of these behaviors are rooted in our recognition of the syntactic form of the sentences involved.

Those of us who posit syntactic rules of inference will surely argue that they too fail to fit neatly into either of Quine's categories. Inference rules do not just describe or "fit" our actual linguistic behavior, since they extend to novel cases. And while they need not always be behavior "guiding" in the manner suggested by Quine, since that requires explicit representation, they are behavior-guiding in a real but less demanding sense. This comes out in all of the various dispositions that underwrite rule-following—we reason in accord with them, accept corrections toward them, frown upon violations, and so on.[75] This doesn't mean that *everything* about the normative role of inference rules is already clear.[76] But it does mean that most of the puzzles and confusions about the normative role of linguistic rules are answered or side-stepped by my approach.

Among our linguistic conventions are linguistic rules of inference. Following these rules—adopting these conventions as opposed to others—is a matter of having complicated dispositions to behave in some ways but not others with sentences. We can see this when dispositionalism about attitudes is combined with dispositionalism about inference and then fed into the dispositional account of rule-following. This is all as it should be, for the purposes of conventionalism; adopting linguistic conventions concerns language use, in a broad sense. The existence of linguistic conventions requires no implausible psychological assumptions, such as explicit representations of the rules that we follow.[77] And since they are rooted in our linguistic behavior, they are fully compatible with a

standard scientific picture of the world—no metaphysical weirdness or cognitive spookiness here.[78] Linguistic conventions, understood and used in the fashion suggested here, are well poised to do useful explanatory work as surely as their morphological and grammatical cousins.

LINGUISTIC CONVENTIONS ARE wholly syntactic linguistic rules of language use, in a broad sense. Among our linguistic conventions are rules of inference. These rules go beyond what is modeled in standard natural deduction systems, and while a full account of human language would need further additions, I will not be concerned with any further extensions here. I think that many philosophers will agree with nearly everything I've said in these first two chapters, given how weak and naturalist-friendly my assumptions are. And yet, on this meager basis I will develop a philosophical theory of both logic and mathematics. I stress the weakness and plausibility of these points because openly loading one's sleeves at the start of a card trick makes it less impressive when the correct card is pulled out at the end. My sleeves are (almost) empty; let the performance begin.

---

1 Lewis (1969), page 78. Lewis also gives a quantitative variant of this definition.

2 See the results reported in Skyrms (2010).

3 See Skyrms (2000) and chapters 1, 3, and 11 of Skyrms (2010); see also Syverson (2002) for similar and perhaps even bolder claims connecting results of this kind to logical conventionalism.

4 Skyrms's related claims about the philosophical importance of his results for the theory of meaning are also overstated—see Birch (2014) and Shea, Godfrey-Smith, and Cao (2017).

5 See page 174 and 208 of Lewis (1969).

6 Lewis (1969), page xii.

7 Lewis's definition also requires that members of a linguistic population have various intentions regarding the regularities in question, and this *seems* to require that these regularities be "explicit" in a way that Quine and others have found problematic—see the next section as well as 7.III for discussion. In response, a Lewisian conventionalist could deny that the expectations and preferences concerning these regularities require explicitness. Alternatively, they could appeal to the just-mentioned results of Skyrms (2010).

8 Quoted from Russell (1921), page 113.

9 As discussed in the previous chapter, the main offender here is Carnap—see his (1934), (1939), (1942), and (1943), among other writings. See also Hellman (1986).

10 A similar example is given in Hirsch (2008).

11 Obviously, there is a sense in which we only directly use tokens, but our dispositions involving tokens specify a "use" for a type. But care must be taken in saying this, because different patterns can cluster around the same syntactic type. See especially chapter 5 for relevant discussion.

12 Boghossian (2014) and Broome (2014*a*) build inference rule-following into the very nature of inference itself, but I think this goes too far. I think it is a conceptual possibility that an agent can perform an inference without following any inference rule at all, even if this isn't the typical case.

13 Those who don't think English speakers follow modus ponens are free to focus on a different example.

14 See Chomsky (1957) and (1965); Chomsky's theories have changed since these initial presentations, see Jackendoff (2002) for a partial overview.

15 See Chomsky and Halle (1968).

16 See Berko (1958); my account follows Pinker (1999), pages 14–15.

17 The symbol " " expresses the concatenation function: so "a" "b" = "ab"—see Quine (1940).

18 In section 4 of my (2017*a*).

19 Peacocke's account of primitively compelling inferences also stresses that we find certain inferences compelling "because they have a certain *form*", quoted from page 135 of his (1992). Here I am not claiming that these inferences need to be primitively compelling, only that they are syntactic. My notion of "basic" rules (see 3.II) more closely corresponds to Peacocke's notion.

20 This terminology is meant to suggest Carnap's (1934) "transformation rules".

21 The famous competence/performance distinction was first made in Chomsky (1965).

22 Wason (1966), (1968) and Wason and Shapiro (1971).

23 See Byrne (1989), Schroyens and Schaekan (2003), and the summary in Oaksford (2005).

24 Quoted from van Bentham (2008).

25 See Cosmides and Tooby (1992) and Davies, Fetzer, and Foster (1995).

26 Natural deduction systems were first presented in Gentzen (1934) and Jaśkowski (1934).

27 Not all, however—Carnap (1942) and Łukasiewicz (1957) treated rejection axiomatically; modern bilateralist treatments (to be discussed below) derive from Smiley (1996).

28 See Frege (1960) and also Geach (1965) for this view; see Ripley (2011) for a general survey of relevant issues concerning negation, rejection, and the

speech act of denial. See Smiley (1996) for critical discussion of Frege's arguments.

29 I have in mind the dialetheism of Priest (1987)—see chapter 3 for fuller discussion.

30 See Smiley (1996) and Rumfitt (2000). A somewhat related but more semantic approach is the "negative conclusion logic" of Harman (1986*a*). Price (1983) and (1990) are also relevant. Interestingly, Frege (1879) included a force indicator for the speech act of assertion in the first presentation of modern formal logic, but later treatments have not followed him in this.

31 See Priest (1979) and (1987).

32 Given their acceptance of contradictions, it might be wondered whether dialetheists might think we can both accept some contradictions while also rejecting all contradictions, but Priest himself has devoted considerable effort to arguing that it is impossible to both accept and reject a sentence at the same time; see Priest (2006) and section V below. See also my discussion of tonk and rejection in chapter 5 and in my (2015*b*).

33 See, for example, Lavine (1994), McGee (2000), Parsons (2001), and Williamson (2006*a*).

34 See Brandom (1994) and Sellars (1953) for some relevant further extensions.

35 Putnam (2015*a*) reports that in the late 1950s Wilfrid Sellars thought that "the notion of a "rule" is a rock-bottom notion that needs no further explanation" (quoted from page 45). Brandom (1994) has an arguably quite similar approach. Boghossian (2014) also thinks something similar, but recognizes his assumption as a threat to naturalism.

36 See Carnap (1955) for a rare explicit discussion by a conventionalist of rules and analyticity in natural, as opposed to formal, languages.

37 These analyses take off from Ryle (1949).

38 Metaphysical modality regained philosophical popularity after Kripke (1980); and essence took off after Fine (1994).

39 The most famous skeptic is Quine; see his (1953) for a representative critique.

40 Empiricists have been wary of modality at least since Hume (1739); for a late statement, after the twentieth-century revolution in modal logic, see the end of Ayer (1982).

41 For an account of how this would go, formally, see appendix 1 in Williamson (2007).

42 For a similar middle position between anti-modal empiricism and metaphysical excess, see Putnam (1992).

43 See Boghossian (2014) and Wright's (2014) commentary. See also Kripke (1982), Lucas (1961), Penrose (1989) and (1994), and Searle (1980).

44 For some discussion, see Broome (2014*a*).

45 Introduced in Russell (1921).

46 Carnap (1947) gave an early, sentence-based account of attitudes. For criticism, see Fodor (1978).

47 Dispositionalism about belief is endorsed by Audi (1972), Marcus (1990), and Schwitzgebel (2013) among others. I also count interpretationalists about belief, such as Davidson (1984) and Dennett (1987), as dispositionalists in this general sense.

48 See Ramsey (1978), Carnap (1958), and Lewis (1970).

49 See Audi (1994), Price (1969), and Searle (1992) for relevant discussion.

50 The canonical pioneering works on functionalism are Armstrong (1968) and Putnam (1960). Functionalism is popular for non-conscious mental states, but is often rejected for conscious mental states—see Block (1978), Searle (1980), and Chalmers (1995) for influential antifunctionalist arguments.

51 This isn't to say that people can't make mistakes and contradict themselves. Nor that people can't mistakenly treat co-referring terms as if they picked out different objects, whether in the same language or across languages—see Frege (1892) and Kripke (1979), for famous cases. There are various strategies that are available. See Burgess (2005*a*) for a simple, dispositionalist-friendly approach to crosslanguage cases.

52 Again, this is true of theoretical inference, but perhaps not practical inference.

53 See Boghossian (2014), Wright (2014), and Neta (2013) for recent discussions.

54 See Boghossian (2014) for further discussion.

55 This example is from Lewis (1980).

56 In unpublished work, I develop a full functionalist approach to inference.

57 See Gettier (1963).

58 Malmgren (2018) argues that (in my terminology) borderline cases of inference are a problem for standard uses of inference in epistemology. I disagree—see the discussion in 6.II.

59 See Martin (1994), Johnston (1992), and Smith (1977).

60 In Kripke (1982), purporting to be a reading of Wittgenstein (1953).

61 In my (forthcoming *a*).

62 Though they don't use this terminology, Blackburn (1984*b*) and Shogenji (1993) have appealed to something like this distinction in answer to Kripkenstein.

63 See Kusch (2006).

64 For a taste see Fodor (1990), Forbes (1984), Forster (2004), and Ginsborg (2011).

65 In addition to Kripke's discussion, see Boghossian (1989) for further development.

66 See Guardo (2012).

67 For more details on this see my (forthcoming *a*). There is also a relevant worry about belief holism that has been championed by Paul Boghossian in his (1989) and (1990). For responses see Miller (2003) and my unpublished work on normal conditions in metasemantics.

68 See Fodor (1990) for an example of the latter reaction.

69 I think that this is basically Boghossian's (2000) approach, though the term "compliance" is my own.

70 In Murzi and Steinberger (2013), which is a reply to Besson (2012), which itself draws on Harman (1986*b*). Besson's discussion is focused on attacking dispositionalism about logical *knowledge*, but it is also relevant to dispositionalism about *following* logical rules.

71 This includes phenomenological states as well, such as a "feeling of going on as before", or the like.

72 For a short overview, see Linaza (1984).

73 Boghossian (2014), Broome (2014*a*), and Brandom (1994).

74 Quine (1970*a*), page 386—there is a typo in this quote in the published version of my (2017*a*).

75 Once again the situation is somewhat similar to functional explanations generally, where it is quite common to see a general disposition explaining or accounting for more particular dispositions as well as singular behaviors. See also the discussion of local regulation in 4.VII.

76 See Rawls (1955), Midgley (1959), and Searle (1969) for relevant discussions. I will return to discuss how conventionalism accounts for normativity in 4.VII.

77 Explicit rule-following arguably leads to a problematic regress; see Carroll (1895) and Quine (1936). Quine's arguments are discussed in 7.III and in my (2017*a*).

78 I've already mentioned that Boghossian (2014) and Wright (2014) at least flirt with the idea that a proper account of inference requires the rejection of naturalism; Brandom (1994) arguably does as well. The second part of this chapter has now explained my alternative to all of these anti-naturalist moves.

PART II

REVITALIZING LOGICAL CONVENTIONALISM

3

Unrestricted Logical Inferentialism

As discussed in chapter 1, logical conventionalism holds that the facts about logical truth, logical falsity, and logical validity in any given language are fully explained by the linguistic conventions of that language. And as discussed in chapter 2, the relevant conventions are syntactic inference rules that are rooted in linguistic behavior. Against this background, the first step toward logical conventionalism is giving a rule-based, inferentialist account of the meanings of natural language logical constants.

In some form or other, inferentialism is a very widely accepted theory of the meanings of logical constants, at least in the philosophical literature. There are different versions of inferentialism, but in any plausible form, inferentialism is a *metasemantic* theory explaining how logical constants like “and”, “not”, and “all” get their meanings. Most forms of inferentialism are gerrymandered to avoid putative counterexamples or to build in partisanship toward a favored logic. I think that such gerrymandering undermines the philosophical motivation for view. I will develop a pure and extreme—yet well-motivated—version of inferentialism. This unrestricted version of logical inferentialism is the foundation upon which logical conventionalism is built.

I Logical Inferentialism

Without worrying yet about what makes them “logical” (that will be addressed in section III), consider the canonical logical constants: expressions like “and”, “or”, not”, “there exists”, “all”, *et cetera*. In the logic classroom we learned that the meaning of “and” is a binary truth function that gives the output *true* when fed the binary sequence of

truth values *true,true* as an input, but gives the output *false* for all other binary sequences of truth values. The logic used in the logic classroom is classical logic, and though it is accepted by most philosophers, classical logic is hardly uncontroversial.[1]

Because it is both the most widely accepted logic and also has the benefit of familiarity and mathematical elegance, I'll assume here that the English logical constants have their classical meanings. This assumption is actually less partisan than it seems, since my inferentialist/conventionalist view is applicable to any other plausible logic as well. This kind of flexibility and pluralism has long been associated with logical conventionalism; I'll say a bit more about it later in this chapter and in chapter 4, but a full discussion of pluralism will have to wait until chapter 5.

So assuming for the time being that "and", "not", and the like have their classical meanings in English, *how* did they get those meanings? How do sounds and marks come to express these particular truth functions? It is important to realize that this question is *metasemantic* rather than *semantic*. The natural semantic question about the meaning of "and" is *what* does "and" mean? For the moment we are assuming, without loss of generality, that the meaning of "and" is its standard truth function. This answers the semantic question, but the metasemantic questions remain.

The metasemantics of logic is particularly vexing because there is less material to work with than elsewhere. Contrast this with the metasemantics of a word like "dog". In some fashion or other, the word "dog" gets its meaning causally (at least in part). This is to say that the word "dog" manages to pick out all and only the dogs partly because language users causally interact with dogs and use (and are disposed to use) the term "dog" in ways that are causally related to dogs.[2] Yet for logical expressions like "and" and "not", there isn't a remotely straightforward causal metasemantic story to tell. Truth functions aren't things in the world that we can causally interact with. So at first glance it is not at all clear how "and" manages to mean what it does.

It was only in the twentieth century that philosophers managed to come up with the start of a plausible metasemantic approach to the

logical expressions. In broadest outline, the approach is summed up in the following principle:

> **Logical Inferentialism**: In any language, the meaning of a logical expression is fully determined by (some of) the inference rules according to which the expression is used.[3]

As chapter 2 argued, we use natural language expressions according to various syntactic inference rules. This principle claims that some such rules, in some so-far unspecified manner, fully determine the meanings of the logical constants. By talking about "(some of) the inference rules according to which the expression is used" I mean only to pick out some subset, possibly but not necessarily improper, of the inference rules we follow that involve the expression in question. What this comes to will be discussed in more detail in the next section.

I've phrased this principle in terms of a logical expression's meaning being "fully determined" by the inference rules according to which it is used. I did this mainly because the principle would sound a bit unwieldy if it said, "that a logical expression has the meaning it does is explained by the expression's being used according to such-and-such inference rules", but both statements come to the same thing. To say that *A* fully determines *B* is just to say that *A* fully and directly explains *B*. The central claim of Logical Inferentialism is that the rules of use for a logical expression explain what the expression means, directly, and without any explanatory contribution from other sources.

The principle implies that, as long as the inference rules of a language remain fixed, whatever else may change, the expression's meaning also remains fixed.[4] In other words, saying that rules "fully" determine the meanings of logical expressions tacitly denies that the rules (merely) partly determine the meaning of logical expressions. In other words, Logical Inferentialism stands opposed to metasemantic pictures of logic that allow inference rules *some* role in determining the meaning of logical constants while also bringing in further elements to complete the metasemantic story. So a metasemantics for logic saying that a logical constant's meaning is the most

metaphysically natural meaning that made valid a majority of the inference rules according to which we use the constant would not count as a version of logical inferentialism.[5]

To illustrate the central idea behind logical inferentialism, let's continue to focus on the logical expression "and". In English, "and" is plausibly used according to the following schematic rules of inference:

$$(\&I)\ \frac{\phi \quad \psi}{\phi \text{ and } \psi} \qquad (\&E)\ \frac{\phi \text{ and } \psi}{\phi} \quad \frac{\phi \text{ and } \psi}{\psi}$$

In these rules "***ϕ***" and "*ψ*" are schematic letters for which any declarative sentence of English may be substituted. Of course, English itself constantly changes and grows via the introduction of new vocabulary, so these schematic letters and, thus, these rules, should be understood in an open-ended fashion (2.IV). If these are the only meaning-determining rules for "and", then they must—according to inferentialists—fully determine the meaning of "and" in English. But in order to understand how this would work, we need to further flesh out the inferentialist picture.[6]

Before doing that though, it is important to ward off a potential misunderstanding. The principle of Logical Inferentialism does not entail anything like its converse. For inferentialists, meaning is determined by rules of use, not rules of use by meaning. It also doesn't entail that every single difference in the meaning-determining rules forces a difference in meaning. It is perfectly conceivable that very different rules can sometimes lead to the very same meanings, at least according to our ordinary, garden-variety concept of meaning.

Inferentialists are committed to using rules of inference for logical expressions to explain their meanings, but by itself this is far too general to give us any idea *how* such an explanation would go. What is truly distinctive of inferentialism is the idea that the rules we follow for an expression determine the expression's meaning by being automatically valid:

> **Meaning Validity Connection (MVC)**: In any language *L*, the meaning- determining inference rules for a logical expression are automatically valid in *L*.

Validity requires necessary truth-preservation in the strongest possible sense.[7] Some have preferred to take "validity" as a primitive notion, but even if taken as a primitive it must be understood in this strong fashion. In any case, validity requires ascribing truth and falsity to sentences. I don't intend this to mark a substantive disagreement with those who prefer to think of propositions or utterances or mental tokens as truthbearers (recall the discussion in 2.V). My claims can be made on any reasonable way of making truth and falsity assignments.[8]

Some philosophers have concerns about natural language notions of validity and truth related to the looming threat of paradox.[9] I would invite those who are actually worried about this to imagine their favored solution to liar-like paradoxes being assumed in the background. Any minimally adequate account must preserve our ability to make the kinds of generalizations I'm making here.[10]

Rules like (*&I*) and (*&E*) are schematic, but we can say that a schematic rule is valid just in case any instance of the schema is valid. I say "any" instance, rather than "all" instances, because schemas like this are open-ended (2.IV).[11] The "instances" that we are disposed to accept go beyond the sentences in our current lexicon. This will be relevant in section IV and will be of crucial importance in part III.

Sometimes I will say that the meaning-determining rules for an expression implicitly define that expression. I will also use the phrases meaning-determining rules and meaning-constituting rules interchangeably. Logical Inferentialism tells us that some of the rules used for a logical expression implicitly define the expression, and the MVC tells us that those meaning-determining rules are automatically valid, where "validity" requires necessary truth-preservation in the strongest sense. They are "automatically" valid because given that they are meaning determining, nothing more is required of them for validity. We can move directly from their meaning-determining status to their validity.

Combined with Logical Inferentialism, the MVC can be used to explain what logical expressions mean. The meanings of “and”, “not”, and other logical expressions in a given language are explained by the automatic validity of their meaning-determining rules. But taken by itself, the MVC is not enough. Inferentialists are also committed to the idea that—in some sense—for meaningful logical expressions, the meaning-determining rules determine all of the valid inferences in which the expression figures. Otherwise it would be possible that a language including all and only the principles of intuitionistic logic was a language in which all and only classically valid inferences were valid! That would violate the standard inferentialist assumption that the logical validities can be read off of the rules of a language.

This assumption was perhaps already implicit in Logical Inferentialism’s claim that the inference rules according to which an expression is used “fully determine” its meaning, but let us make it completely explicit in the form of the following principle:

> **Totality**: In any language *L*, if a logical inference involving a logical expression is valid in *L*, then its validity is fully determined by the automatic validity of the meaning-constituting rules for the expression.

Restricted to logic, something like Totality is assumed by all inferentialists, but it is often left implicit. Perhaps this is because it is close to being a byproduct of accepting Logical Inferentialism and the MVC. Still, the principle is important and powerful in its own right, so is worth treating as a separate component of inferentialism. Understanding Totality requires clarifying the notion of a “logical inference” and discussing two different ways that a logical inference’s validity can be “determined” by the automatic validity of meaning-constituting rules.[12]

Say that an inference rule is a *logical inference* just in case it is a schematic rule that contains only logical expressions non-schematically. Section III will discuss the question of which expressions are “logical”, but assuming as we have that “and” is a logical constant, the following is a logical inference:

$$(\&\&) \quad \frac{\phi \text{ and } \psi}{\psi \text{ and } \phi}$$

Totality demands that if a rule like (&&) is valid, its validity is determined by the automatic validity of the meaning-determining rules for "and". But the principle itself is completely neutral on which rules are meaning-determining rules for "and". It is even neutral on whether they are all and only schematic rules containing *only* "and" non-schematically. This and related issues will be discussed in depth in the following section.

The restriction to rules involving only logical expressions non-schematically helps to ensure that Totality doesn't lead to consequences that some would find implausible. Take the apparent fact that the following rule is valid, even though it contains no logical terminology:

$$(\textit{Gold}) \quad \frac{\alpha \text{ is gold}}{\alpha \text{ has atomic number 79}}$$

The natural kind term "gold" rigidly designates the substance with atomic number 79, so the rule (*Gold*) seems to be necessarily truth-preserving.[13] But it isn't obvious that this rule is valid in the sense of "validity" used in the MVC and Totality. *That* notion of validity required necessary truth-preservation in the strongest sense possible. I would call it "logical" validity, but I think that conceptual entailments and mathematical entailments are valid in the very same way. So let me instead call it strong validity. Strong validity may well be more demanding than truth-preservation in any metaphysically possible world. Perhaps it requires truth-preservation in every single scenario that we can coherently describe and reason about.[14]

Some think that we can coherently reason about scenarios in which (*Gold*) isn't truth-preserving, even if it might be more natural to describe such scenarios in other terms (perhaps metalinguistically).[15] Others think that a conventionalist picture of the necessary *a posteriori* can be developed along the lines of conventionalism about logic and mathematics.[16] I am sympathetic to both of these ideas, but

we can sidestep all of these issues here—Totality makes no ruling either way about the strong validity of (*Gold*) and related inferences. Even if (*Gold*) isn't an example of a non-logical but strongly valid rule, there clearly are such rules—mathematical rules for example (see part III). Some valid mathematical arguments even contain logical expressions like "and". From "the number of *F*s and *G*s = *n*" and "*n* > *k*" you can infer that "the number of *F*s and *G*s > *k*". But this argument, though strongly valid and containing "and", isn't a "logical inference" of the kind spoken of in Totality, so there is no demand that its validity is fully determined by the meaning-determining rules for "and".

Obviously, inconsistent rules would lead to every inference rule in the language being valid, at least in systems like classical logic in which inconsistency leads to triviality. But I don't think accommodating this requires any emendation of Totality. My reasons for thinking this will be explained briefly in section VII and then throughout chapter 5, so I won't go over them here.

There are at least two different ways for the strong validity of a logical inference *I* involving logical expression *c* to be determined by the automatic validity of the meaning-determining rules for *c*. The first is the trivial case, where *I* is one of the meaning-determining rules for *c*. Then by the MVC, *I* is *automatically* valid, so its validity can be given no deeper explanation. The other is for the validity of the meaning-determining rules for *c* to require the validity of *I* as well. In this case, the validity of the meaning-determining rules explains the validity of *I* without *I* itself being one of them. The next section will explain one way that this can happen.

In essence, the Totality principle enshrines another aspect of the connection between meaning and use. So understood, it is very easy to motivate. It encodes the key inferentialist assumption that the validity of the meaning-determining rules for a logical expression is the only source of the expression's meaning. Since nothing but our use determines the meaning of a logical expression, the validity of a logical inference must be witnessed in our use. Put this way, I suspect that many readers are nodding along. But Totality is quite strong. The strength of the MVC and Totality, as well as their connection to

famous metamathematical results like soundness and completeness, will be discussed in the next chapter (see 4.I and 4.III especially).

The stunning feature of logical inferentialism is that it reverses the traditional order of metasemantic explanation. That traditional order takes the meanings of expressions like “and” as given, and then uses those meanings to explain which rules of inference are valid for sentences containing “and”. Inferentialism reverses this by seeing meaning as determined by rules rather than vice versa. The basic inferentialist thought was first clearly endorsed by the Wittgenstein of *Philosophical Grammar* and the Carnap of *The Logical Syntax of Language*. Here’s Wittgenstein on negation:

> There cannot be a question whether these or other rules are the correct ones for the use of “not” (that is, whether they accord with its meaning). For without these rules the word has as yet no meaning; and if we change the rules, it now has another meaning (or none), and in that case we may just as well change the word too.[17]

And here is Carnap on the general reversal being pulled off:

> Up to now, in constructing a language, the procedure has usually been, first to 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 any rules of inference be chosen arbitrarily; then this choice, whatever it may be, will determine what meaning is to be assigned to the fundamental logical symbols.[18]

At around the same time that Wittgenstein and Carnap hit upon this philosophical idea, the great logician Gerhard Gentzen, inventor of both natural deduction systems and sequent calculi, was also toying with the idea that the natural deduction introduction rules for a connective (or operational rules in a sequent calculus) “define” the connective.[19] Shortly thereafter, philosophers like Karl Popper and William Kneale explicitly endorsed and began to develop early

versions of logical inferentialism.[20] Since that time many influential philosophers have been attracted to some form of inferentialism about logic, including Paul Boghossian, Michael Dummett, Ian Hacking, Christopher Peacocke, and Neil Tennant among many others.[21]

For explicitness, I'll call anyone who endorses Logical Inferentialism, understood according to the MVC and Totality, a logical inferentialist. Some so-called "logical inferentialists" in the literature have flirted with or even endorsed the idea that a radically new semantics is needed for the standard logical constants, perhaps one that eliminates all reference to semantic properties like truth and truth functions. This is unduly radical. The less-revisionary metasemantic approach that I've adopted is more in line with the original guiding insights of Wittgenstein, Carnap, and Gentzen. But even aside from its historical pedigree I think that inferentialism *should* be understood as a metasemantic view. In this book, logical inferentialism is a view about *how* logical expressions get their meanings, not a view about what the meanings of particular logical expressions are.

Given our assumption that "and" has its classical meaning, we can now see how logical inferentialists can explain this fact using the rules of inference for "and": by Logical Inferentialism and our previous assumption that the rules (*&I*) and (*&E*) are the inference rules according to which "and" is used, the meaning of "and" is determined by the rules (*&I*) and (*&E*). And by the MVC, these rules are valid, so by (*&I*), when ϕ and ψ are both true, then so is $\ulcorner \phi$ and $\psi \urcorner$, and by (*&E*), if either ϕ or ψ or both are false, then so too is $\ulcorner \phi$ and $\psi \urcorner$, else the (*&E*) rules wouldn't be valid. This means that "and" has its standard truth-functional meaning, and we have just fully explained this fact on inferentialist grounds. Unfortunately, the inferentialist story for the other connectives is more complicated. That story will be told in section IV, but some further philosophical groundwork is required first.

II Unrestricted Inferentialism

Logical inferentialism comes in many forms. The most important differences between versions of inferentialism concern which collections of rules can be used to determine a meaning for an expression. We have already (perhaps?) seen a difference on this point between Wittgenstein and Carnap illustrated in the quotes above. Wittgenstein suggested that some rules for "not" might fail to determine a meaning, while Carnap stated that meaning-determining rules could be "chosen arbitrarily". I implicitly allowed for this difference in the MVC by applying it only to meaning-determining rules. This leaves it open that some rules of use may fail to implicitly define a meaningful connective.

This allowed for a general characterization of inferentialism in section I, but I accept a Carnap-style unrestricted inferentialism. I have already defended the view in print, but I'll provide an independent development and defense here.[22] The distinctive unrestricted inferentialist claim is summed up in the following principle:

> **Meanings Are Cheap (MAC)**: Any collection of inference rules that can be used for an expression can (in principle) be meaning determining for the expression.[23]

This principle needs to be clarified in a number of ways. At its heart, it simply expresses the very natural idea that meanings (or concepts) are cheap. We can't bring unicorns into existence with our thoughts and practices, but the *concept* of unicorn is another matter entirely. The MAC principle tells us that there are no *metaphysical* constraints (so to speak) on our linguistic or conceptual toolbox.

The Logical Inferentialism principle talks about the "inference rules according to which an expression is used" and the MAC principle similarly talks about "inference rules that can be used for an expression". These phrases require some unpacking. Basically, an inference rule *R* is used for expression *c* just in case *c* occurs in *R* (when *R* is schematic, this requires that *c* occurs in all instances of *R*). When *R* is schematic, we can also say that *R* is indirectly for *c* when some but not all of its instances are directly for *c*, in the just-

given sense. We can also extend this notion to collections of rules, for both the direct and indirect senses.

This means that the schematic rules (*&I*) and (*&E*) are also used for "not", indirectly, since many conjunctions in our language contain "not"—"I went to the movies and Rudolf did not". Because of this, the possibility is open that the schematic (*&I*) and (*&E*) rules are also (partly) meaning-constituting for "not", in addition to "and". There is nothing mysterious about this possibility, it is just expressing the obvious point that there will be valid *particular* inferences involving "not" that are instance of these rules. So following these rules is part of how we use "not", so they may well be relevant to what "not" means. This will be of more central importance later in this section and in chapter 5. At present though, the central import of this way of formulating MAC comes from the other direction—the MAC principle is not so broad as to allow the possibility that the castling rule in chess can be used to determine a meaning for "not". That rule is not even indirectly for "not", so it has nothing to do with its use.

I said that MAC is motivated by a denial of metaphysical constraints on our conceptual toolkit, but that doesn't mean that there aren't any constraints at all. Obviously there are *practical* constraints on which rules we can actually follow. The practical impossibility of following a rule can come about in a number of ways. One possibility is that the rule would require us to have non-algorithmic cognitive powers in order to follow it. Following such rules is incompatible with naturalism, according to 1.IV. Another possibility is that a conflict between two rules could make it conceptually impossible to follow both of them at the same time. This will happen when the dispositions that constitute following one rule can't be had jointly with the dispositions that constitute following the other rule. This is analogous to the impossibility of following a collection of chess-like rules that allow castling while also, at the same time, following a collection of chess-like rules that don't allow castling. We could play these different games in succession, or could make mistakes when trying to follow one collection of rules, but we couldn't be really and truly following both rules at the very same time. This is an important point, the relevance of which will be discussed in 5.I and 5.II.

Since the MAC principle concerns only rules that *can* be used for an expression, for the moment we can limit our attention to collections of rules that language users could, in principle, follow. This isn't really a restriction in this context though—if some possible beings can follow the rule *R* for logical expression *c*, then we can allow *R* to be meaning constituting for *c* and automatically valid in their language by the MVC.

We still face the question of which rules in a language an inferentialist should take as the meaning-constituting rules. When trying to answer this question, it is important to keep sight of the distinction between rules that are directly versus indirectly for a given logical expression. Let's turn back to the example of "and". Section I claimed that the rules (*&I*) and (*&E*) are plausibly the meaning-constituting rules for "and" in English. But as noted in the discussion of Totality, these aren't the only valid rules for "and" in English. Consider again the following rule:

$$(\&\&) \quad \frac{\phi \text{ and } \psi}{\psi \text{ and } \phi}$$

This rule, expressing the order symmetry of the "and" connective, is clearly valid. Is it also a meaning-constituting rule for "and"? Different answers to this and similar questions lead to an important division among inferentialists.

On one hand, *holist* inferentialists will see any valid rule for a connective, including (&&) for "and", as meaning-constituting for the connective. According to the holist, the meaning-constituting rules for "and" include all of the rules we may use for "and", including, in addition to (*&I*) and (*&E*), the rule (&&) and countless other rules as well. Holist inferentialists still accept Totality, but think that all of the valid logical inferences are meaning-constituting and thus automatically valid. By contrast, *non-holist* inferentialists think that some but not all valid rules for a connective are meaning constituting. Non-holist inferentialists typically think that (*&I*) and (*&E*) are meaning constituting for "and" but that (&&) is not.

I'm not a holist, in this wide sense. But by allowing implicitly defining rules that are only indirectly for an expression I opened the

door to a narrower form of holism. According to this narrow holism, a logical expression like “not” has its meaning partly determined by schematic rules that are only indirectly “for” the expression. This contrasts sharply with the just-explained wide holism, which takes every valid schematic rule that is directly for an expression to be meaning constituting for it. These two kinds of “holism” cut across each other. It is possible to be a holist in both the narrow and the wide sense, but it is not required. They cut across each other because narrow “holism” is an ersatz form of holism.

Many philosophers have found wide holism—that is, legitimate holism—deeply problematic.[24] But I don’t think that these worries carry over to narrow-holism. However, narrow holist (and non-holist) inferentialists face a challenge that wide holists do not. The wide holist paints with a broad brush, and sees every rule directly involving a connective as meaning-constituting for it. By contrast, those of us who distinguish the meaning-constituting rules from the non-meaning-constituting rules must make this distinction in a non–ad hoc and philosophically satisfying manner.

Happily, this can be done straightforwardly in a well-motivated way. There is a very natural distinction between inference rules that are followed directly and those that are only followed indirectly, as a byproduct of directly following other rules. Let’s call the directly accepted rules basic rules and the indirectly accepted rules derivative rules. Very roughly, the basic rules are those that speakers accept without any epistemic reason for doing so, while the derivative rules are those that they accept because they accept other rules that allow them to indirectly infer the rule’s conclusion from the rule’s premises.[25] We might say that speakers accept the basic rules simply *because they do*, since they don’t try to epistemically justify their acceptance on some other grounds. Causally speaking, as discussed in 2.III, speakers endorse instances of a basic rule like modus ponens because they recognize it as an instance of modus ponens, formally. But their endorsement of modus ponens itself is not something that rests on some special evidence. The basic rules are instead the foundational moves in our language game.

The dispositionalist account of rule-following from chapter 2 allows us to make the distinction between basic and derivative rules quite

clearly. Following an inference rule required, among other things, a disposition to infer according to the rule if given a chance, if certain other conditions were also in place (see 2.VIII and 2.IX). These dispositions to infer can be be simple, exercised directly without intermediate actions. Or they can be complex (or even composite), exercised, if at all, only by way of intermediate actions. This crucial distinction was introduced to deal with Kripkenstein's problem of error, but it is also applicable here.[26]

For some rules but not others we are disposed to infer according to them only by way of being disposed to take various intermediate steps. These rules are derivative. Rules we infer according to directly, when we do, are basic.[27] Quite plausibly rules like (*&I*) and (*&E*) are followed by all speakers who use the word "and" without any intermediate steps, actions, or activities involved. By contrast, a rule like (&&) is initially followed only by way of following more basic rules.

In the kinds of cases we're considering, what you can prove with the derivative rules can also be proved without them. For the rule (&&) we can see this by way of the following proof:

$$\frac{\begin{array}{cc}\phi \text{ and } \psi & \phi \text{ and } \psi \\ \psi & \phi\end{array}}{\psi \text{ and } \phi}$$

This proof uses two instances of (*&E*) followed by an instance of (*&I*) to establish (&&) as a derived rule (it also tacitly assumes that provability is transitive—this assumption will be discussed at the end of this section). As discussed in chapter 2, formal natural deduction proofs like this model inferences that are sanctioned by our natural language inference rules. Section IV will say more about the relationship between formal models and our inferential practices (see also the next chapter). The proof here *shows* that following (*&I*) and (*&E*) as basic rules also requires following (&&) as a derivative rule.

Of course, the details of this particular example aren't very important. What matters is that there is a useful and non–ad hoc distinction between the basic and derivative rules for an expression in terms of the behavioral/psychological factors underlying inference

rule-following. It is also not important that (&&) is a very simple derivative rule, so simple that some may even argue that it is itself a basic rule. Those who think that (&&) is a basic rule are free to replace the example with a more complicated rule involving “and”, since the very same points will apply.

It is clear that we should also count as “derivative” rules that are followed directly only *after* something like an argument or a proof from the basic rules has been found or at least shown to exist. This is probably the best way to account for various highly complex rules that we follow (it also neatly matches mathematical practice, with its constant appeal to theorems and lemmas), but it works even for simple rules. Once speakers realize that a rule like (&&) can be proved using (*&I*) and (*&E*), they can then start skipping the intermediate steps when applying (&&). But even when this happens (&&) is better seen as derivative because the speakers only started following it directly after they became aware that it could be proved from the basic rules of the language. In this way, the behavioral/psychological distinction between basic and derivative rules gives rise to a related but broader quasi-historical distinction.

Of course, which rules in a real natural language are actually basic will likely vary somewhat from speaker to speaker. But I suspect that there will be a great deal of overlap between any two speakers of the same language. Perhaps most importantly, the distinction between rules that are simply followed and rules that are followed only by way of various intermediate steps is a real one reflected in language use. It is not merely projected upon natural language by the logician’s codification.[28] The distinction is realized in the different kinds of dispositions that underwrite inference rule-following, for particular rules. For our purposes here, we can assume that there can be no deep conflict between complex and composite dispositions.[29] If a composite disposition conflicts with a complex disposition, one or the other must go, so at most one can be stable in the sense required for rule-following (see 2.VII). The dispositions to infer according to the basic rules are simple, but not so for derivative rules.

The distinction between basic and derivative rules is not itself a directly epistemological distinction. But in chapter 6 I will show that it does give rise to an important epistemological distinction. Even so it

is not itself normative, but is instead psychological-behavioral-historical. The initial distinction is psychological-behavioral; it concerns the way in which someone moves from the premises of a rule to the rule's conclusion when following the rule. Among the rules that are "basic" in this sense are some that are now followed directly but only as shortcuts. For the purposes of meaning constitution I count as *basic* only those rules that were always followed directly by speakers of the language, even before any proof of the rule's validity was found. The pseudo-basic rules are those that seem basic but would be given up if the earlier proof of their validity was shown to be faulty.

With this distinction in hand, we can give a very plausible account of meaning-constituting rules. In any language the meaning-constituting rules for a logical expression are the basic inference rules that are *for* that expression, either directly or indirectly. The basic rules are the primitive facts of use that constitute a language.

Why include rules that are only indirectly "for" an expression among its meaning-constituting rules? Because we want to include *all* of the primitive moves that involve the expression. So the meaning of "not" (for example) is *partly* based on the instances of other basic rules that include "not". This is narrow (not wide) holism, and there is nothing mysterious about narrow holism when viewed in this way. I will be understanding the various inferentialist principles like the MVC and MAC so that only the basic rules *for* an expression (in the extended sense) are meaning-constituting for the expression.[30]

With this clarification made, we can now see that Totality claims that only the logical inferences that are basic or derived rules in a language are strongly valid.[31] The basic rules are followed directly, the derivative rules only indirectly, yet both are followed. The validity of the indirectly followed derivative rules is itself a trivial byproduct of the automatic validity of the basic rules. But since the validity of the derivative rules is forced by the validity of the basic rules, we are free to take the basic rules for—either directly or indirectly—a given logical expression, as the only meaning-determining rules for the expression. The validity of a derivative rule in a language is explained by the validity of some subset of the basic rules of the language (7.IV will

include more discussion of the relationship between basic and derived rules).

One other relevant point that is worth addressing concerns the role of structural rules in the proof of (&&) from above. Structural rules can be put in purely schematic terms using no object language vocabulary. In natural deduction systems they are typically left implicit, but they are made fully explicit in a sequent calculus. If we express validity using the semantic turnstile ("$\models$"), the semantic versions of the standard structural rules can be stated as follows (with Γ and Δ sets of sentences, ϕ, ψ, and χ single sentences):

ASSUMPTION/REFLECTION: $\phi \models \phi$
THINNING (LEFT): If $\Gamma \models \phi$ then $\Gamma, \psi \models \phi$
INTERCHANGE: If $\phi, \psi \models \chi$ then $\psi, \phi \models \chi$
CONTRACTION: If $\phi, \phi \models \psi$ then $\phi \models \psi$
TRANSITIVITY/CUT: if $\Gamma \models \phi$ and $\Delta, \phi \models \psi$ then $\Gamma, \Delta \models \psi$

As the name suggests, the structural rules concern the structure of proofs in a formal system. In terms of natural language inference, they reflect very general properties of our inferential practices. Nothing in the "logical" inferentialist principles suffices to explain their validity. But if we apply inferentialist insights more generally, then the validity of each of these rules and meta-rules can be explained in terms of language use (this is true even without requiring unrestricted inferentialism).

This is because the (strong) validity of these rules is guaranteed if they are all basic rules we "follow". In fact, all of these rules are key to our inferential practices, so it seems clear that we "follow" these schematic rules. To argue from our following these rules to their validity requires generalizing the MVC and Totality so that they hold for all of our basic rules, even these entirely schematic rules. The general version of the MVC is straightforward. The general version of Totality claims that rules that we don't "follow" are not strongly valid. With these general principles in hand, the simplest way to argue for the *semantic* versions of the structural rules is by assuming the *syntactic* versions of the structural rules. This is perfectly fine, but I

think it's also instructive to see how we might argue for these rules directly, by appeal to the facts about rule-following and inference:

- ASSUMPTION: Presumably, we trivially follow the rule with ϕ as the conclusion and lone premise (see the mention of limiting cases of inference in 2.VI), and so, by the generalized MVC, the rule is valid.
- THINNING (LEFT): Assume that $\Gamma \models \phi$, then by generalized Totality we follow the rule with premises Γ and conclusion ϕ, but then we must also follow the rule with added premise ψ, so by the generalized MVC, $\Gamma, \psi \models \phi$.
- INTERCHANGE: Assume that $\phi, \psi \models \chi$, then by generalized Totality, we follow the rule with premises ϕ,ψ and conclusion χ, and this is the same thing as following the rule with premises ψ,ϕ and conclusion χ, so by the generalized MVC, $\psi, \phi \models \chi$.
- CONTRACTION: Assume $\phi, \phi \models \psi$, then by generalized Totality, we follow the rule with premises ϕ,ϕ and conclusion ψ, but this is the same thing as following the rule with premise ϕ and conclusion ψ, so by the generalized $\mathbf{MVC}, \phi \models \psi$.
- TRANSITIVITY/CUT: Assume both $\Gamma \models \phi$ and $\Delta, \phi \models \psi$, then by generalized Totality, we follow both the rule with premises Γ and conclusion ϕ and the rule with premises Δ,ϕ and conclusion ψ, and so we also follow the rule with premises Γ,Δ and conclusion ψ, and so by the generalized MVC, $\Gamma, \Delta \models \psi$.

Let me stress again that these brief arguments are not appealing to the MVC or Totality as stated in section I. Instead they are appealing to generalized versions of these principles that I have not defended. Here I'm just giving a sense of why metasemantic inferentialists won't think these structural validities fall completely outside of the overall inferentialist program.

The structural rules encode aspects of our practice of inference rule-following. Proofs in a formal natural deduction system reflect similar aspects of our inferential practices. As such, it is not surprising that all of my arguments for structural validities make a crucial appeal

to the nature of inference and inference rule-following. We trivially follow ASSUMPTION because inferences can be trivial. We follow THINNING because inference and inference rule-following require general dispositions that abstract away from other aspects of a rule-follower's mental state. This means that, if adding an extra premise blocked us from following the righthand rule, then we couldn't have really been following the lefthand rule. We follow INTERCHANGE AND CONTRACTION because, in terms of inference rule-following, there is no difference at all between following the lefthand rule and following the righthand rule. And we follow TRANSITIVITY because the dispositions required for inference rule-following could be either simple, complex, or composite (2.VII, 2.VIII, and 2.IX). Having the dispositions that constitute following both lefthand rules constitutes having a (more complex) disposition to follow the righthand rule.[32] Since we follow all of these rules and meta-rules, then, by generalized inferentialist principles, they are all valid. So I don't think that use of structural rules requires anything that conflicts with the spirit of inferentialism.[33]

The structural rules and the basic rules lead to the derivative rules. The inferentialist story applies to the basic rules of a language. This is intuitively plausible, since the basic rules are the rules that are followed by language users not because they have found some proof of the rule's validity or have been given some other reason for using the rule, but simply because they have *decided* (perhaps implicitly) to use sounds and marks in such and such a way. The basic rules are, in essence, the foundational conventions of a language. They are the regularities that sit at the heart of a Lewis-style account of the social conventions of language (2.I). And in a manner of speaking, they are the only conventions of the language.

Let's now say that any inferentialist who endorses MAC is an unrestricted logical inferentialist. Unrestricted inferentialists endorse all of Logical Inferentialism, the MVC, MAC, and Totality. It is "unrestricted" because it puts no substantive constraints on the meaning-determining rules of a language. In effect, the rules of a language are entirely self-justifying. The basic rules of use in any language, whatever they are, are *automatically valid*. Unrestricted inferentialism is an incredibly simple, neat, and tidy theory of the meanings of logical notions. I think a version of the theory was tacitly

endorsed by Carnap, Ayer, and many other early logical conventionalists, at least some of the time. Despite this, as far as I know, I am the view's only current defender. In a sense, I'm the view's only *defender*, ever, since the positivists didn't really defend the view, they just assumed it to be true because it was so intuitively plausible and theoretically attractive. I'll be indirectly defending unrestricted inferentialism throughout the book by developing theories of logic and mathematics on its basis, but my direct defense of unrestricted inferentialism against restricted inferentialism will come in section V, and then at greater length in chapter 5.[34]

Before that, I'm going to show how the unrestricted inferentialist theory I have developed gives rise to a satisfying theory of the classical logical constants. First discussing which expressions are "logical" (section III) and then showing that inferentialism explains how logical expressions manage to get their meanings (section IV).

III Which Expressions Are "Logical"?

Thus far I've characterized the logical constants only by appealing to a standard list ("and", "either … or", "if … then", "not", "all", …) while saying nothing at all about what unifies the expressions on the list and makes them worth theorizing about. There is nothing wrong with this procedure, as such—what are you interested in? This, that, and the other—but any comprehensive theory of logic should be able to say something about what it is that makes certain expressions but not others "logical". I'll do so here.

One feature shared by all of the logical expressions is that they are *non-empirical*. As noted briefly at the start of section I, learning an expression like "dog" isn't only a matter of learning standard inferences involving the expression, it also requires an ability to reliably recognize dogs of various kinds as objects in the world.[35] By contrast, some expressions, including the logical expressions, require for their mastery only the use of certain pure inference rules involving whole sentences. Mastery of such expressions doesn't require interactions with the world, as mastery of "dog" does: call such expressions non-empirical.[36]

This rough account of empirical expressions is less than ideal. Imagine that the predicate "*F*" is a paradigmatically empirical term, and imagine that we introduce "*G*" as a synonym for "*F*", so that "*F*" and "*G*" are terms analogous to "cat" and "feline". In this case, "*G*" would be non-empirical and "*F*" would be empirical according to a strict reading of the criterion for empiricality just given. Intuitively though, the expressions are synonyms. One response would be to beef up the account of "non-empirical" so as to exclude the possibility that an empirical expression, of the same grammatical type, is inter-derivable with any non-empirical expression according to the rules of our language.[37] But even this procedure doesn't do away with every intuitive counterexample. If a "b-cat" is a cat that is big, "b-cat" will be non-empirical according to the amended criterion. Again, this seems counterintuitive; after a little bit of practice the term could be applied directly to an item in the world.

To provide an inferentialist response to this, let's assume for the moment that all expressions have their meanings determined by inference rules. For this to be at all plausible, the notion of an inference rule needs to be broadened a bit. One way of broadening it allows rules that sanction accepting or rejecting a particular sentence when the world is in a certain state, or vice versa.[38] Even leaving the details vague, against this background we can give an inferentialist characterization of "non-empirical" expressions that avoids the problems from above.

Say that an expression is deeply non-empirical just in case its meaning-constituting rules don't connect to the world in either direction. This means that they don't trace back to the world, and don't lead forward to the world. In other words, non-empirical expressions can't be defined, or even partially defined, in terms of empirical terms *even ancestrally.* Nor do they have any empirical consequences even ancestrally. Roughly, we can imagine going through an expression's meaning-determining rules and tracing them backward and forward using the meaning-determining rules for any expression used at any given step. If the iterated application of this procedure always leads to the world (so to speak) after some finite number of steps, then the initial term is empirical. Otherwise, it is deeply non-empirical.

Why must the procedure “always” lead to the world for the expression to be empirical? Because when meaning-determining rules are schematic, they will have many instances involving empirical sentences, predicates, and names that trace back to the world. We don’t want this to force us to mischaracterize schematic logical rules as empirical, so we require that the procedure always leads back to the world. In impoverished languages we might also require that the procedure always traces back to the same *state of the world*, or the same type of state—surely, introducing a connective like “and” into a language with only empirical sentences doesn’t thereby make “and” an empirical notion. There are issues with this amended approach that I won’t pursue here.

The suggested procedure leaves us a rough but intuitive picture. Some expressions are implicitly defined by rules that connect directly to the world, and these are empirical expressions. Expressions whose implicitly defining rules explicitly include empirical expressions are also empirical, and so on and so forth. The deeply non-empirical expressions are those whose implicitly defining rules do not directly involve any empirical expressions (to again use the distinction between two different ways a rule can be “for” an expression, introduced in the previous section).

A complete account of the application of this procedure would require a complete inferentialist metasemantics for the declarative portion of natural language. But even absent a complete account, we can see that a logical expression like “and” will be deeply non-empirical. This is true even if “and” is partly implicitly defined by schematic rules for “not”. None of the canonical logical constants are defined in terms of empirical expressions, nor in terms of expressions so defined. All intuitively “logical” expressions are deeply non-empirical, but are all non-empirical expressions intuitively “logical”? Unfortunately not. Take a mathematical abbreviation, for example. If a number is a “4-prime” if it it is either 4 or a prime number, “4-prime” will be deeply non-empirical but it is not a “logical” expression, at least not in a strict sense.

To characterize the logical notions without including mathematical notions as well, we need more than deep non-empiricality. What else should be required? The natural candidate is another condition with a

distinguished history in discussions of logical constants: *topic neutrality*. The logical constants are topic neutral in that they can be used whether the topic is stoicism, sex, or stamp collecting. In the literature it is common to spell out the notion of topic neutrality using model-theoretic criteria of permutation invariance or some related feature, but I won't be attempting this.[39] In fact, I don't think any absolute criterion of this kind is likely to perfectly distinguish the intuitively logical from the intuitively non-logical, since it seems likely to me that topic neutrality comes in degrees.

The sentential connectives are topic neutral in the sense that their rules have instances in every topic involving declarative, truth-apt sentences. But the quantifiers require a topic in which objects of some kind are discussed. This narrows things somewhat, but not too much. The quantifier rules are still extremely general. By contrast, a term like "4-prime" has a very specific topic: the natural numbers. There are also many borderline cases—should deontic operators like *ought* count as logical? What about operators expressing alethic modalities? I don't think we should expect there to be a clear cut answer in every case. But even allowing that topic neutrality comes in degrees we can ask how it should be understood if not in terms of some model-theoretic criterion of permutation invariance, however refined.

For a global inferentialist, the natural approach again appeals to the meaning-constituting rules for an expression. If those rules have applications in discourse of all kinds, they count as topic neutral. The wider the compass of these implicitly-defining rules, the more topic neutrality an expression possesses. Non-empirical expressions whose rules have a sufficiently high degree of topic neutrality deserve the honorific "logical"; others do not. But I don't think it is all that helpful to try to find a precise cutoff point for the degree of topic neutrality a non-empirical expression's rules must possess in order for it to count as "logical". We shouldn't expect perfect determinacy in a matter like this, but this shouldn't seriously trouble us. Nor should the context-sensitivity of "sufficiently high degree".

The familiar terms "and", "either…or", "if…then", "not", "all", … , *et cetera*, all count as *logical* notions because they are implicitly defined by deeply non-empirical rules that are highly topic neutral. This is

enough to make them an important and theoretically interesting class of expressions for philosophers to focus on, whether or not they form a natural kind carved into nature by the gods.

To sum this discussion up, inferentialists can characterize the "logical" constants in the following way:

> **Logicality**: An expression is a logical constant just in case its meaning-constituting rules are both (*i*) deeply non-empirical and (*ii*) highly topic neutral.

Can we jettison (*i*) and stick simply with (*ii*)? We can if topic neutrality entails deep non-empiricality. It does seem prima facie plausible that empirical terms are not topic neutral. But on reflection it isn't obviously the case. Color terms, for example, might be both empirical and topic neutral to a quite high degree. Of course, color terms aren't maximally topic neutral—numbers and other abstract objects have no color—but they are still applicable very, very widely. In any case, I think inferentialists can roughly single out the canonical logical constants on purely inferentialist grounds using Logicality.[40]

In addition to the points already made, one further benefit of this approach is that it allows us some flexibility when it comes to novel and borderline cases of putative "logical" expressions. The vagueness here might worry some readers, but it doesn't trouble me at all. I think that any more precision on these matters is as likely to distort as to illuminate. This section's inferentialist account of logicality is sufficient for my purposes, but there are still a couple of points that need to be clarified.

Some may worry about inferentialism's ability to distinguish a canonical logical constant like "and" from a truth-functionally equivalent but distinct connective like "but". I am not worried about this, but something should be said for those that are. A full account would require a completely worked out global version of metasemantic inferentialism, but I think the broad outlines of an inferentialist response are visible even at this level of detail. No matter exactly how it is done, global inferentialism should be able to distinguish "and" and "but" by drawing a distinction between canonically *semantic* and canonically *pragmatic* features of use.

Whatever the precise details, this will be best done by appealing to further rule-governed aspects of language use.[41]

The rules of use—that is, the linguistic conventions—concerning "but" differ from those concerning "and", in that "but" requires, for correct usage, that there is a pragmatic, contextually salient contrast between the embedded sentences. I don't think that "pragmatic" rules of this kind can plausibly be taken as basic, since they will usually involve context-sensitive assessments of expectations and communicative context. And some of them probably don't even rise to the level of "rules" at all, being mere regularities that we are disposed to conform to only in a weak and easily canceled sense. This is also why utterances of "it is what it is" or "it is and it isn't" aren't problematic counterexamples to our rules governing logical truths and falsehoods. Utterances of these sentences usually express or communicate something owing to pragmatic features of language use. If taken literally by some audience a speaker could and would quickly clear things up. So these points aren't a problem for my account of logicality, nor for inferentialism generally. No more than irony is a problem for general theories of meaning.

My approach to logicality also cleanly distinguishes between canonical logical expression and the non-logical expressions that prominently figure in conceptual or analytic truths, more broadly. To illustrate: it is extremely natural to think that many common non-logical words, including familiar terms like "bachelor", are implicitly defined by basic rules in a manner analogous to how logical expressions like "and" are implicitly defined:

$$(BI)\ \frac{\alpha \text{ is unmarried} \quad \alpha \text{ is male}}{\alpha \text{ is a bachelor}} \qquad (BE)\ \frac{\alpha \text{ is a bachelor}}{\alpha \text{ is unmarried}} \quad \frac{\alpha \text{ is a bachelor}}{\alpha \text{ is male}}$$

Despite this, the term "bachelor" isn't deeply non-empirical, since its meaning-determining rules directly involve other terms ("unmarried" and "male") that themselves trace back to the world. Nor is the term "bachelor" highly topic neutral, since it has a particular subject matter related to human social conventions and traditions. So it is overdetermined that "bachelor" is not a logical constant. Yet "bachelor" stills figures in truths that have the same trivial character,

and the same conventional source, as the logical truths—"all bachelors are unmarried", for example.[42] These so-called *analytic* or *conceptual truths* are like the logical truths in being necessary, *a priori*, and trivial. While not all conceptual truths are logical truths, I do think that all logical truths are conceptual (and analytic) truths, in this sense. Conceptual truths will be discussed a bit again in part IV, in 13.I, where I discuss and assess conventionalist slogans.

Assuming classical logic, it is a logical truth that any expression of our language is either logical or non-logical. But this doesn't mean that every expression of our language is either logical *or factual*, logical *or descriptive*, and so on. Some may think that these alternative notions extensionally coincide with the non-logical expressions, but it is not at all obvious that this is so. I think that mathematical expressions are both non-logical and non-descriptive (see 11.I for discussion). And it is also worth keeping in mind that, as we just saw with "bachelor", empirical or factual terms can figure in *truths* that are themselves non-empirical.

At present, I will continue to focus on the canonical logical constants. These expressions are neither magical nor metaphysically special, but in being both deeply non-empirical and highly topic neutral, they comprise an interesting, important, and fairly well-demarcated part of our conceptual scheme, well worth serious philosophical investigation.

IV The Meanings of the Logical Constants

At the end of section I, I showed that inferentialists could explain how "and" means what classical logicians assume that it means by using the rules of (*&I*) and (*&E*) and the metasemantic assumptions of logical inferentialism. Here I will provide similar accounts of the other standard constants of sentential and first-order logic.

The ability to account for the meanings of logical constants in terms of inference rules is philosophically important in part because it provides a wholly naturalistic metasemantics for logical expressions. The possibility of doing this is a central reason for logical inferentialism's popularity in philosophy.[43] Oddly though, it is far less

well-known than it should be that if we want to provide an inferentialist theory that will deliver classical meanings for all of the canonical sentential connectives then we need to go beyond the familiar acceptance-based rules of standard natural deductions systems. This was first shown by Rudolf Carnap in the 1940s.[44]

Though my concern is with English logical constants ("and", "not", "all", and so on), in order to approach this matter more clearly I'm going to work with a formal language meant to accurately model the relevant aspects of natural languages. This is entirely standard. The only wrinkle is that the formal models I will work with here are what I call *full languages*—a full language is a standard formal language together with a natural deduction proof system for the language. These are sometimes called "systems" in the literature, but since that word is used in many different ways, I prefer not to use it.

The package of natural deduction rules for a full language together give rise to a relationship of provability (or derivability) of a sentence in the language from other sentences in the language. The exact formal definition of a "proof" will vary somewhat based on the formalism used.[45] In the two-dimensional layout I've been using, proofs will be finite trees containing the conclusion sentence at the root, the premises at the terminal nodes, and proceeding from node to node using only the rules of the system. In linear systems proofs will be finite sequences of sentences (or sequences of ordered *n*-tuples—sentences together with various items used for bookkeeping) where each item in the sequence is either a premise (possibly discharged) or arises from earlier items in the sequence using one of the rules of the system.

Usually, formal treatments work with schematic surrogates for object-level proofs, and occasionally proof theory texts reserve the word "proof" for the meta-level objects, reserving the word "derivation" for what I have been calling "proofs".[46] Here I will use the word "proof" when talking about both proper proofs and also proof schemas. I will also use the syntactic turnstile ("⊢") to express facts about provability. My procedure will be to work with a full language that is assumed to accurately represent a significant fragment of English and show that inferentialist principles force each logical constant to have its standard meaning in this full language.

Let's start with sentential logic. Assume that we're working in a language *L* for sentential logic. A valuation *v* is an assignment of truth values (*T* or *F*) to all sentences in the language; call a valuation *Boolean* if it respects the standard truth functions assigned to the canonical sentential connectives ($\neg,\wedge,\vee,\rightarrow,\leftrightarrow$). Consider a standard collection of single-conclusion natural deduction rules *R* for the sentential connectives together with the standard structural rules of deducibility *S* (see the end of section II) and say that a valuation *v* is *inadmissible* if it makes one of the rules in $R \cup S$ non-truth-preserving. In other words, *v* is inadmissible if there is some instance of one of these rules where *v* assigns every premise *T* but the conclusion *F* and *admissible* otherwise. We want our collection of inference rules $R \cup S$ to ensure that all admissible valuations are Boolean by ruling non-Boolean valuations inadmissible. Metasemantically, this involves trying to use the inferentialist principles from sections I and II, applied to single-conclusion standard natural deduction rules, to determine the standard classical meanings for the sentential connectives.

In the 1940s Carnap showed that this cannot be done:

Theorem. *(Carnap 1943) For any collection of standard single-conclusion natural deduction rules R and structural rules S, there is at least one non-Boolean valuation that is admissible relative to* $R \cup S$.

Proof. Consider a valuation *v* that assigns *T* to every sentence in *L*, clearly *v* will make all of our rules truth-preserving, since there is no possibility of all premises of a rule being assigned *T* while the conclusion of the rule gets assigned *F*, since *v* does not assign *F* to any sentence in the language; so *v* is admissible relative to $R \cup S$. However, *v* is not Boolean, since for any sentence ϕ in *L*, $v(\phi) = v(\neg\phi) = T$, so negation will not receive its standard truth functional meaning.

—

Things are worse than this, since even if we bar trivial valuations that assign all sentences *T*, we won't be able to use standard natural deduction rules to force negation, disjunction, or the conditional to be truth-functional—the truth values will be unconstrained when all component sentences are assigned *F*, for example.[47]

The theorem and these associated results show that the inferentialist idea of using some standard set of rules (and only the rules in said set) to force the standard meanings for the connectives

doesn't work. If these rules can't show non-Boolean valuations to be inadmissible, then they can't plausibly explain why each connective has its standard truth-functional meaning. In fact, among the canonical sentential logical constants, standard single-conclusion rules force the proper meaning only for conjunction, so my example above was deftly chosen.

Some inferentialists, including Carnap himself, solved the problem by (in effect) moving to a sequent calculus formulation of our logic and allowing for multiple conclusions in addition to single conclusions.[48] But most inferentialists have felt that multiple-conclusion systems don't correspond to our actual reasoning practices.[49] Happily, if we move to bilateralist inference rules, as suggested in 2.IV, we can provide a simple inferentialist account of the classical connectives.[50] As discussed before, this is not an ad hoc move introduced to overcome a philosophical obstacle; instead it is independently motivated by the features of natural language reasoning that inferentialists are attempting to model.[51]

I believe that this was first noted by Timothy Smiley; the approach was then further developed by Ian Rumfitt. Given that bilateralist rules are needed to model natural language reasoning, it shouldn't be that surprising to find that by modeling only a proper portion of our inferential behavior with terms like "or" and "not" we aren't able to adequately explain what they mean. In 2.IV I noted that we can develop a bilateralist natural deduction system using the force indicators "+" and "-" where the plus is an acceptance indicator, the minus is a rejection indicator.[52] Once again though, it must be kept in mind that despite how they look, these force indicators are not logical operators or connectives and as such they do not embed, so a string like ⌜ +ϕ ∨-ψ ⌝ is not any part of our language. I mentioned some of the philosophical advantages of bilateralism in chapter 2, but perhaps its most important advantage is the ability to cleanly solve Carnap's problem.

Instead of formulating natural deduction rules using sentences, we'll be using *signed sentences*, which are our familiar sentences of sentential logic prefixed with one of our force indicators. Our formal system will have all of the standard structural rules, together with one new structural rule. Let Γ be a set of signed sentences and α and β

be variables for signed sentences and let α^* be just like α except with the opposite sign (so if $\alpha = +\phi$ then $\alpha^* = -\phi$). Our new structural rule is structural reductio:

S-REDUCTIO: If $\Gamma,\alpha \vdash \beta$ and $\Gamma,\alpha \vdash \beta^*$, then $\Gamma \vdash \alpha^*$

This rule is so-named because it is something like a structural version of the familiar reductio rule for negation. But it is important to keep in mind that this is a structural rule, not a negation rule. At the end of section II I gav inferentialist arguments for the semantic validity of all of the standard structural rules. The semantic version of *S*-reductio can be established similarly. Before doing this though, the more general notion of validity for inferences involving signed sentences (signed inferences) must be explained.

A signed inference is *valid* if and only if every valuation which makes all plus-signed premises true and all minus-signed premises false also makes the conclusion true if it is plus-signed but false otherwise. Philosophically I have already noted that the inferentialist motivation for the MVC continues to apply to this slightly expanded notion of validity. And it is trivial to extend the MVC so that it covers this slightly more expansive notion of validity. When this is done, the semantic version of the structural rule of *S*-reductio can be established on inferentialist grounds.

As before, the basic idea is to imagine having in hand *general* versions of our inferentialist principles like Totality and the MVC. I have neither argued for nor clearly formulated these here, but they would be justified by the same kinds of considerations that justified their logical offshoots. We then combine these general principles with the evident facts about our inferential practices:

S-REDUCTIO: Assume both $\Gamma,\alpha \models \beta$ and $\Gamma,\alpha \models \beta^*$, then by generalized Totality, we follow (perhaps derivatively) both the rule with premises Γ,α and conclusion β, and also the rule with premises Γ,α and conclusion β^*, and so we must also follow the rule with premise Γ and conclusion α^*, so by the generalized MVC, this rule is valid, that is $\Gamma \models \alpha^*$.

We could also offer a more involved direct argument, as I did above for the standard structural rules. Like the standard structural rules, the *S*-reductio rule reflects general features of our (bilateralist) inferential practices.

To solve Carnap's problem, we also need bilateralist rules for the standard connectives. Here are some simple bilateralist rules for the connectives ¬, ∧, and ∨:

$$(\wedge I)\quad \frac{+\phi \quad +\psi}{+(\phi \wedge \psi)} \qquad (\wedge E)\quad \frac{+(\phi \wedge \psi)}{+\phi} \quad \frac{+(\phi \wedge \psi)}{+\psi}$$

$$(\vee I)\quad \frac{-\phi \quad -\psi}{-(\phi \vee \psi)} \qquad (\vee E)\quad \frac{-(\phi \vee \psi)}{-\phi} \quad \frac{-(\phi \vee \psi)}{-\psi}$$

$$(\neg I)\quad \frac{-\phi}{+\neg\phi} \qquad (\neg E)\quad \frac{+\neg\phi}{-\phi}$$

$$(\vee I)\quad \frac{-\phi \quad -\psi}{-(\phi \vee \psi)} \qquad (\vee E)\quad \frac{-(\phi \vee \psi)}{-\phi} \quad \frac{-(\phi \vee \psi)}{-\psi}$$

Obviously, in the bilateralist proof system, instead of trees (or sequences) of *sentences*, proofs will be trees (or sequences) of *signed sentences*.

Given a bilateralist understanding of proofs, this simple system of sentential logic is semantically complete (I give a completeness proof for a bilateralist system of first-order logic in the appendix to chapter 4). The key to its power is noting that the structural reductio rule gives rise to a powerful principle of reversal: if $\Gamma, \alpha \vdash \beta$ then $\Gamma, \beta^* \vdash \alpha^*$.[53] Bilateralist systems of classical logic like this have a number of nice syntactic features that are lacking in standard single-conclusion formulations of classical logic (but possessed by many single-conclusion formulations of intuitionistic logic). For example, the bilateralist system is *separable*—every proof of sentence ϕ from set Γ

uses rules only directly for connectives that occur either in ϕ or in some sentence in Γ (aside from uses of structural rules). It is also harmonious according to many definitions. Accordingly, many restricted inferentialist restrictions on MAC that single-conclusion classical systems fail are passed by bilaterlist systems (this issue has been explored in detail by Rumfitt).[54] As an unrestricted inferentialist, I don't put much stress on these features, but they are nice to have when you can get them.

To see that the bilateralist system solves Carnap's problem, note that the deviant valuation *v* from above, which assigned *T* to every sentence in the language, will fail to make the rule of ($\neg E$) valid in the new sense, since that valuation will assign truth to the rule's positive, negated premise but will also assign *T* to the rule's negatively signed conclusion. In symbols: $v(\neg\phi) = T$ and also $v(\phi) = T$ and so the rule is not valid in the extended sense. All other non-Boolean valuations will be similarly ruled out by these rules, which means, metasemantically, if the above system represents natural language logical rules, then each connective will be assigned its standard meaning.

To prove this: since the rules for "and"/"$\wedge$" are the same as the standard rules, the argument at the end of section I above suffices to show that the above introduction and elimination rules for "$\wedge$" and the MVC force the familiar truth function for "and". For "or"/"$\vee$": the rule of ($\vee I$) and the MVC force $\ulcorner \phi \vee \psi \urcorner$ to be false when both ϕ and ψ are false; and if either of ϕ or ψ is true, and $\ulcorner \phi \vee \psi \urcorner$ false, then one of the ($\vee E$) rules will fail; together these two points force the standard truth-functional meaning upon "$\vee$". For negation, the rule of ($\neg I$) forces $\ulcorner \neg\phi \urcorner$ to be true when ϕ is false; and if ϕ was true and $\ulcorner \neg\phi \urcorner$ true, then ($\neg E$) would be invalid, so $\ulcorner \neg\phi \urcorner$ must be false when ϕ is true. Together these force the standard truth-functional meaning upon "$\neg$". So, since this set of connectives is truth-functionally complete, the bilateralist rules force the standard truth-functional meanings on all of the sentential connectives.[55]

Of course, this account of how our rules force the standard connectives to have their familiar, classical truth-functional meanings doesn't float in the air.[56] We start with valuations assigning either truth (*T*) or falsity (*F*) to each sentence in the language, not both, not neither. Given these valuations and the assumptions of inferentialism,

the meanings of the classical connectives are secured by the bilateralist rules. This starting point might be questioned. Some people think that there are truth values beyond the two familiar ones, or that sentences can be assigned both truth and falsity. Of course, these are minority views that many of us find literally unthinkable. Somewhat more plausibly, we might imagine that valuations could be partial functions from the sentences of our language to the standard truth values. This also extends the more standard notion of a valuation, but not as radically. Partial valuations can be handled on inferentialist grounds, but doing so adds some irrelevant complexity and goes beyond the assumption of classical logic that I've been working with, so I won't pursue the matter here.[57]

Thus far I have only discussed sentential logic, but this system can be extended to the quantifiers by adding the following rules (where "*c*" is a term and "*v*" a variable):

$$(\forall I)\ \frac{+\phi(c)}{+\forall v\phi(v)} \qquad (\forall E)\ \frac{+\forall v\phi(v)}{+\phi(c)}$$

$$(\exists I)\ \frac{-\phi(c)}{-\exists v\phi(v)} \qquad (\exists E)\ \frac{-\exists v\phi(v)}{-\phi(c)}$$

Where "*v*" generalizes upon "*c*" and in ($\forall I$) and ($\exists I$) and "*c*" isn't in $\phi(v)$ or in any assumptions used to derive the rule's premise. This system can do everything that standard systems can do. The universal quantifier rules are simply the standard rules transferred into a bilateralist framework. The standard ($\exists I$) rule follows from my ($\exists E$) rule by reversal and the standard but more complicated elimination rule for "$\exists$" can also be derived from these rules.[58]

Does a version of Carnap's problem plague the standard quantifier rules? Yes, in a sense. It isn't usually connected to inferentialism, but worries have been raised about the power of standard quantifier rules to force the classical interpretation of our quantifiers. The most famous of these worries are related to versions of Skolem's paradox.[59] The key to answering these worries is not bilateralism, but

open-endedness (2.IV). The most general and most troubling versions of these concerns will be discussed in chapters 9 and 10, but at present it is worth showing that the rules above, understood in a fully open-ended fashion force "∀" and "∃" to have their standard semantic values.

This can be demonstrated most perspicuously by using so-called *generalized quantifier theory*, which sees quantifiers as properties of properties or, speaking extensionally, sets of sets.[60] According to this approach, the extension of "∃" on any given domain *D* is the set of all non-empty subsets of *D*, and the extension of "∀" on *D* is the unit set of *D*. Even when our current language *L* lacks a constant "*c*" witnessing a given quantificational claim, the open-ended validity of the quantifier rules ensures that "*c*" (or a translation of "*c*") can be added to give us an expanded language *L*+ in which the quantifier rules continue to be valid, by open-endedness. So by starting with non-empty domain *D* and letting "*ext*(*x*)" pick out the extension of "*x*" we can prove that *A* is a non-empty subset of *D* if and only if $A \in ext(\exists)$.[61] Similarly, we can show that $A = D$ if and only if $A \in ext(\forall)$.[62] These proofs both assume (unproblematically) that extensions are defined over *D*. They also assume (more problematically) that we can add to our language a name for any object in *D*, and a predicate for any subset of *D*.

This key assumption is bound up with the *philosophical* idea of open-endedness and will be discussed and defended in part III, especially in chapter 10. The central point of interest, *logically speaking*, is that inferentialists aren't committed to a substitutional account of the quantifiers. Instead the quantifiers can be interpreted in a standard, objectual/referential fashion. Issues about quantification over *absolutely everything* are more controversial. They will be discussed indirectly in part III (see 11.II and 11.III).

Combining the results about the sentential connectives with the results about the quantifiers, the validity of the bilateralist rules forces the standard classical meanings on the canonical logical constants. If these are an adequate model of our English conventions, then a bilateralist, open-ended inferentialist approach can explain how our logical constants come to have classical meanings. In chapter 4, I will argue that this unrestricted inferentialist approach also leads to

logical conventionalism. And in chapter 5 I will show that my approach extends well beyond classical logic. Before doing either of these things I want to step back and briefly discuss some of the broader philosophical virtues of inferentialism before discussing some reasons for preferring unrestricted to restricted inferentialism.

V Inferentialism and Understanding

Over the next three chapters, I am going to argue that unrestricted logical inferentialism leads to logical conventionalism, logical pluralism, and a naturalistically acceptable epistemology for logic. More accurately put: an unrestricted inferentialist theory of logic *is* a version of logical conventionalism. But first I want to outline two important virtues of inferentialism that can be dealt with briefly, one in this section and one in the next.

Logical inferentialism naturally coheres with a use-based, "know how" approach to the psychology of meaning. In this type of account, a speaker's understanding of a logical constant like "and" consists in his ability to use "and" in the right way. Of course, many different patterns of use for "and" are compatible with understanding "and", so we need to abstract away from various messy details and be somewhat holistic. But the account of logical constants assumed by inferentialists already builds the needed abstractions into the idea of a basic linguistic rule of inference.

On this basis, inferentialists will typically endorse something like the following account of understanding a logical constant:

> **Basic Understanding (BU)**: If someone uses a logical expression in accordance with its meaning-constituting rules of inference, then they have a basic understanding of the expression.

One of the central virtues of this account is that it doesn't *over-intellectualize* linguistic understanding. Many fairly young children in English-speaking countries understand "and" even though they are completely unable to give any type of systematic theory of the meaning of "and".[63] In fact, many of them, and perhaps many adults

as well, can understand the meaning of an expression like “and” without having a refined and developed concept of *meaning* at all. The inferentialist account of basic understanding, as summed up in **BU**, allows for this.

This is not to say that this account of understanding cannot be elaborated upon. To allow for the possibility of elaboration and further development, **BU** concerns only “basic” understanding (as the name suggests). What is more, **BU** provides only a sufficient condition for basic understanding. I don’t think that this sufficient condition is also necessary—you can have a basic understanding of an expression that you don’t use. Presumably this often happens with slurs of various kinds; we have a basic understanding of the slur without using it ourselves.

BU claims that using a logical expression in the right way suffices for having one kind of basic understanding of that expression. The door is left open for inferentialists to offer more involved and elaborate notions of linguistic understanding by building upon this simple foundation. I won’t endorse or discuss in detail any of these more elaborate accounts here. I think that a fully nuanced account of understanding, in all of its forms, will likely come in degrees. We could even modify **BU** so as to allow for partial understanding, if needed. It is also plausible that many distinct psychological notions are being employed when we talk about “understanding” an expression or “knowing what it means”. More and more elements can be added to allow for more and more robust types of linguistic understanding. But no matter how much else is added, a basic, use-based account will remain a central component of our theory of understanding.

VI Inferentialism and the Metaphysics of Logic

What metaphysical picture of the logical constants does unrestricted inferentialism lead to? There is a tradition in the philosophy of logic, intimately related to inferentialism, that sees logical constants as analogous to punctuation marks. This general line of thought

developed out of the early Wittgenstein's approach to logic and came to fruition in the work of Carnap and the middle Wittgenstein.[64]

A prominent way of developing this idea in contemporary discussions sees logical notions as, in some fashion, *indicating* aspects of the structural rules of proof.[65] Formally this approach often trades on the fact that when we use a multiple-conclusion sequent calculus to analyze the patterns of logical inference in a language, some of the differences between alternative logics seem to be pushed into the structural assumptions. The operational (non-structural) rules for the conditional in classical, intuitionistic, and relevance systems can be shown to be identical from some such perspectives. Results like this are often seen as supporting a philosophical position whereby the various competing logics all agree on what the conditional and other logical constants mean, whatever their other disagreements.[66]

By contrast, the natural-deduction-based approach I have been using sees different logics as disagreeing in what is meant by the conditional and negation and whatnot. This is because in a natural deduction setting, each of these logics will have different basic rules for these constants. Owing to this feature, my approach more naturally coheres with a form of logical pluralism (this type of pluralism will be discussed in detail in chapter 5). Despite this seemingly important disagreement, my view of the metaphysics of logical constants is quite similar to the "punctuation mark" approach. I think that logical constants indicate or express aspects of our use of language and patterns of reasoning. The sense in which this is so is perhaps best captured by Robert Brandom's suggestive phrase that logic involves *making explicit* what is already, in some sense, *implicit* in our practice of inferring.[67] This idea can be made clear without moving to a multiple-conclusion system.[68]

I'll illustrate this by discussing negation, conjunction, and the conditional. The notion of negation can be seen, in a sense, as a *reflection*, within our language, of the attitude of rejection. Of course, this is not to say that "not" *means* "is rejected" or anything like that. Rather, it merely glosses what is apparent from the bilateralist rules for negation:

$$(\neg I)\quad \frac{-\phi}{+\neg\phi} \qquad (\neg E)\quad \frac{+\neg\phi}{-\phi}$$

As can be easily seen, for any sentence, $-\phi \dashv\vdash +\neg\phi$ and $+\phi \dashv\vdash -\neg\phi$, so the inferential role of negation is intimately linked to the role that rejection plays in our bilateralist formalization. This is what it means to say that negation is an object-level reflection of the notion of rejection. Because of the validity of these rules, negation is a flexible tool for, in some very general and not all that exact sense, *expressing* the rejection of a sentence. And because the negation sign, unlike the force indicator for rejection, can be embedded, the notion allows for sophisticated and varied contents to be entertained and judged.

The notion of conjunction also brings into our language a general feature of our inferential practices. It "makes explicit" the implicit fact that contents expressed by distinct sentences may be jointly accepted. This can be seen by looking again at the rules for conjunction:

$$(\wedge I)\quad \frac{+\phi \quad +\psi}{+(\phi\wedge\psi)} \qquad (\wedge E)\quad \frac{+(\phi\wedge\psi)}{+\phi} \quad \frac{+(\phi\wedge\psi)}{+\psi}$$

These rules illustrate the role of conjunction by connecting a conjunction to its conjuncts in the appropriate fashion. When I say that conjunction imports general or structural features of our inferential practice into the language, this is all that I mean.

Similarly for the conditional. In the full language of section IV the conditional is defined, but we could also take it as a primitive with the following rules:

$$(\rightarrow I)\quad \frac{+\psi}{+(\phi\rightarrow\psi)} \quad \frac{-\phi}{+(\phi\rightarrow\psi)} \qquad (\rightarrow E)\quad \frac{+\phi \quad +(\phi\rightarrow\psi)}{+\psi}$$

The elimination rule here is simply modus ponens, and the introduction rules lead to the standard conditional proof rule.[69] The conditional reflects into our language the relationship between the

premises and the conclusion of an inference rule. Again, care must be taken, since this is not to say that “if ... then” *means* the same thing as deducibility in a formal system or logical consequence or anything like that. It is just that logical operators like these reflect, somewhat imperfectly, aspects of our rule-governed linguistic behavior. They are, in a broad sense, shadows cast by the syntactic rules of our language. This is but one illustration of the book’s guiding metaphor. There are more to come.

VII Unrestricted versus Restricted Inferentialism

Section IV’s inferentialist account of the meanings of logical constants can be adopted by all unrestricted inferentialists. It can also be adopted by liberal restricted inferentialists who allow the bilateralist, open-ended rules of that section to be meaning-constituting rules. This also applies to both of the virtues of inferentialism discussed in the two preceding sections. But in the next chapter, when I move from logical inferentialism to logical conventionalism, it will be important that unrestricted rather than liberal but restricted inferentialism is adopted (see 4.II).

Despite the popularity of inferentialism in the philosophy of logic, as far as I’m aware, unrestricted inferentialism has been rejected by every single post-Carnap philosopher who is not identical with the present author.[70] Restricted inferentialists accept Logical Inferentialism, the MVC, and Totality, but they don’t think that meanings are cheap, so they don’t accept MAC. Different versions of restricted inferentialism will offer different restrictions of the MAC principle. One popular restriction requires that newly added meaning-constituting rules conservatively extend our starting language (where a full language $L+$ is a conservative extension of a full language L just in case for all sentences ϕ in the shared language, if $\vdash_{L+}\phi$ then $\vdash_{L}\phi$).[71] But conservativeness isn’t the only constraint that has been proposed. Other restricted inferentialists demand that introduction and elimination rules must be in *harmony* with each other, and go on to give an account of what it means for rules to be in harmony.[72] There are almost as many versions of restricted inferentialism as

there are restricted inferentialists. Many of the constraints that have been proposed are quite restrictive, even to the point of ruling out classical logic.[73] Others are more permissive, allowing for classical logic and, in some cases, significant portions of mathematics to be defended on inferentialist grounds.[74]

In general though, restricted inferentialists propose, in advance, constraints on which rules can found meaning. I find this unappealing. Clearly beings like us can, in actual fact, follow the rules of classical logic and mathematics and more. In the process of doing so we, just as clearly, manage to communicate information to each other, form beliefs about the world, and successfully use the information imparted. Given this, to deny that these rules can be meaning constituting is dangerously close to baldly denying evident realities. I prefer instead a naturalism-rooted approach to inferentialism, trading on the key metasemantic idea that *meaning is ultimately determined by language use*. In contrast, some restricted inferentialists proceed as if they can move directly from recherché formal constraints on a natural deduction system or a sequent calculus to claims about meaning constitution. This is a highly dubious way to proceed—there is no mathematical substitute for metasemantics.[75] If a version of restricted inferentialism flowed from natural metasemantic principles and constraints that would be one thing, but it is unrestricted inferentialism, rather than restricted inferentialism, that best coheres with important results and trends in metasemantics.

To illustrate this, note that Logical Inferentialism, the MVC, Totality, and MAC combined lead to the following general principle that sums up much of unrestricted inferentialism:

> Charity to Use: All of the basic inference rules for a logical expression are automatically valid and only rules whose validity is required by those basic inference rules are valid.

For an unrestricted inferentialist, if speakers follow some rules of inference, then those rules are automatically valid, no matter whether they are conservative, harmonious, contradictory, or what have you.[76] I have named this principle “Charity to Use” because it is closely related to a cluster of principles in metasemantics and the theory of

interpretation that have often gone under the name “charity”. Such principles are endorsed by anyone who thinks that meaning is directly connected to language use. I think that unrestricted inferentialism should be welcomed not only by those who focus on charity or rationality in interpretation, such as Donald Davidson, Daniel Dennett, David Lewis, and W.V. Quine, but also by those who endorse conceptual or inferential role semantics, such as Ned Block, Robert Brandom, Hartry Field, Gilbert Harman, and Wilfrid Sellars, and by use theorists, influenced by the writings of the later Wittgenstein, such as Paul Horwich and Wittgenstein himself.[77]

This provides a central reason for accepting unrestricted inferentialism in general and the crucial MVC in particular. That is, the MVC can be defended by appeal to any of these general approaches to metasemantics. What is more, I believe that any approach to metasemantics that ties meaning closely to use will end up having to endorse the Charity to Use principle, once it is properly understood. If this is right, then any acceptable approach to metasemantics will vindicate the MVC and Charity to Use. This is a powerful reason for supporting unrestricted inferentialism over restricted inferentialism. The theoretical virtues of unrestricted inferentialism that this book aims to demonstrate, when taken in total, provide another powerful reason for accepting unrestricted inferentialism. I will return to a global, big-picture discussion of the correctness of my unrestricted inferentialist approach in part IV, especially in 14.II, but many related points will be made on the way to that final chapter, especially in chapter 5.

Even many proponents of restricted inferentialism will acknowledge the weight of the points I’ve been making in favor of unrestricted inferentialism, they just don’t think they establish that we should endorse unrestricted inferentialism. Instead, they take these points to establish that *if* unrestricted inferentialism is coherent, then it should be endorsed. However, they also think that—alas!—unrestricted inferentialism is not coherent. They think this because of the apparently powerful arguments that originally moved otherwise sympathetic philosophers away from the pure, unrestricted form of inferentialism in the first place.

I will discuss the most important of these arguments in detail in chapters 5 and 7 but it is worth briefly saying something of relevance here (if only because many readers will have been furiously muttering these arguments to themselves while reading this chapter). Charity to Use entails that if some other language includes a rule that is invalid in our language as a basic rule, then that rule will be valid *in that language*. Let's illustrate this briefly with the rule of affirming the consequent:

$$(ATC) \quad \frac{\psi \quad \phi \rightarrow \psi}{\phi}$$

In our language this rule is invalid, but there is nothing to stop someone from using this rule as a basic rule in an alternative language. If they did, then by MAC it would be meaning-constituting, and by the MVC it would be valid. In this case, clearly "→" in that language wouldn't mean the same thing as "→" in our language. There is nothing mysterious or problematic about this as it stands. Speakers often use the same symbols in different ways, thus ending up with different meanings. This much follows from the idea that meaning is determined by language use, and it is nicely summed up in the tenets of unrestricted inferentialism. Chapter 5 will discuss the types of logical pluralism that unrestricted inferentialism does and does not give rise to.

Perhaps many will admit the justice of my response to this particular argument, but a distinct yet closely related putative counterexample is largely responsible for unrestricted inferentialism's unpopularity: A.N. Prior's infamous tonk connective.[78] Tonk is stipulated to be a binary sentential connective with the following introduction and elimination rules:

$$(tI) \quad \frac{\phi}{\phi \text{ tonk } \psi} \qquad (tE) \quad \frac{\phi \text{ tonk } \psi}{\psi}$$

Adding the tonk rules to English—call the resulting language "Tonklish"—would enable us to prove *any* sentence from *any other*

sentence. This is a disaster, to say the least. In Tonklish, by MAC, the meaning of "tonk" is given by (*tI*) and (*tE*); and so by the MVC both (*tI*) and (*tE*) are valid in Tonklish. Yet this seems absurd. If unrestricted inferentialism were correct, then it seems that we could simply move to Tonklish and put our hands on a runabout inference ticket allowing us to establish *anything.* Want to prove the Riemann hypothesis or that *P* = *NP*? Easy—just add the tonk rules, get to your very short work, and wait for the glory to roll in.

The tonk rules are thought to be invalid because, from "2 + 2 = 4," it seemingly follows, via a tonk introduction followed by a tonk elimination, that Saul Kripke was born before Plato. But while the sum of 2 and 2 is 4, Saul Kripke was not born before Plato, and so the tonk rules are invalid. Something like this basic argument seems to have convinced every single philosopher who has considered it, but it is a non sequitur. From the fact that the sentence "Saul Kripke was born before Plato" is false in English, *nothing whatsoever* follows about the truth value of the syntactically identical sentence in Tonklish. Adding the tonk connective to English and following the tonk rules drastically changes the meaning of every sentence in the language.[79]

So while there is a recipe for establishing the *sentence* "*P* = *NP*" in Tonklish, that isn't at all the same thing as establishing the sentence "*P* = *NP*" in English. It would be just as good if we could also establish that "*P* = *NP*" should be translated homophonically into itself from Tonklish into English, but clearly it shouldn't be. Tonklish is a trivial language in which every sentence is true. This makes the language worthless and stupid but it doesn't provide any counterexample to unrestricted inferentialism. I will say much more about this in chapter 5, but for now I think simply pointing out the error (or, at least, the unjustified assumption) of homophonic translation that lies at the heart of all of our tonk-induced fears, as well as less frightening bugaboos, is enough to undermine the widely held view that unrestricted inferentialism is a non-starter. In my published works I have called this error *the translation mistake*, and it is, I think, perhaps the most dangerous and widely made mistake in all of contemporary philosophy, in this context and others. I'll discuss the

mistake in detail in chapter 5, after establishing the conventionalist credentials of unrestricted logical inferentialism in chapter 4.

≈≈≈≈≈

UNRESTRICTED INFERENTIALISM IS the correct metasemantic theory of the logical constants. All resistance to unrestricted inferentialism can be explained away and put aside. The reasoned resistance of philosophers is almost entirely based on fundamental misunderstandings, such as those unearthed in the discussion of tonk. The unreasoned resistance that some readers may still feel is based, I think, on the feeling that unrestricted inferentialism makes things *too easy*. This is philosophy, and philosophical problems, perhaps especially those of meaning and content, are supposed to be depressingly hard. It may feel like I have stepped up to this Gordian knot and "untied" it with one clean, swift stroke of my sword, like a would-be Alexander. But once this emotional reaction has dimmed, I hope it will be admitted that theoretical simplicity and the resolution of hitherto impossibly difficult problems is a strength of the theory rather than a drawback. This theory, once accepted, leads to a version of logical conventionalism, as characterized in Part I. But this needs to be demonstrated, not just asserted.

---

1 In a recent, wide-ranging survey of philosophers, 51.6% of respondents indicated support for classical logic with only 15.4% indicating support for some non-classical logic: see Bourget and Chalmers (2014).

2 To say this is not to endorse a simple causal account of meaning of the kind endorsed by Fodor (1990). Instead it is just to note that the metasemantics of *reference*, at least, must be causal in some fashion or other. This much has been widely recognized since the pioneering discussions by Kripke (1980) and Putnam (1975). For criticism of Fodor's approach see Boghossian (1990).

3 Something like this is called the *meaning inference connection* in my (2015*a*) and (2015*b*).

4 As we will see in 3.II and especially in chapter 5, the situation here is somewhat subtle, since we can sometimes change the inference rules according to which an expression is used by adding rules to our language that don't *explicitly* involve the expression in question when stated in their general form.

5 Metaphysical or objective naturalness has been popular in metasemantics since Lewis's (1984) adoption of an idea proposed by Merrill (1980) in response to Putnam (1980).

6 As I'll discuss, "and" is an easy case, the other connectives are much harder to handle inferentially. They will be handled in section IV. Here I'm only trying to explain the basic philosophical idea of inferentialism, so have chosen to focus on the simplest case.

7 In section IV of this chapter and again in chapter 4 a more expansive notion of validity will be used.

8 Compare the start of Field (1986).

9 For instance Field (2009) worries about characterizing validity in terms of necessary truth-preservation on these grounds.

10 For more detail on this type of issue, see the discussion of Montague's paradox in 12.IV.

11 The distinction between "any" and "all" goes back to Russell (1908).

12 Despite how it may sound, Totality does not enforce a separability constraint—see the discussion in the next section and in 5.II.

13 See Kripke (1980). The expression "α" in (Gold) is a substance term, rather than a singular term, so (Gold), though plausible in natural language, can't be formalized in standard systems.

14 It is possible that these two characterizations come to the same thing, and the apparent variation is merely a matter of how we are describing the world or scenario in question. I mean for my remarks to be neutral on this point.

15 Compare the writings of two-dimensionalists such as Chalmers (1996) and Jackson (1998), as well as Putnam (1992).

16 For one such directly conventionalist approach, see Sidelle (1989). This type of approach is not unrelated to the two-dimensional analysis of necessary *a posteriori* that was just mentioned.

17 Wittgenstein (1974), X–133.

18 Carnap (1937), page xv.

19 In Gentzen (1934); Gentzen's formal work and scattered philosophical remarks have exercised an important influence on the development of inferentialism.

20 See Popper (1947) and Kneale (1956).

21 See Boghossian (1996), Dummett (1991*a*), Hacking (1979), Peacocke (1987), and Tennant (2002) for a small sampling of broadly inferentialist writings in philosophy (as opposed to proof theory or logic proper).

22 In my (2015*b*) and, implicitly, from an entirely different direction, in my (2018). As mentioned in the preface, despite publication dates, the first versions of these papers were written in 2011.

23 In my (2015*a*) and (2015*b*) I called a version of this principle the *meaning rules connection* or the MRC in analogy to the MVC. Here I have opted for a more suggestive name.

[24] See Fodor and Lepore (1992) for a book-length assault on semantic holism of various sorts.

[25] Compare Peacocke (1987) and (1992) on *primitively compelling* inferences.

[26] See my (forthcoming *a*) for a full discussion.

[27] Of course, for astronomical instances of even putatively basic rules like modus ponens we have only composite dispositions to reason with them. However, even in this case, the disposition to (for example) accept the conclusion upon accepting the premises is itself simple, whereas for other rules with astronomical premises, the linking disposition between premises and conclusion will be complex (or composite if the chain is long enough).

[28] My distinction between basic and derivative rules is also related to Harman's (1986*a*) psychological notions of immediate implication and immediate inconsistency. I was unfamiliar with Harman's notions when writing the initial version of these paragraphs, but in a rough, idealized sense it might seem that Γ immediately implies ϕ just in case the rule with premises Γ and conclusion ϕ is basic. The only caution is that it isn't clear that Harman's more directly psychological notion allows for idealization without distortion (consider astronomically large sentences used in a simple inference rule).

[29] This may not be fully accurate in general, but I won't discuss this here since the potential counterexamples I have in mind are paradoxical cases like the liar.

[30] One consequence of this is that invalid but inductively strong rules cannot be basic rules in our language, or indeed, in any language. Though I won't discuss this here, I think this is the correct result for metasemantic reasons related to those discussed in section V of this chapter and then again in chapter 5. See also the discussion of strong validity in the previous section.

[31] Part III will relax this somewhat in the case of mathematics—see chapter 10.

[32] If transitivity ever fails, it will fail when our complex dispositions come apart from our composite dispositions. Since I am assuming (see earlier in this section, including footnote 29) that this never happens, it is not something I am going to discuss in detail here.

[33] Since my arguments here appeal to my account of what it is to follow an inference rule, some might worry that I am forced to reject the very possibility of so-called sub-structural logics—logics that reject one of the standard structural rules (see Restall (2000) for an introduction to sub-structural logics). In a sense this might be right, but there is nothing stopping us from admitting the possibility of communities that "reason" in a way that conforms to a sub-structural logic. I just think that those communities aren't well described as "inferring" or "following inference rules", nor perhaps even as "accepting" and "rejecting" various claims. Instead these imagined communities will engage in inference-like and acceptance-like behaviors. I won't say anything more about this possibility here, since for our standard concept of inference rule-following the structural rules surely hold. The general account of conceptual pluralism in

5.III can be used to make sense of alternative "languages" with only inference-like behaviors.

34 As mentioned at the start of this section, my (2015*b*) also defends the view at length, focusing on the famous tonk-objection from Prior (1960).

35 Proponents of conceptual role semantics dispute whether conceptual roles "stop at the skin," as Block (1986) insists, or reach out to the world, as Harman (1982) insists. I think most of the differences here are merely verbal (Block is also sensitive to this possibility).

36 Obviously, the talk of interactions with the world isn't meant to rule out skeptical scenarios such as vat worlds. I don't think that taking account of this requires anything special, since talk of "the world" can refer to the world "in the image" in Putnam's (1981) vivid terminology.

37 Inferentialists will ultimately explain the fact that an expression α is of type-X (predicate, universal quantifier, *et cetera*) in terms of the expression's inferential behavior. This will be further discussed in part III—see 9.II.

38 This is akin to the language-entry and language-exit rules of Sellars (1953) and Brandom (1994).

39 This approach was initiated by Tarski (1986); McGee (1996) is an important modern contribution to this literature; Woods (2014) emends this type of procedure to allow indefinites to count as "logical".

40 Other broadly inferentialist approaches to the demarcation of logic are often extremely complex; for example, Hacking (1979) takes a constant to be logical just in case its operation rules in a sequent calculus formulation have the sub-formula property; and can be proved to be conservative with respect to the antecedent facts about deducibility using finitistically acceptable reasoning.

41 Compare the approach of Searle (1969), which leans heavily on rules.

42 Hanson (1997) also seems to think that logical truths and conceptual truths are necessary in roughly the same way, and of course this was widely believed by the logical positivists. Hanson also suggests, as I did above, that the notion of logicality might be somewhat imprecise.

43 Not all inferentialists are naturalists, for example, Brandom (1994) develops a version of inferentialism employing primitive normative notions. I am unsure of my interpretation though, since Brandom's writings are extremely difficult to follow (including his "introduction" (2000)); see also Peregrin (2014). Restall (2002) briefly claims that the inner-workings of Brandom's approach mean that he isn't really an *inferentialist*, as the term is usually used. I think I agree but find it amusing, since Brandom seems to have invented the term "inferentialism".

44 See Carnap (1943); it was apparently independently rediscovered by McCawley (1981) and perhaps several others. Although the problem was discussed by specialists in the interim, Raatikainen (2008) recently reintroduced the problem to a general philosophical audience. For a book-

length discussion of the problem and some of the options, see Garson (2013). The standard rules do fix the classical meaning for “and”, as was shown above.

45 For a good survey of different natural deduction systems and notations, see chapter 6 of Bostock (1997).

46 See page 22 of Troelstra and Schwichtenberg (2000), for example.

47 This restriction on valuations in endorsed by McCawley (1981), though he is aware that it doesn’t solve the problem. A similar restriction is also used for many of the results in Garson (2013).

48 See Carnap (1943), Shoesmith and Smiley (1978), and Restall (2005)—Restall’s approach is a mixture of multiple conclusions and rejection-based approaches.

49 For just one example of a complaint of this kind see Rumfitt (2000).

50 Again, see Smiley (1996) and Rumfitt (2000) for bilateralist systems—the name “bilateralism” is Rumfitt’s.

51 Could Totality alone be used to solve the problem? Not directly. Totality requires that the only logical inferences that are valid are those whose validity follows from the automatic validity of the basic rules. Formally, validity requires truth-preservation over all valuations, so the problematic truth-only valuation v used in the proof above doesn’t automatically lead to any new validities (it does lead to many further truth-preserving inferences, including logical inferences, in v, but that is a different matter). This is related to the fact that there are at least two definitions of “validity” that might be employed here, a natural “global” version and a “local” version; see Garson (2013) for discussion. This issue also pops up in supervaluationist treatments of vagueness, where the choice between global and local validity is of crucial importance after the introduction of a “definitely” operator into the language; see chapter 5 of Williamson (1994) for discussion.

52 My formalization is largely based on one given in Incurvati and Smith (2010), which is itself based on the work of Smiley and Rumfitt.

53 *Proof.* Assume $\Gamma,\alpha \vdash \beta$, so also $\Gamma,\alpha,\beta \vdash \beta$ by weakening on the left, and since $\beta \vdash \beta$ by reflexivity, by multiple applications of weakening on the left, $\Gamma,\alpha,\beta \vdash \beta$, so by *reductio* $\Gamma,\beta \vdash \alpha$ as desired (if we were working with sequences instead of sets of formulas on the left we would also have to appeal to a permutation rule).

54 In Rumfitt (2000).

55 Murzi and Hjortland (2009) argue that the bilateralist approach fails to fully solve Carnap’s problem since we can consider extended valuations that generate a version of Carnap’s problem at a higher level. They do this by treating the force indicator “-” as, in essence, an unembeddable negation sign, drawing on an equivalence result about such negations proved in Bendall (1979). But as Incurvati and Smith (2010) argue in response, these extended “valuations” aren’t valuations in any standard sense, since they don’t merely assign truth values to sentences, but also take into account force indicators. To

fail to treat these signs as force indicators is simply to reject the bilateralist framework altogether. I think that section 5 of Kroedel (2012) also fails to treat "-" as a force indicator.

56 Compare Hacking's (1979) discussion of "semantic frameworks" as the base of his "do-it-yourself" semantics.

57 See Rumfitt (1997); his solution uses both multiple conclusions and bilateralism and, on top of that, only works if a certain type of compactness result holds for the base logic. I think there is a simpler approach available if we move from *bi*lateralism to *tri*lateralism by adding another force indicator "=". This third attitude can be thought of as expressing a non-epistemic attitude of neutrality toward a sentence. From there we can determine semantic values of (for example) the strong three-valued semantics introduced in Kleene (1952) by adopting an appropriate trilateralist natural deduction system, along with suitable further constraints on how the natural deduction rules constrain admissible valuations. Some complications arise, but I think they can be dealt with without too much trouble—I attempt this in an unpublished paper.

58 Proof of the standard (∃E) rule: assume that Γ, +ϕ(c) ⊢ α and that "c" isn't in Γ or α, then Γ,α ⊢-ϕ(c) by reversal, so then by (∃I) rule, Γ,α ⊢-∃vϕ(v) so then by reversal again, Γ, +∃vϕ(v) ⊢ α, as desired.

59 See the classic Skolem (1922); see also Antonelli (2013).

60 See Mostowski (1957) and Lindström (1966).

61 *Proof.* Left to right: assume that A is a non-empty subset of D and let "F" be a monadic predicate we add to our language such that ext(F) = A and assume for *reductio* that Aext(∃). Since A is non-empty, it has some member o, so add constant "α" such that ext(α) = o. We are now in an expanded language in which "∃xFx" is false, but "Fα" is true, contradicting the open-ended validity of (∃E). Right to left: with subset A of D a member of ext(∃), assume that A is empty, for *reductio*. Let "F" be a monadic predicate we add to our language such that ext(F) = A; also add a constant "α" to the language and note that no matter the extension assigned to "α", ext(α)ext(F), since, by assumption, A is empty. We are now in an expanded language in which "Fα" is (arbitrarily) false but "∃xFx" is not, contradicting the openended validity of (∃I).

62 *Proof.* Right to left: assume A ∈ ext(∀) and let "F" be a monadic predicate we add to our language such that ext(F) = A and assume for *reductio* that A≠D. Since A≠D, there is some object o in D but not A (since A ⊆ D it can't be the other way around), add a constant α to the language such that ext(α) = o, we are now in an expanded language in which "∀xFx" is true but "Fα" is not, contradicting the open-ended validity of (∀E). Left to right: let A = D but assume, for *reductio*, that Aext(∀). Let "F" be a monadic predicate we add to our language such that ext(F) = A and "α" be any constant such that ext(α) = o for some member o of D. We are now in an expanded language in which "Fα"

is true for arbitrary "α" but "∀xFx" is not, contradicting the open-ended validity of (∀I).

63 Some—for example, Gelman (2003)—have argued that fairly young children are quite sophisticated, even *metaphysically* speaking. I don't mean to deny this possibility.

64 See the discussion of the connection between Wittgenstein's *Tractatus* and Carnap's view of logic in Friedman (1997) and the other essays in part III of Friedman (1999).

65 Perhaps the first systematic development of this idea is in Došen (1989).

66 This type of approach to meaning change and logic change is developed by Restall (2002) and Paoli (2003), and can be seen as a formal, proof-theoretic approach to working out thoughts endorsed by Putnam (1962) and Haack (1974).

67 Hence the title of Brandom (1994). I am not at all confident in my Brandom exegesis, so my appropriation of his slogan here might have changed its original sense. The Hempel quote in 1.I also used a similar idea of "making it explicit".

68 The approach of Restall (2005) nicely attempts to account for a multiple-conclusion notion of consequence as characterized by primitively incoherent patterns of assertion and denial; see Woods (2016) for critical discussion.

69 Proof: Assume that Δ, +ϕ ⊢ +ψ, then since -(ϕ → ψ) ⊢ +ϕ by (→ I) and reversal, by transitivity, Δ,-(ϕ → ψ) ⊢ +ψ, and by the other (→ I) rule, +ψ ⊢ +(ϕ → ψ), so by reversal and thinning, Δ,-(ϕ → ψ) ⊢-ψ, so by *S*-reductio, Δ ⊢ +(ϕ → ψ), as desired.

70 Skorupski (1993), section VII, says things that sound somewhat like an endorsement of unrestricted inferentialism, but discussion later in the essay supports a non-inferentialist interpretation (rejecting the MVC). Peacocke (1993), responding to Skorupski, doesn't interpret him as an unrestricted inferentialist, but as what might be called an extremely permissive restricted inferentialist—placing only *consistency* as a constraint on MAC.

71 In effect, this is one of the restrictions offered in the very first version of restricted inferentialism, in Belnap (1962). There Belnap also notes that his conditions are context-sensitive; for further discussion of this, focusing on the case of tonk, see Cook (2005).

72 The canonical discussion of harmony is in Dummett (1991*a*); other influential versions of restricted inferentialism are found in Belnap (1962), and Prawitz (1965) and (1971). For a more recent discussion see Read (2010).

73 See Dummett (1991*a*) and Tennant (1997) for example.

74 Permissive proposals include the harmony proposals of Weir (1986) and Read (2000), the cut-elimination approach of Hacking (1979), and some of the approaches in Garson (2013).

75 The allusion to Kripke's (1976) methodological sermon is intended.

76 This principle is related to "logical charity" from my (2018).

77 See Block (1986), Brandom (1994), Davidson (1984), Dennett (1987), Field (1977), Harman (1982), Horwich (1998*a*), Lewis (1974), Quine (1960*b*), Sellars (1953), and Wittgenstein (1953). Charity principles have figured in some of my published work; perhaps my most extensive discussions are in my (2016*c*) and (2018).

78 Prior (1960).

79 For a full defense of unrestricted inferentialism from the terrors of tonk, see my (2015*b*) and sections 5.I and 5.II.

4

Logical Conventionalism

Against the background set out in part I, chapter 3 developed an unrestricted inferentialist theory of the meanings of the logical constants. But this doesn't yet tell us that all logical facts are fully explained by our linguistic conventions. Here I will show that this unrestricted inferentialist theory leads to logical conventionalism. Sometimes, both here and earlier in the book, I say that unrestricted logical inferentialism *entails* logical conventionalism, but it's really more accurate to say that unrestricted logical inferentialism *is a version of* logical conventionalism.

Despite this, inferentialism and conventionalism have rarely been paired together in our post-Carnap world. Why the divorce? Because it is only *unrestricted* inferentialism and not *restricted* inferentialism that truly leads to conventionalism, and unrestricted inferentialism has been almost universally rejected in recent philosophy. After presenting conventionalist explanations of logical validity, truth, falsehood, and necessity, I will critically discuss a popular putative alternative explanation of the logical facts and then round out my discussion by showing that conventionalism leads to an attractive metaphysical picture of logic (in a broad sense of "metaphysical") as well as smoothly allowing for the normativity of logic.

I From Inferentialism to Conventionalism

There are many different ways of characterizing the logical truths. In 3.III I characterized the logical constants or expressions as those that are both ancestrally non-empirical and highly topic neutral and claimed that the canonical list of logical constants from first-order logic all count as "logical" according to this standard. The logical

rules in a given language *L* are the rules that *implicitly define* logical notions in *L*. We can then say that the logical truths of *L* are the sentences of *L* that are provable using only the logical rules of *L* together with the background structural rules.

Why the detour through "logical rules"? Why not just say that the logical truths of *L* are the sentences of *L* that are provable using highly topic neutral and non-empirical rules in *L*? Because we can imagine expressions whose introduction rule is non-empirical, but whose elimination rule is empirical. Talking of logical rules ensures that cases like this are avoided, so that nothing we may prove traces to the world, so to speak, in either direction. So understood, the logical truths are the *particular* logical truths. To express *general* logical truths, we need a language with sufficient devices of generalization.

Given this account of logical truth, conventionalism is easy to establish on the assumption of unrestricted inferentialism. Assume that *p* is a logical truth in *L*; this means virtually by definition that *p* is provable in *L* using only the logical rules. These rules are basic, by assumption, so by the MVC they are automatically valid. That *p* is provable from these rules thus explains *p*'s truth and necessity (since the rules are valid). And that *p* is provable from logical rules shows that it is a logical truth. The conditions on rules induced by Logicality are themselves features of our conventions, so the fact that a rule is "logical" is also a fact about our total package of conventions. So all of the facts in this explanation concern only the linguistic rules of *L*, so the explanation is fully conventionalist.

An analogous explanation can be given for the logical falsehoods of *L*. Define a *logical falsehood* as a sentence that is refutable using the rules of *L* alone. Given the bilateralist perspective, this means that -*p* can be proven from the empty set using only the logical rules of *L*, in our full language representation. That the sentence *p* is false, as well as logically false, is also fully explained using only the rules of *L*.

Of course, at this level of generality, it feels like a rabbit is being pulled out of an empty hat. In particular, the definition of "logical truth" in terms of provability from logical rules already very closely connects the logical truths to our rules, and some may be worried

about that. But any plausible alternative definition will be extensionally equivalent in this context (but see section III below). Explanation is hyperintensional (1.III), but this extensional equivalence will mean that the logical truths are provable from—and thus explained by—our basic, meaning-constituting rules. Let me illustrate this by discussing an alternative approach.

Let's say instead that a sentence *p* in a language *L* is a "logical truth" of *L* just in case the sentence is true and any sentence that results from uniformly substituting for each non-logical expression in *p* with an expression of the same grammatical type drawn from *L*'s lexicon, is also true. This definition definitely isn't building in any *conventionalist* assumptions. In fact, a version of this definition was favored by Quine, the the canonical opponent of conventionalism.[1] This type of definition is usually seen as a cashing out of the idea that logical truths are true in virtue of their logical form.

Yet the conventionalist explanation goes through even if we start with this definition. Let *p* be a logical truth of *L* according to this second definition. Let p^S be the result of replacing the non-logical expressions in *p* with schemata of the appropriate grammatical type. Now, the schematic rule with no premises and p^S for a conclusion is valid in *L*. Thus, by Totality, this rule is a valid logical inference (but not necessarily a *logical rule*) and so is provable using the meaning-constituting rules for the logical expressions occurring in p^S.[2] From this point, the account is exactly the same as above. That *p* is true and necessary is explained by its being provable from the basic, meaning-constituting rules of *L*. That it is a logical truth by this definition is also guaranteed and explained by the associated schema being provable with no premises. A directly analogous account also goes through for logical falsehood.

For the sake of definiteness, it will be worth going through these arguments in more detail, not in general, but in particular. In chapter 3 I assumed that the logical constants in English are the classical constants and then showed how to vindicate that assumption on inferentialist grounds. I want to stress again that I'm not making this claim with too much seriousness—as the next chapter will discuss, logical conventionalism entails logical pluralism, so my assumption here isn't partisan in any deep sense. So let us again assume that

we speak a language where the logical rules are given by the classical bilateralist system from chapter 3.

The logical truths of this system correspond to the classical logical truths. All that this means is that if we are looking at a formal representation of English, the sentences in the formal representation that the standard semantics says are logical truths will correspond to English logical truths and, similarly, when the standard semantics says that sentence ϕ is a logical consequence of some set of sentences Γ, this logical relationship holds between the associated English sentences.[3] For the moment we won't concern ourselves about the portions of natural language that don't correspond to first-order logic (for this, see section III below).

Let's briefly review some notation that was already used at times in chapter 3: if a sentence ϕ is provable from a set of sentences Γ according to our rules, we say that ϕ is *provable* from Γ or, in symbols: $\Gamma \vdash \phi$. And if a sentence ϕ is provable from the empty set of premises we say that ϕ is a *theorem*, or, in symbols: $\vdash \phi$. When a sentence ϕ is true in every interpretation in which every sentence in the set Γ is true we say that ϕ is a logical consequence of the set of sentences Γ (or that ϕ *follows from* or *is entailed by* Γ) or, in symbols: $\Gamma \models \phi$. And if a sentence ϕ is true in all interpretations *simpliciter* we say that ϕ is a *logical truth* or, in symbols: $\models \phi$.

It is a familiar but remarkable fact about the standard semantics for first-order logic that it admits of a recursive, complete proof procedure. In fact, many different natural deduction systems for classical logic are both sound and complete with respect to the standard semantics. In symbols, where Γ is a set of sentences and ϕ a sentence, $\Gamma \models \phi$ if and only if $\Gamma \vdash \phi$.[4] In chapter 3, following Smiley and others, I used a bilateralist natural deduction system to solve Carnap's categoricity problem for the sentential connectives. I then extended this treatment to deal with similar worries about the quantifiers. Since such systems use, not sentences, but signed sentences—sentences prefixed by either one or the other of our two force indicators "+" or "-"—the standard completeness theorem doesn't directly apply to them. But there is a natural way to extend

the standard semantics to cover bilateralist systems. This was discussed in 3.IV, but let me go into a bit more detail about it here.

Let *V* be the set of truth valuations for standard first-order logic, understood standardly, and define a set of correctness valuations, *C*, each member of which is induced by the truth valuations in *V* according to the following two correctness clauses:[5]

$$(C1) \qquad v_c(+\phi) = 1 \text{ if and only if } v(\phi) = T$$

$$(C2) \qquad v_c(-\phi) = 1 \text{ if and only if } v(\phi) = F$$

In effect, (*C*1) is standard, and (*C*2) says that a valuation is *correct* for a sentence prefixed by the rejection sign just in case the prefixed sentence is false. Correctness valuations can now be used to define a notion of validity for signed arguments, in parallel to the standard, unsigned, definition of validity:

(VALIDITY) For a set of signed sentences Γ, and signed sentence *α*, $\Gamma \models_c \alpha$ if and only if for every correctness valuation v_c, if $v_c(\beta) = 1$ for all $\beta \in \Gamma$, then $v_c(\alpha) = 1$ as well

For this more general validity relation $\models_c$, the bilateralist natural deduction system of chapter 3, which gives rise to the bilateralist provability relation, $\vdash_b$, is both sound and complete (see this chapter's appendix for a proof).

By the completeness of our bilateralist system for the logical constants and our assumption that the standard semantics is extensionally adequate for capturing natural language logical notions, it follows that the meaning-constituting rules for the logical expressions of our language suffice for proving every single logical truth and establishing any particular logical entailment.

I mention these facts about completeness explicitly, but the inferentialist metasemantic principles accepted in chapter 3 swiftly led to an informal version of the completeness theorem. Say that a given schematic logical inference is valid in a language

L—$\Gamma \models \phi$. Then by Totality, the rule having each member of Γ as a premise and ϕ as a conclusion is derivable from the basic, meaning-determining rules of *L*—Γ ⊢ ϕ. In a sense, Totality corresponds to completeness, the MVC corresponds to soundness, and MAC allows the scope of these principles to apply as widely as possible.

Inferentialism makes this informal completeness result unsurprising. Logical truth matches provability from our logical rules, generally, just as the validity of a rule like modus ponens is automatic, given that modus ponens is a basic rule of our language. On the other hand, I don't see how hardcore realist views about logic can make much sense of this matchup between logical truth and our logical rules. For the inferentialist, logical validity is tied to our use, so it cannot outstrip our use. Since some collections of rules, like ours, allow that some sentences are provable using only our valid rules from no assumptions, certain sentences end up being true as a direct byproduct of the inferentialist account of meaning. This isn't a cheat; it's an explanation. Things that seem mystical on other approaches are nearly trivial on this approach. This is a good thing, not a bad thing.

It leads to a conventionalist explanation of any given logical truth. If "*p*" is provable from our basic logical rules without any premises, then according to MAC, these rules of ours will be meaning constituting, and so according to the MVC they will be valid. So, since "*p*" is provable from the empty set using these rules, then, trivially, "*p*" will be a necessary truth in a standard formal system. In the bilateralist system, similarly, what matters is whether +ϕ or -ϕ is provable from the empty set; ϕ will be a logical truth if the former is provable and a logical falsehood if the latter is provable. Both logical truth and logical falsity are wholly explainable in terms of the logical rules of our language. Of course, there might be other senses in which the truth of a particular logical truth can be "explained" (see 7.II). The fact that it is raining ensures, given our conventions, that "either it is raining or it is not" is true. In most contexts, I prefer not to call this an "explanation", but in any case, it doesn't conflict with conventionalism. And my discussion here shows that a conventionalist explanation is always available, for both particular

and general truths, and without tendentious metaphysical assumptions, no matter our evidential situation.

We can fully explain, on conventionalist grounds, the facts concerning validity in our own language. This means explaining *why* sentence ϕ is a logical truth, logical falsehood, or logically contingent, and why rule *R* is valid or invalid. The source of logical truth and validity, the only source in fact, is the automatic validity of our basic, meaning-constituting rules. This is absolutely central, so I'm going to further illustrate how these explanations go for specific sentences in our own language before generalizing the explanation to arbitrary languages. (I'd rather beat a dead horse than leave important things unclear.)

Why is every sentence in our language of the form $\ulcorner \phi \vee \neg\phi \urcorner$ true? Because every signed sentence of the form $\ulcorner +(\phi \vee \neg\phi) \urcorner$ is provable using the basic rules of our language alone. Why is each sentence of that form *logically* true? Because the proof of each sentence of that form uses only the logical rules of our language. Why is every sentence of the form $\ulcorner \phi \wedge \neg\phi \urcorner$ false? Because every signed sentence of the form $\ulcorner -(\phi \wedge \neg\phi) \urcorner$ is provable from the basic rules of our language and is thus false. Why is each sentence of that form logically false? Because proofs of signed sentences of this form use only the logical rules of our language. Why is the sentence "*p*" logically contingent? Because "*p*" is neither provable nor refutable using only the logical rules of our language.

Why is an inference rule like modus ponens ($\phi, \phi \rightarrow \psi \vdash \psi$) valid in our language? Because a version of modus ponens is (let us presume) a basic rule of our language and so is trivially valid.[6] Why is a more complicated rule like modus tollens ($\neg\psi, \phi \rightarrow \psi \vdash \neg\phi$) valid in our language? Because its validity can be derived using the basic rules of our language and is thus valid as a derivative rule. Why is a rule like affirming the consequent ($\psi, \phi \rightarrow \psi \vdash \phi$) invalid in our language? Because it is neither a basic nor a derivative rule of our language and so given its status as a logical inference and the Totality principle, it is not valid. All such questions can be given analogous answers; and all such answers make substantive appeal only to the rules of our language, that is, *to our linguistic conventions.*

So far my discussion has been aimed at the natural language (English) in which I write and you read. But what was said in the preceding would go just as well for *any other language we may encounter.* The inferentialist metasemantic principles introduced in chapter 3 are constitutive principles of linguistic meaning rather than parochial claims about our own language. If Martians speak a language adequately modeled by intuitionistic logic, then we can explain why sentences of the form ⌜ $\phi \vee \neg\phi$ ⌝ are not logically true in the Martian language by pointing out that not every sentence of this form is provable using the basic rules of the Martian language and then appealing to Totality. To say this, of course, is not to suggest that every sentence of this form (or even *any* sentence of this form) is refutable using the rules of the Martian language. The conventionalist explanation of logical truth and logical validity that I'm offering is not parochial. It is applicable to any language that we might encounter, of whatever complexity (but see section III for a caution on this score—related matters will be discussed in detail in chapter 5).

Not only do these points provide a conventionalist account of logic in the sense of chapter 1's Logical Conventionalism principle by explaining logical truth, logical falsehood, and validity, but they also explain the deep *necessity* of logical truths. Why is ⌜ $\phi \vee \neg\phi$ ⌝ *necessarily* true? Because ⌜ $+(\phi \wedge \neg\phi)$ ⌝ is provable using our logical rules, and these rules are all basic, meaning-determining rules that are, by the MVC, automatically valid. And as discussed in 3.I, the type of validity in question is strong validity; it requires necessary truth-preservation in the strongest sense. So it follows from this that ⌜ $\phi \vee \neg\phi$ ⌝ is not just true, but necessarily true in the strongest sense there is. An analogous explanation can be given for any necessary falsehood, as well.

Care must be taken so that these points about necessity are not misunderstood. Our logical rules apply in all situations that we can coherently describe and reason about, including all counterfactual scenarios.[7] But it must be stressed that this *does not* mean that in any possible scenario, we (or our counterparts in that scenario) have adopted rules of inference that allow ⌜ $+(\phi \vee \neg\phi)$ ⌝ to be proved with no undischarged assumptions (an objection to conventionalism

along these lines will be discussed in 7.I). In some counterfactual scenarios we (or our counterparts) will have adopted different linguistic conventions. This is not a problem for the conventionalist. My claims about necessity concern only what we express in our current language, not the language we (or our counterparts) may speak in some described but non-actual scenario. In this way, logical necessity is *projected onto* the counterfactual scenarios we dream up, *rather than found within them*. Logical necessity too, is a shadow cast by our practice. A shadow that falls upon any counterfactual scenario we are able to dream up.

A related potential confusion involves a failure to distinguish between (*a*) the necessity of a given truth and (*b*) the truth of a given necessity claim. Thus far, my discussion has only concerned (*a*) and not (*b*). Expressing necessity claims themselves involves using a necessity operator, standardly written as "$\Box$". Explaining why a sentence ϕ is necessarily true is distinct from explaining why a sentence of the form $\ulcorner \Box\phi \urcorner$ is true. So far I haven't discussed a full language that includes a necessity operator, and while I won't get into this in detail, it would be natural to think the standard necessitation meta-rule is at least partially meaning constituting for "$\Box$":[8]

This is a meta-rule, but meta-rules can be basic too (standard conditional proof is a meta-rule). With this principle (presented in a non-bilateralist language for simplicity) we can explain the truth of $\ulcorner \Box(\phi \vee \neg\phi) \urcorner$ by noting that $\ulcorner \phi \vee \neg\phi \urcorner$ is a theorem, and so by necessitation so too is $\ulcorner \Box(\phi \vee \neg\phi) \urcorner$.[9] And since $\ulcorner \Box(\phi \vee \neg\phi) \urcorner$ is provable using only the basic, meaning-constituting and thus automatically valid rules for the language, $\ulcorner \Box(\phi \vee \neg\phi) \urcorner$ is true. In general, when we have established that ϕ is a logical truth, we can appeal to necessitation, which can be meaning constituting by MAC. Then we can use the MVC to explain the truth of $\ulcorner \Box\phi \urcorner$ without any added difficulties. Logical necessity raises no special puzzles for conventionalists, even with all relevant distinctions noted.[10]

Quite generally, unrestricted inferentialism explains the facts about logical truth, logical falsity, logical validity, and logical necessity in any given language, and does so wholly in terms of the linguistic

conventions of the language in question. In other words: unrestricted logical inferentialism leads directly to logical conventionalism. Better put: unrestricted logical inferentialism is a form of logical conventionalism. But in order to fully secure this crucial point it is worth taking the time to establish the conventionalist credentials of the explanations I've been offering.

II Are These Really *Conventionalist* Explanations?

It might seem questionable whether we truly have a conventionalist explanation in the sense required by the Logical Conventionalism principle. The explanations in section I made appeal to the MVC and MAC, as well as Totality (to ensure generality), with background appeal to Logical Inferentialism throughout. And these, being generally applicable metasemantic principles, go beyond the particular conventions of any given language. In order to defend my approach as fully conventionalist, my use of these principles needs to be accounted for.

In reply, the central point is that—as will be recalled from 1.III's discussion of explanations—laws and general principles might not need to be part of ordinary explanations. Accordingly laws and general principles of metasemantics, like Logical Inferentialism, the MVC, Totality, and MAC need not be thought of as being among the *explaining facts* in our explanations. The only things included among the explaining facts are, appropriately enough, facts concerning the basic rules of language that the relevant group of language users have settled upon. In other words: facts about the relevant language's linguistic conventions. For this reason, the explanations I have given are fully conventionalist, despite the role that general metasemantic principles played in my discussion.

I think this is untendentious. But even if it is denied, so that my explanations include metasemantic principles, this doesn't trivialize my "conventionalist" approach, since everybody else still thinks that *the explanation I'm giving fails*.[11] That is to say: virtually everybody else thinks that logical truths are connected to the world in a more

substantial fashion than I do, regardless of whether a principle like the MVC is or is not part of my explanation. As such, criticism along these lines is, at bottom, merely terminological. It is also misbegotten, since if my theory of logic doesn't count as a brand of conventionalism, then no other view does either, *even the views of historical conventionalists.*[12]

Being a fully conventionalist explanation means, in particular, that an explanation of the truth of "the world is flat or it is not" does not need to appeal to any fact about the world, including the supposed fact that the world is flat or it is not. That the sentence "the world is flat or it is not" is provable using the basic rules of English (we're assuming) and that the basic rules of English are valid and thus necessarily truth-preserving is all that is needed to explain the truth (and, indeed, the necessary truth) of this sentence. This contrasts sharply with a sentence like "snow is white": for this sentence, the rules of language only get us to the point of saying that *if* snow is white, then "snow is white" is true, but this doesn't account for the truth of the sentence. For that, the world's contribution must also be factored in. This difference is absolutely central to the distinction between logical claims and standard empirical claims. Logic leaves no room for the world, so to speak.[13] Taken simply, there is room for confusion about what this involves, so my discussions in 4.V, 4.VI, and 7.II attempt to add some nuance.

Another prima facie worry concerns circularity. More specifically the worry is that my conventionalist explanations tacitly appeal to logic and, in so doing, become viciously circular in some diffuse manner. As one (not fully adequate) answer to this, recall from chapter 1's discussion of explanation that the relationship between the explanation and the fact to be explained isn't merely that of logical entailment. But there is a deeper reason that this worry is groundless. An explanation's entailing the fact to be explained is not the same thing as an explanation making substantive appeal to some unexplained logical fact. More explicitly: even if it is a necessary condition on *E*'s explaining *p* that *E* (suitably supplemented) logically entails *p*, that doesn't mean that the fact that *E* logically entails *p is itself part of the explanation of p*. Instead, that this entailment holds is a necessary condition on explanation that is

external to the particular explanation itself and so is not among the explaining facts. Imagine how silly it would sound if we attempted to add facts about logical relationships to ordinary causal explanations. But if it is silly in that type of case, then it is just as silly here. There is no problematic logical circularity lurking in the conventionalist explanations I've offered.[14]

Unrestricted inferentialism leads to conventionalism. But does *restricted* inferentialism also lead to conventionalism? The issue is a bit subtle. Even extremely permissive versions of restricted inferentialism require that the basic rules of our language meet some external condition *C* in order to be meaning constituting. This condition might be mere consistency, or conservativeness over some reasonable base theory, or harmony, or something else. The points I will make apply to all such constraints, so I will continue to speak in generalities. That our rules meet condition *C* is itself a particular matter of fact that is not a matter of convention. So if this fact is itself part of the restricted inferentialist's explanation of the truth (or logical truth, or necessity) of some sentence of our language, then we might worry that the provided explanation is not a purely conventionalist explanation.

But it isn't obvious that the fact that our rules satisfy *C* needs to be part of our explanation of logical truth. Another option is to see this fact as an external, enabling condition on our explanations. The thought here is that when things go well—that is, when our rules meet *C*—then the rules alone are a sufficient explanation. If this is right, then it doesn't matter that the rules meeting *C* is a non-conventional fact, since this fact will not be part of our explanation of any given logical fact. This move on behalf of the restricted inferentialist is analogous to my argument above that an explanation's logically entailing *p* need not be part of the explanation. There is something in this analogy, but it is somewhat strained. There is a difference between necessary conditions on explanations themselves and enabling conditions that allow particular facts to explain other facts.

The principle of Logical Conventionalism required a *full* explanation of the logical facts in *any* language. The demand for a "full" explanation suggests that we need to include *all* of the relevant

particular facts in our explanation. Otherwise, we would have only a partial explanation. It is plausible that this demand requires us to make note of all enabling conditions as well. A full explanation of why I'm allowed by law to vote in the United States is not given merely by citing the fact that I am an adult citizen of the United States. It also needs to be added that I meet the other conditions for voting as well—for example, not being a convicted felon. So perhaps restricted inferentialists must include the fact that our rules meet *C* as part of their full explanation of logical facts.

The question then becomes whether this fact, itself, goes beyond our linguistic conventions. For some versions of *C* that we can imagine, the answer is clearly "yes". For instance, if the condition is that the rules align with reality's objective logical structure, metaphysically speaking.[15] Or if *C* is that the rules are favored by God. These constraints are clearly "external", in that they can't simply be read off of the rules themselves, even in the context of the language as a whole. But conditions of this kind violate the spirit and probably the letter of Logical Inferentialism. Most of the constraints actually proposed by restricted inferentialists are internal to the conventions themselves. So if *C* is that the rules are consistent with the other rules of the language, and consistency is a purely syntactic notion—it is not the case that both ϕ and $\ulcorner \neg\phi \urcorner$ are provable using these rules—then the fact that the rules meet the condition *C* is merely a fact about the conventions themselves. When the restricted inferentialist condition is internal in this sense, rather than external, then restricted inferentialism can plausibly be extended to conventionalism where it applies.

Still, we wouldn't want to say that the unrestricted inferentialist and the restricted inferentialist offer *the very same* explanation of the logical facts, even when both explanations go through. The unrestricted inferentialist's explanation of logical truth in language *L* is merely that the *L*-speakers have adopted certain linguistic conventions, while the restricted inferentialist's explanation is that the *L*-speakers have adopted certain linguistic conventions *and* that these conventions meet the conditions for being meaning conferring. It is merely a verbal matter whether this is a difference in the explaining facts or a difference in the enabling conditions required,

but either way, the full explanations differ. And the unrestricted inferentialist's account is simpler and cleaner, even if both theories count as versions of logical conventionalism.

The discussion thus far has concerned the letter of conventionalism, not its spirit. Historically, conventionalism has been associated with the idea that we are utterly free to choose our language as we see fit. This was famously summed up in Carnap's principle of tolerance, and it was very widely believed by conventionalists.[16] A suggestive illustration is provided by considering an alternative language. Imagine a Martian language that uses classical logic. Imagine for the moment that we are intuitionists who accept a version of restricted inferentialism according to which classical negation fails to be meaningful. From our perspective, many supposed Martian logical truths will be utterly meaningless. Yet the Martians follow the rules of classical logic. Even for anti-classical restricted inferentialists, it is clearly possible for there to be such a language. So the Martians will follow ancestrally non-empirical, topic neutral rules that fail to be "logical" because they fail to confer a meaning. This possibility is deeply at odds with the spirit of conventionalism, if not its letter.

Only unrestricted inferentialism leads to a conventionalism with the freedom and uniformity that has historically been associated with the view. So even if and when restricted inferentialism ends up as "conventionalism", it is an impoverished conventionalism. It has little hope of applying uniformly to alternative languages, and it blocks extending the account from logic to mathematics, as I will do in part III. Only unrestricted inferentialism leads to a canonical form of logical conventionalism. For all of these reasons, for conventionalists, unrestricted inferentialism is much to be preferred. Accordingly, when I talk about logical conventionalism, I am talking about the pure, unrestricted version.

The next chapter will explore and clarify the freedom and pluralism that conventionalism leads to. Before that, the next two sections discuss two general worries that some might have about logical conventionalism, one technical and one philosophical. After dealing with these concerns, the final three sections of this chapter set out

my conventionalist approach to the metaphysics and normativity of logic, respectively.

III The Role of Semantic Completeness

Take a particular logical fact, for example, that "either the *Tractatus* was published before 1950 or it was not" is a logical truth. The conventionalist explanations of why this sentence is true, necessary, and logically true don't need to rely on anything like Totality or the completeness theorem. The role of those principles is more general: they ensure us that no logical truths are left unexplained.[17]

More formally: if we assume that logical truth is extensionally characterized in a language *L*, semantically, by $\models$, then the conventionalist account requires a proof relation in *L*, $\vdash$, spelled out in terms of finite proofs using the rules of language, that suffices for capturing everything captured by $\models$. If completeness fails, then there will be some set of sentences Γ and sentence ϕ, for which $\Gamma \models \phi$ but $\Gamma \nvdash \phi$. This requirement immediately raises a number of serious concerns about incomplete extensions of and incomplete alternatives to classical logic.

A famous and essentially incomplete extension of classical logic is second-order logic with its standard semantics.[18] This logic admits of no recursive, complete proof procedure, and so (seemingly) no rules of inference that we could possibly learn can ever fully match up with the semantics. In addition to incomplete *extensions* of classical logic like this, there could also be *alternatives* to classic logic that are incomplete. However, issues concerning alternatives are complicated by the interaction of the logic of the metatheory with the logic of the object theory, and so often aren't as straightforward to assess (some related technicalities will be discussed in part III; see 11.I and all of chapter 12).

To illustrate: as classical logicians, we can provide a possible worlds semantics for intuitionistic logic and prove a completeness result for said semantics, but such a proof would be of dubious worth to a real intuitionist, since our completeness proof itself uses logical

rules that intuitionists don't accept.[19] Generally though, I am committed to tying logical validity and truth in *any* language to the rules of inference followed in that language. I don't think that this is a deeply partisan assumption, even if a completeness proof is difficult to carry out for a given alternative logic in a metatheory using said logic.

A related issue is that Gödel famously showed that any consistent recursive system containing a rudimentary amount of arithmetic will be negation incomplete. That is: there will be sentences G in the language of the theory T for which T ⊬ G and also T ⊬ ¬G; but by bivalence either "G" is true or "¬G" is true, so in any recursive system of arithmetic there will be true claims unprovable in that system.[20] Gödel's theorem holds even for weak first-order theories like Robinson Arithmetic (Q) and weak non-first-order theories like Primitive Recursive Arithmetic (PRA), and since a great many mathematical theories can relatively interpret Q, incompleteness holds for many non-arithmetical theories as well.[21] This means that standard assumptions about mathematical truth, together with Gödel's theorems, make *mathematical* conventionalism on the model of section I a non-starter. Nothing like Totality is true even for basic arithmetic, at least if it is interpreted as above, where the validity of derivative rules is guaranteed by provability using the basic rules.

All of this is exactly right. If these cases can be given a conventionalist explanation it won't be on exactly the same lines as the approach in section I. Despite this, I think that inferentialism and conventionalism can be extended to cover such cases. I will argue for this in part III. To set up the discussion in that part, chapter 8 will spell out the challenges involved in extending conventionalism from logic to mathematics in much greater detail. A central point will be whether Totality must be given up, or if some alternative method of determination can be used while staying within the bounds of scientific naturalism (1.IV).

All of these point will be addressed in detail in part III. But for the remainder of part II I will focus on logics for which the issues of incompleteness don't arise. For such logics—and there are many—the simple version of logical conventionalism outlined in section I will

be directly applicable, in some form or other. The philosophical issues concerning these alternative logics and logical pluralism will be discussed in detail in the next chapter.

IV Conventionalism versus Deflationism

A philosophical worry about my inferentialist/conventionalist explanations of logical facts concerns whether they are really needed given the apparent existence of even simpler, non-conventionalist explanations of the very same facts. I have in mind here so-called *deflationary* approaches to truth and meaning which are often contrasted with and taken to be competing with *inflationary* approaches—including inferentialist approaches.[22]

One particularly popular deflationist view endorses *disquotationalism* about truth, holding that, in some sense, all that there is to the semantic property of truth (if it even counts as a property) is contained in the disquotational *T*-schema:[23]

$$(T) \qquad T[\phi] \text{ iff } \phi$$

Brackets here are a name-forming device of an appropriate kind. Deflationists often argue that disquotational principles like this can provide satisfying answers to the metasemantic questions that I have answered using inferentialism. To illustrate, in an influential discussion, Hartry Field invites us to consider:

> …the problem of saying what determines the 'referent' of a logical operator like 'or' or 'not' or 'all'; or to put it less Fregeanly, the problem of saying what determines the contributions that such a logical operator makes to the truth conditions of sentences that contain it.[24]

Field claims that deflationism provides an answer to these questions. To illustrate the supposed deflationist answer to this type of question, consider the case of natural language conjunction, as

expressed by “and”. In this case, deflationists offer something like the following argument, which uses only instances of (*T*) together with basic logic:[25]

1. *T*[*ϕ* and *ψ*] iff *ϕ* and *ψ*
2. *T*[*ϕ*] iff *ϕ*
3. *T*[*ψ*] iff *ψ*
4. *So*: *T*[*ϕ* and *ψ*] iff *T*[*ϕ*] and *T*[*ψ*] (1,2,3)

The conclusion, which follows from the three premises by standard classical reasoning, seems to express the fact that an “and”-claim is true just in case the two conjuncts are both true, and this seems to be an oblique way of stating that “and” refers to (or means) the standard truth function for conjunction. For this reason, deflationists claim that arguments like this one answer metasemantic questions about logical expressions.[26]

Given this, deflationists are likely to react to inferentialist accounts of logical constants by asking why they are even needed. The deflationary answers to the very same questions seem—to them—better and simpler than the inferentialist answers. But this is mistaken, I think. Despite how it seems at first glance, deflationism and inferentialism are not answering the same questions. I have no issue with truth deflationism in most forms, but I don’t think that deflationism manages to answer the metasemantic questions that I’m interested in, despite what Field and other deflationists have sometimes suggested.[27]

Somewhat related points have been made in the critical literature on deflationism, often when discussing the contrast between sentential versions of deflationism that appeal to the *T*-schema with propositional versions that appeal to a related, propositional equivalence principle.[28] But the precise point has largely been missed in the literature on meaning because there are two distinct but related questions that can be put using almost the same words. One is *semantic*: it aims to answer the questions “what does ‘and’ mean/refer to?” or “how does ‘and’ contribute to the truth conditions of sentences?”. I agree that the deflationary argument answers these semantic questions. I think the argument might also answer the

question in Field's second statement of the problem—"what determines the contributions that such a logical operator makes to the truth conditions of sentences that contain it"—at least if *that it means this particular truth function*, is an acceptable answer to that question. In other words, the argument above answers Field's questions if they are understood semantically. But I don't think the semantic understanding is the most natural understanding. In any case, the semantic questions aren't the questions I have been trying to answer here.

The questions I am interested in are *metasemantic*, rather than semantic. In the case at hand they concern how our word "and" manages to mean that familiar truth function, and what explains that it does. This reading corresponds more closely to Field's first statement of the problem. While Field seems to take both statements to be equivalent, the second has a semantic reading as well as a metasemantic reading, but the first arguably does not. There is no plausible reading of "what determines the referent of 'and'?" upon which *that it means this particular truth function* is an acceptable answer. And while the deflationary argument provides a satisfying answer to the semantic question, it is powerless against the associated metasemantic question. If someone was puzzled about how our word "and" managed to mean what it does, the argument above would not alleviate their puzzlement in the slightest.

As chapter 3 discussed at length, the metasemantic question (that I claim deflationism is powerless against) is the one that inferentialism is needed to answer. Similarly for questions about *why* a particular rule of inference is valid in our language. Supposed deflationist answers would, once again, miss the metasemantic point at issue. The deflationist argument provides an answer to first-order semantic questions like "what does 'and' mean?" and "what truth function does 'and' express?", but it sheds no light on metasemantic questions like "*why* does 'and' express that familiar truth function?" and "*how* does 'and' manage to express the truth function that it does?".

Failure to distinguish between semantic and metasemantic questions can only lead to confusions, in both directions. On the one hand, failure to answer the metasemantic question is not a cogent

reason for rejecting deflationism about truth, understood semantically. And on the other hand, the semantic success of deflationism does not undermine more involved metasemantic theories, including metasemantic inferentialism.[29] For this reason, deflationary theories of truth and other semantic notions are fully compatible with inferentialist approaches to metasemantics, *even as applied to predicates like "true"*.[30]

Seeing deflationism as competing with inferentialism is a mistake. Deflationism is not a metasemantic theory and nor is it metasemantically relevant to the questions I am interested in here. However, there are important affinities between my conventionalist theory and the notoriously attractive minimalist metaphysics associated with deflationism, and these are worth setting out in detail.

V The Metaphysics of Logical Facts

What *metaphysical* picture of logic does my conventionalist theory leave us with? Here I am using the word "metaphysical" in the very broadest way possible, so that even logical positivists have "metaphysical" views, in the relevant sense. In fact, the conventionalist quotes paraded in 1.I put some of these positivist metaphysical views on display. In those heavily rhetorical quotes, several conventionalists suggested that logical conventionalism led to some form of logical *non-factualism*. The idea being that the conventionalist theory shows that there are no logical facts.

Is the version of conventionalism developed here a version of logical non-factualism? The answer to this question depends upon what exactly is meant by "logical facts" and "non-factualism". There is a metaphysically lightweight, everyday notion of a fact according to which it is a fact that *p* just in case "*p*" is true. According to this notion, my version of conventionalism is fully factualist. All of the logical truths are true, so since "either the Red Sox are the best team in baseball or the Red Sox are not the best team in baseball" is true then it is a fact that either the Red Sox are the best team in baseball or the Red Sox are not the best team in baseball. In addition, as

we've seen (1.III), I've already helped myself to a metaphysically lightweight and ultimately dispensable notion of "logical facts" in formulating the target of conventionalist explanations. These "facts", even if "about" the world in a thin sense, are utterly empty.[31] Still, in this thin sense, of course there are logical facts, but only in a sense reminiscent of deflationary views of facts (and truths).[32] In being "about" the world, we might sometimes say they are made true by (parts of) the world. This doesn't undermine conventionalism (see 7.II).

When making their non-factualist claims the historical conventionalists weren't concerned with this lightweight notion of fact, but instead with a more robust, inflationary, and heavyweight notion. Say that some true sentence expresses a *fact* in the heavyweight sense just in case it holds objectively, is a key part of reality, and is fully independent of all of our cognitive and social practices. According to conventionalism, logical truths are simply byproducts of our using language in the way that we do—given the way that we use conjunction and negation symbols, we find that every instance of the schema $\ulcorner \neg(\phi \wedge \neg\phi) \urcorner$ is provable and therefore true. The logical truths are, in this sense, reflections of the rules of our language. So the logical truths and logical validities are not completely independent of our cognitive and social practices and therefore, using the metaphysical notion just adumbrated, there are no logical facts. Of course, even for conventionalists, there is still a sense in which logic is independent of our practices, and logic is still necessary in the broadest sense. It is just that there are different ways to understand counterfactuals concerning alternative conventions in our language, and conventionalists admit an important role for our practices in explaining logical truth and validity (see 7.I for further discussion).

Yet on all conventionalist views, logic is not part of the superstructure of the world awaiting discovery. Negation isn't the type of thing that is *out there* in the world like tables, electrons, and Oreo cookies. The laws of negation aren't *describing* some ethereal realm of ghostly logical facts. If God so made the world, there was no day on which he said, "Let there be negation"! And having already said, "Let there be light!", he did not then have to add, "It is not the

case that there is no light". Logical truths are semantic reflections of the syntactic theorems of our language, so yet again, they are merely reflections of rules, shadows of syntax.

Does this mean that conventionalism entails that logic is subjective or that we *make* or *create* the logical truths? No, of course not. It would be misleading to characterize conventionalism by saying that we "make" or "create" logical truths. Logic is not simply a projection of our ephemeral cognitive whims. Not everything goes and there is a real distinction between good reasoning and bad. Some people are poor reasoners and we all make foolish mistakes sometimes. It is badly misleading to say that we "make" or "create" logical truth, but it is just as misleading to say that we "find" or "discover" logical truth. Conventionalism avoids the metaphysical weirdness of a radically subjectivist approach to logic, but it also eliminates the need for a spooky heavyweight metaphysics of logic.

Logical conventionalism, built on a solid foundation of logical inferentialism, avoids both the Scylla of logical hyper-realism and the Charybdis of extreme logical subjectivism, leaving us with a moderate metaphysics of logic that does justice to the insights of both extreme camps while avoiding their excesses. This concerns the logical truths, but there are also logically necessary connections between contingent facts about the world. These connections seem to be objective and mind independent, so conventionalists must account for them, as well.

VI Logically Necessary Connections

According to the previous section's metaphysics of logic, there is at least one good sense in which there are logical facts, and there are even senses in which these logical facts are objective and mind-independent too. But on the other hand, there are also more robust notions of factuality according to which conventionalists will deny the existence of logical "facts". Under no circumstances do conventionalists accept logical facts as substantive, independent parts of reality.

Some may feel that the moderate view of the metaphysics of logic is fine for facts about logical *truths*, but does it apply to logically necessary connections between non-logical facts, as well? To address this, it will be helpful to consider a particular example. Take the two sentences, "Jared's copy of *Logical Syntax* is green" and "either Jared's copy of *Logical Syntax* is green or it is red"—both of these sentences are true but they easily could have been false (since there are non-green and non-red copies of *Logical Syntax* that I could have acquired instead).[33] Each of these sentences expresses a contingent fact about the world that has nothing directly to do with logic. Yet there is a logically necessary connection between these two contingent facts: if my copy of *Logical Syntax* is green, then necessarily, my copy of *Logical Syntax* is either green or red.

I don't think that this type of connection needs to be considered factual in a heavyweight sense. The second sentence is a disjunction, and disjunction is a logical notion. Logical atomists like the middle Bertrand Russell and the young Ludwig Wittgenstein famously denied that there were necessary connections between atomic facts, which were all contingent and independent.[34] But even the atomists allowed logically necessary connections between sentences expressing contingencies, provided that at least one of them was non-atomic.[35] In this book I deny connections of *logical* necessity between atomic facts, but am not trying to argue that all necessary connections are logical.[36]

The logically necessary connection between "Jared's copy of *Logical Syntax* is green" and "either Jared's copy of *Logical Syntax* is green or it is red" is not a robust feature of the world awaiting our discovery. But this does not mean that it is merely a projection of our desires or whims. There is instead a very real sense in which this necessary connection is objective—it must be recognized, and in this form, by anyone who speaks our language. There is also a very real sense in which the connection is mind-independent—the necessitation relation between these and other contingent facts is counterfactually robust, holding no matter what we do or think.[37] In this way, conventionalists can admit that necessary connections between contingencies are real and that, in a perfectly standard

sense, they are both objective and mind-independent. There is a sense in which the fact that my copy is green “makes it true” that my copy is either green or red. And this fact also “makes it true” that my copy is either green or not green. But even applied to logical truths, this doesn’t undermine conventionalism (as noted in 4.I above, see 7.II for further discussion).

It may be an objective fact that my copy of *Logical Syntax* being green logically necessitates that my copy is either green or red, but it is not a fact in exactly the same way that contingent facts—like the fact that my copy of *Logical Syntax* is green—are facts. Contingent truths did not have to be the case and they can easily be imagined to have been false. By contrast, a fact about a logically necessary connection between two contingent facts simply could not have been otherwise, in any sense. This is not just to say that these connections themselves hold necessarily, but that they are built into the very fabric of our language. Given the rule of “or”-introduction:

$$(\vee I) \quad \frac{\phi}{\phi \vee \psi}$$

Any fact expressed in our language by a disjunction is entailed by any fact expressed by one of its disjuncts. The necessary connection in the world between two such claims is a shadow cast by this rule. This is not to say that the shadow is not *real*, just as it is not to say that the necessary connection does not exist. The connections are real and they do exist. But they are insubstantial, wholly different in character from contingent truths which, though stated in our language against a backdrop of our conventions, are not automatically true as a byproduct of those very conventions.

Conventionalists don’t need to deny the existence of necessary connections in the world, even between contingent facts. Nor does the conventionalist allow these connections to be real only because we have put them there ourselves with our whims and desires. Instead these connections *must* be there, given how we use language. This doesn’t mean, however, that the connections themselves simply *are* our linguistic rules. That would be a

use/mention confusion. Our rules are our rules, and the logically necessary connections are their unavoidable byproducts, explanatorily speaking. When creating the world, God didn't have to decree that it was a fact that my copy of *Logical Syntax* being green necessitated my copy being green or red. Logically necessary connections instead come along for the ride; they are a metaphysical free lunch.[38]

The moderate metaphysical picture of the previous section applies even to wordly necessary connections between contingencies, and it does so without collapsing into non-factualism or subjectivism. In light of these points about logical metaphysics, it might be wondered whether these subjectivist-adjacent elements of conventionalism mean that conventionalists cannot adequately account for the all-important *normativity* of logic.

VII The Normativity of Logic

At a basic level, the normativity of logic is explained in the exact same way by every plausible approach to the philosophy of logic: logically valid rules are truth-preserving in the strongest sense. They are strongly valid. So in any situation, reasoning according to these rules, but not invalid rules, will lead only from true beliefs to further true beliefs. They will never mislead you under any hypothesis you can dream up. In this they are unlike inductive or abductive rules. Even when those rules are very likely to lead from truths to truths, there is still a small risk of them leading you astray. Not so with logical validity. Thus at bottom, there is no special puzzle about the normativity of logic, given that logical rules are necessarily truth-preserving in the strongest possible sense.[39]

Of course, nothing in this picture requires assuming that non-logical—even non-valid—principles aren't normative in a similar but weaker sense. And this explanation of the normativity of logical rules carries over to valid but non-logical rules as well. In general, principles and rules that give rise to true beliefs can give rise to various norms of reasoning in a familiar and untendentious, broadly

externalist fashion. One component of this will be explored in detail in chapter 6's discussion of the epistemology of logic.

An externalist perspective is perfectly compatible with there also being various other norms in force at the same time, of relevance or whatnot. The normativity of logic is tied to correctness in reasoning, where correctness involves accepting only true claims and rejecting only false claims. Of course, there is no rational requirement that *all* true claims be accepted or that *all* false claims be rejected. Related points have been made at length by Gilbert Harman in his influential discussions of "reasoned changes in view".[40] Harman pointed out that many straightforward construals of the role of logic in reasoning violate an intuitive "clutter avoidance" principle according to which one shouldn't clutter one's mind with things one isn't interested in. Instead, having a reason to add a belief (or disbelief) requires a need or desire for attitudes about the matter at hand. A broadly externalist perspective on the normativity of logic doesn't conflict with this at all, but something should be said about the proper form of logical norms. I'll focus on the case of modus ponens, for simplicity, but the points I will make are entirely general.

Perhaps the most natural (but ultimately mistaken) way to accommodate clutter avoidance, normatively, is by seeing the normativity of logic as directly involving norms of *permission*, rather than norms of *obligation*.[41] A logical rule like modus ponens is normative not in that whenever you accept the premises of an instance of modus ponens you are thereby obligated to accept the conclusion. It is rather, according to this idea, that whenever you accept the premises of an instance of modus ponens, you are thereby permitted to accept the conclusion. The permission-approach does get around clutter avoidance, but it has a number of problems.

One is that you are not automatically all-things-considered permitted to accept the conclusion of a modus ponens argument when you accept the premises. There are various other norms in play that may bar acceptance of the conclusion in a given context. Perhaps a madman has captured my family and credibly threatens to execute them if I ever again use modus ponens to form a new belief. In that type of situation, while we might say that I am rationally or

epistemically permitted to form a new belief via modus ponens, there is a clear sense which I am not permitted to, when everything is taken into account. And it isn't even obvious that we are always rationally or epistemically permitted to accept the conclusion. Consider a case where I have accepted the premises for bizarre or silly reasons. In this type of case you may think it irrational for me to accept the conclusion, and so may deny that I am rationally permitted to do so.

Perhaps even in this kind of case I have a *wide-scope* permission —to either reject one of the premises or accept the conclusion, but this seems far too weak to capture the full normative force of a logical rule. This points to another issue with the permission-approach. While I have been highlighting ways in which it is sometimes too strong, it is also sometimes seems too weak. We tend to think that modus ponens does more than simply give us permission to accept the conclusion upon accepting the premises. It also seems to somehow rationally push us toward acceptance of the conclusion. So the obligation-approach is far too strong and runs up against clutter avoidance, and the permission-approach is sometimes too strong and sometimes too weak. We need a completely different approach.

The way out is to express the normative force of logical rules in different terms. My preferred strategy is to say that accepting the premises of an instance of modus ponens gives (some) justification for accepting the conclusion (I stated this without defense in 2.IV). The important point about this way of talking is that having "some justification for" accepting the conclusion does not automatically mean that you are fully justified in accepting the conclusion, nor even that you are fully rationally or epistemically justified in accepting the conclusion. Just that you have some justification for doing so based on your acceptance of the premises.[42] An especially nice thing about this approach is that in some contexts, the justification you have on the basis of accepting the premises might give you permission to accept the conclusion, while in others it might not. In still other contexts, the justification you have might obligate you to accept the conclusion. Permission and obligation are a function of the overall justification you have in a given context, according to this picture.

Alternatively we might instead talk in terms of “reasons”, saying that the premises provide a defeasible reason for accepting the conclusion. I think that the “justification” approach and the “reasons” approach come to basically the same thing. Both approaches seem to overcome the defects of the simplistic permission and obligation approaches to logical normativity. Whether or not you are permitted to accept the conclusion upon accepting the premises involves factoring in everything about a given situation. And the same goes for your obligations.

This is a step forward, but it can’t be the end of the story. In addition to these positive norms, there are also negative norms associated with logical principles. Anyone who follows modus ponens is obligated not to accept the premises of an instance of the rule while at the same time rejecting the conclusion of that same instance. We can put this more naturally by saying that we are “not permitted to” do this, instead of saying that we are “obligated not to”. It is important to understand that this is very different than saying that if you follow modus ponens you are obligated to either reject a premise or accept the conclusion for every instance of the rule. That is far too demanding.

Putting these two points together with unrestricted inferentialism, a natural first-pass inferentialist account of the normative force of modus ponens is given by saying both that (*a*) following modus ponens and accepting the premises provides you with (some) justification for accepting the conclusion (not necessarily all-things-considered justification) and (*b*) when following modus ponens you are not permitted to accept the premises while also rejecting the conclusion, in any circumstances. Both (*a*) and (*b*) can perhaps be derived from more general norms concerning evidence and truth, against the backdrop of unrestricted inferentialism, but I won’t try to do so here.

Of course, fully disentangling all of the strands that make up our norms of rationality is out of the question here. I will discuss the epistemology of logic in great detail in chapter 6, but beyond that, I won’t pursue distinctively normative concerns further. Not because they aren’t interesting or important, but because at around exactly this point the issues no longer concern the philosophy of logic,

proper. Instead they concern normativity, more generally. By and large, logic is rationally normative because of its relationship to truth, where this relationship is sometimes epistemologically convoluted. Beyond this, I don't think there are any special problems about the normativity of logic, as such, for conventionalists. But this has been called into question.

VIII Rules and Roles

The central challenge that I want to discuss here is related to Kripkenstein's normativity challenge (discussed in 2.VII), though it long predates Kripkenstein, having originally arisen during the heydey of rule-based approaches to language.[43] Translated into modern terminology, the basic issue is that there is an apparent tension between the dual roles that inference rules play in inferentialism. On the one hand they play a constitutive role by implicitly defining expressions of the language in question, but on the other hand, they play a normative or regulative role. This key distinction between *constitutive* and *regulative* rules is, supposedly, incoherently blurred by the inferentialist/conventionalist picture I've been developing.[44]

The basic problem is straightforward: constitutive rules are part of what creates a given game, practice, or—in the case at hand—language. But regulative rules, which are what give rise to prohibitions and norms, can only regulate preexisting types of behavior. To illustrate the distinction: the castling rule in chess is constitutive of the game itself, but the rule barring chess players from yelling at their opponents during the game is merely regulative. The inferentialist/conventionalist picture seems to get into trouble by taking linguistic rules to be both constitutive and regulative at the very same time, in the very same fashion. If following these rules constitutes the linguistic practice in question, then how can it also be that these rules regulate moves within that same linguistic practice? Such regulation seemingly requires a pre-existing linguistic practice, with well-defined meanings already established. So the

conventionalist picture of the normativity of logic is incompatible with the very foundation of the view, or so goes the challenge.

I think this challenge is based on a distorting simplification of the conventionalist picture. Say that the rule of modus ponens is meaning constituting in our language, so that it is a constitutive rule of our language that implicitly defines the meaning of "if" and related terms. As stressed throughout chapter 2, the notion of inference rule-following is dispositional, rooted in the patterned use of language. As such, there will be many times when modus ponens functions like a regulative rule *locally*, since even those who follow modus ponens don't do so perfectly and invariably in every situation. This isn't incoherent at all, because only certain local patterns of acceptance and rejection are being regulated here, and these are, in fact, explanatorily rooted in preexisting behavior patterns, as discussed in chapter 2. There I claimed that acceptance and rejection and inference are themselves are rooted in dispositions to use syntactic items in certain ways but not others.

This is not to say that the behavior patterns themselves don't involve items with content, of course. It is just to say that the behavior patterns can themselves be directly regulated—don't use "if" that way, *et cetera*. So there is no real conflict between these rules being globally constitutive and, on occasion, locally regulative. Their local regulative function does not conflict with their constitutive function, since what is directly regulated isn't anything that presumes a priorly given specific meaning for "if" and other terms.

That also isn't to say that the inferences that are regulated are not movements of thought that involve operating on the meanings or contents involved, at least at some level of description (see 2.VI). This is especially important to note since recently a similar variant of the older challenge I've been discussing has gained some currency in the literature. It has even led one prominent inferentialist, Paul Boghossian, to publicly recant:

> …something that should have been obvious, but that is often lost sight of, including by me … is that reasoning is an operation on thought *contents* and not on symbols (that have content). That immediately implies that the usual ways of presenting programs of 'inferential role semantics' are confused—a logical constant's role in inference must be explained by its content; and its content cannot be explained by its role in inference.[45]

Here, instead of focusing on the regulative role of inference rules, Boghossian is focusing on the very idea that inferences are operations on contents. The problem is supposed to be that inferentialism explains content or meaning in terms of inference rules, which govern inferences, which presuppose content, so the inferentialist approach goes in a circle and so must be rejected.

Apart from Boghossian, many opponents of inferentialism have also been tempted by this and similar criticisms. They have recently been forcefully presented by Nick Zangwill.[46] But I think my discussion above already shows why this line of criticism is mistaken.[47] Take again the chess analogy from above. Surely in moving chess pieces according to the rules, we are operating with chess pieces, rather than with mere pieces of wood. In other words, they are *meaningful*, and nobody doubts that the meaning of a piece —its role in the game—is constituted by the rules of chess. The rules must be in place for a piece of material to play the role of a knight, for example, and this is not at all mysterious. Inferentialists think that something analogous is happening with language, so see no need to deny that inferences are movements of thought that are naturally explained as operating on meanings or contents. This descriptive claim is not in conflict with the explanatory claims of inferentialism. In fact, they are completely in harmony.[48]

I think the feeling of discord is induced by trying and failing to imagine a system of rule-governed operations on meaningful items being in place *before* meanings are generated. But this is a caricature of inferentialism. The rules and the meanings defined by the rules grow up together, walking hand in hand. Perhaps Boghossian's claim that reasoning is "an operation on thought contents and not on symbols (that have content)" is meant to draw

some other distinction. If so, I do not see what that distinction could be. I think that reasoning and inference are most naturally described as operations on symbols that have content (as discussed in chapter 2), and that what gives these symbols content is the overall pattern of their use, including the behavior patterns involved in inferences, dispositionally speaking. If we so desire, we can even talk about "reasoning" as operating directly on contents or propositions with no mention of symbols either in the world or in the head. Explanatorily though, I think that this sense of "reasoning" is usually derivative, but there is no special problem in talking this way, even for an inferentialist.

Inferences are operations on meaningful items, but the items have meaning because of the rules that are followed for their overall deployment. Inferentialism is not, and never was, committed to the idea that there must first be some senseless syntactically isomorphic "inference"-like activity in place temporally before meanings get generated. This is as silly for the human activity of inferring as it is for the human activity of playing chess. A word like "and" plays the conjunction role based on the rules we follow in deploying it, just as a little carved horse figurine plays the knight role in chess based on the rules we follow for deploying it. In neither case did the role have to predate the rules.

Likewise, as argued above, the semantically loaded regulative function of constitutive rules like modus ponens arises only in an indirect fashion. While these norms do involve assumptions about the meanings of the sentences involved, that is not a problem. The constitutive rules of language, in the form of basic, directly followed inference rules, give rise to various regulations and prescriptions derivatively. In chapter 6 I'll show exactly how this goes for standard epistemological norms. As noted previously, the logical norms of rationality are byproducts of the constitutive role of the rules and, more particularly, the relation of this constitutive role to validity and, thereby, to truth. Just as the constitutive rules of chess indirectly give rise to goals and sub-goals that define correct and rational chess play, so too do the constitutive rules of language indirectly give rise to goals and sub-goals that define correct and rational thought and language use (given background information about interests,

desires, and the like). Once again, this is not because the norms here are particular to logic, but instead because of the relation of constitutive rules of language to truth. A relationship which, for conventionalists, is *explained by* the constitutive role of the rules themselves, and so the normative function piggybacks upon the constitutive function. All in all, when viewed rightly, there is no conflict between the normativity of logic and logical conventionalism.

≈≈≈≈≈≈

UNRESTRICTED INFERENTIALISM LEADS to logical conventionalism. Far from being a weakness of the theory, this is a significant strength. A metasemantic theory of the meanings of the logical constants that grows naturally out of the most popular family of approaches to metasemantics also leaves us with a full, naturalistically acceptable explanation of the logical facts. The resulting theory is metaphysically satisfying, avoiding the problematic excesses of other approaches to logic. I will show in chapter 6 that conventionalism leads to a simple and plausible epistemology of logic that avoids the need for mysterious faculties of reason or intuition. All of this, together with the extreme theoretical simplicity and tidiness of the view, makes unrestricted inferentialism sound almost too good to be true. I will discuss whether it is too good to be true in detail in chapter 7 when I rebut several important objections. But first I need to discuss conventionalism's relationship with logical pluralism. We have our own logical rules, but Martians may have distinct logical rules. Surely the laws of metasemantics apply to them just as they apply to us. So what sense could it make for us to say, in our language, that the Martians are logically mistaken?

Appendix: Completeness for Bilateralist Validity

In 3.IV and again in 4.I I appealed to the notion of a correctness-valuation for a bilateralist language, $\models_c$, and claimed that there are bilateralist natural deduction systems, $\vdash_b$, such as the one I

presented in 3.IV, that are sound and complete with respect to $\models_{c}$. I prove that here. The strategy of the proof is to use a translation between signed languages and unsigned languages, together with the standard completeness theorem, to induce completeness in the extended sense. We'll need several definitions:

(t) Let t be a translation from the signed language to the unsigned language such that $t(+\phi) = \phi$ and $t(-\phi) = \neg\phi$

(Γ^t) If Γ is a set of signed sentences, let $\Gamma^t = \{t(\alpha) | \alpha \in \Gamma\}$

(*) Let * be a translation from the unsigned language to the signed language such that (i) if $\phi \neq$ "¬"$^\frown\psi$ for some sentence ψ, then $*(\phi) = +\phi$; (ii) if $\phi =$ "¬"$^\frown\psi$ for some sentence ψ, then $*(\phi) = -\psi(\Gamma^*)$ If Γ is a set of unsigned sentences, let $\Gamma^* = \{*(\alpha) | \alpha \in \Gamma\}$

One point of note that will be relevant below is that * is not exactly the inverse of t, since consider $+\neg\phi$: $t(+\neg\phi) = \neg\phi$, but $*(\neg\phi) = -\phi$, so $*(t(+\neg\phi)) \neq +\neg\phi$

Now, four lemmas will establish the completeness theorem.

Lemma 1. *if* $\Gamma \models_c \alpha$ *then* $\Gamma^t \models t(\alpha)$

Proof. By (VALIDITY) from 3.I, if $\Gamma \models_c \alpha$ then for all $v_c \in C$, if $v_c(\beta) = 1$ for all $\beta \in \Gamma$, then $v_c(\alpha) = 1$, and $v_c(\beta) = 1$ for all $\beta \in \Gamma$ if and only if $v(t(\beta)) = T$ for all $\beta \in \Gamma$, since if $\beta = +\phi$, the result follows since $v_c(+\phi) = 1$ if and only if $v(\phi) = T$ by ($C1$) from above. And if $\beta = -\phi$, by ($C2$) $v_c(-\phi) = 1$ if and only if $v(\phi) = F$, and $v(\phi) = F$ if and only if $v(\neg\phi) = T$. And by the same reasoning, if $v_c(\alpha) = 1$, then $v(t(\alpha)) = T$. So for any valuation v, if $v(\beta) = T$ for all $\beta \in \Gamma^t$, then $v_c(\beta) = 1$ for all $\beta \in \Gamma$, and so by $\Gamma \models_c \alpha$, $v_c(\alpha) = 1$ as well, and so $v(t(\alpha)) = T$, that is, $\Gamma^t \models t(\alpha)$.

Lemma 2. *if* $\Gamma^t \models t(\alpha)$ *then* $\Gamma^t \vdash t(\alpha)$

Proof. By the completeness theorem for standard first-order logic.

Lemma 3. *if* $\Gamma^{t} \vdash t(\alpha)$ *then* $\Gamma^{t*} \vdash_{b} *(t(\alpha))$

Proof. A case-by-case inspection shows that each standard inference rule transforms into a *b*-provable rule under *, and each structural rule of the standard system is also a structural rule of *b*, so our result follows. __

Lemma 4. *if* $\Gamma^{t*} \vdash_{b} *(t(\alpha))$ *then* $\Gamma \vdash_{b} \alpha$

Proof. The only case in which $*(t(\beta)) \neq \beta$ is when $\beta = +\neg\phi$, and as we saw above, $*(t(+\neg\phi)) = -\phi$, but we know from 2.IV's rules of ($\neg I$) and ($\neg E$), that $+\neg\phi \dashv_{b} \vdash_{b} -\phi$, from which our result follows. __

Theorem. *(Completeness) if* $\Gamma \models_{c} \alpha$ then $\Gamma \vdash_{b} \alpha$

Proof. Immediate from lemmas 1, 2, 3, and 4. __

---

1 See chapter 5 of Quine (1970*b*).

2 As I discussed briefly in 3.II and will detail at length in chapter 5, the meaning-constituting rules for a logical expression may well include rules that are only indirectly for the expression—schematic rules that don't contain the expression, non-schematically.

3 Standard semantics are presented in intermediate logic textbooks, for example, Boolos and Jeffrey (1989) or Enderton (2001).

4 Soundness is relatively straightforward; completeness was first proved in Gödel (1929).

5 My terminology of "correctness-valuations" follows Murzi and Hjortland (2009), who cite Humberstone (2000*a*) for their discussion.

6 Of course, the formal system introduced in chapter 3 didn't have the conditional as a primitive notion, but this was merely for the sake of simplicity (3.V gave bilateralist rules for the conditional).

7 See the remarks on the nature of mathematical necessity in Burgess (2013).

8 I call it a "meta-rule" because it is important to distinguish necessitation as stated from the analogous (and invalid) inference rule with premise ϕ and conclusion $\ulcorner \Box\phi \urcorner$. In 12.IV I will discuss the formal meta-rule of necessitation in a very different context.

9 In a bilateralist framework necessitation and related principles would deal with signed sentences, telling us that if $\vdash +\phi$ then $\vdash +\Box\phi$, for example.

[10] Some of the points about necessity made here aren't unrelated to philosophical points about necessity that have been made in discussions of two-dimensional modal logic—see Evans (1979) and Davies and Humberstone (1980).

[11] In this discussion I have been leaving aside the issue of whether or not the general metasemantic principles are themselves conventional. I think that it is plausible that they are constitutive principles of meaning in some sense, but this isn't exactly to say that they are true by convention. A full discussion of these matters would require a full theory of meaning. As Hellman (1986) notes, there is a difference between claiming (a) that logic is true by convention and claiming (b) that logical conventionalism is true by convention. Whatever the status of these principles, for the reasons given in the text, I think that (a) is secure. For further discussion of the status of conventionalism, see 14.I.

[12] The only somewhat plausible objection to this that I can imagine is that conventionalism requires radical conventionalism, such as was (seemingly) endorsed by Wittgenstein (1956). I discuss objections along these lines in 7.IV.

[13] I won't discuss *a posteriori* necessary and *a priori* contingent claims here, but note that nothing I've said so far commits me to taking a stand on these matters. And conventionalism, as developed here, certainly doesn't force identification or even extensional coincidence of the necessary, the *a priori*, and the analytic.

[14] Compare Lewis's (1984) remarks about how conforming to a constraint C, as opposed to merely making C-theory true, can ward off Putnam's (1980) model-theoretic argument against realism. Related points will also be discussed in the next section and, in much more detail, in chapter 11 and at greater length in 12.III.

[15] A metaphysician like Sider (2011) could potentially offer a theory like this.

[16] Tolerance is from Carnap (1934).

[17] A similar point about the role of completeness is noted by Hellman (1986).

[18] See Enderton (2001) and Shapiro (1991).

[19] The dialectical subtleties of this general type of situation are explored in Burgess (2005*b*) and Woods (2019).

[20] Gödel (1931).

[21] For relative interpretability and Robinson arithmetic see Tarski, Mostowski, and Robinson (1953).

[22] This distinction comes from Field (1994*a*).

[23] Introduced in Tarski (1933*a*), though it should be noted that Tarski wasn't a deflationist. Various broadly deflationary views of truth are endorsed in, for example, Ayer (1946), Horwich (1990), Grover, Camp, and Belnap (1975), Ramsey (1927), and Quine (1970*b*).

[24] Field (1994*a*), page 256.

25 See Field (1994*a*), page 258. Here I have presented the argument in schematic form; in his discussion Field considers several different ways of formalizing arguments like this so that they have sufficient generality.

26 More carefully, Field-style deflationists make these claims. Others—Horwich (1990), for example—do not, at least explicitly. For related context on Field's metasemantic views and their development see his (1986) and (2017).

27 In his discussion Field allows that there is a sense in which deflationists can allow that logical expressions express the truth functions they do "because of" their conceptual roles (1994*a*, pages 259-260). Because of this, I am unsure whether or not Field would object to the distinction I make below. McGee (2006) interprets deflationists not as answering metasemantic questions with disquotational arguments, but as holding that no story connecting meaning to use can be told.

28 See Jackson, Oppy, and Smith (1994) for example. Field is the most prominent roughly sententialist deflationist; Horwich (1990) is the most prominent propositional deflationist.

29 Although they don't put things in the same way, I think my point here is related to both Harman's (1982) remarks on Davidson-style (1967) approaches to meaning and Rumfitt's (2000) remarks on the truth axioms approach of McDowell (1997).

30 I think that the two views very neatly cohere, once the semantic/metasemantics distinction has been made. So much so that a combined view—inferentialist deflationism—is, I think, the best overall approach to matters of meaning. Field seems unsympathetic, but Horwich's (1998*a*) and (1998*b*) combine into a similar view, I think.

31 A related distinction is made by Unger (2017*a*). See 7.II for some related discussion.

32 For a general discussion of creeping minimalism about truth, facts, and other metaphysical notions, see Dreier (2004).

33 I actually have more than one copy of *Logical Syntax*, but my original copy is green!

34 For logical atomism see Russell (1919) and Wittgenstein (1922)—I have here used the terminology from the original Ogden/Ramsey English translation of the Tractatus in talking of "atomic facts". It is likely that both Russell and Wittgenstein wanted to restrict the atomic facts so that high-level "facts" about macroscopic objects could not be atomic.

35 Arguably, the atomist project came to ruin on this point; see Wittgenstein (1929) on the color exclusion problem. But perhaps some gerrymandering can salvage the project; see Moss (2012).

36 In fact, I suspect that there are objective connections of nomological necessity, but nothing in my argument hangs on this assumption. Interestingly, Putnam (1992) reports that even Carnap believed in objective (non-epistemic) nomological connections of this kind.

37 As I will discuss in 7.I, care must be taken to distinguish standard counterfactuals from counter*conventional* conditionals, in both directions; see also Einheuser (2006).

38 The discussion in the last three sections is relevant to the question of which theory of truth conventionalism is committed to. The answer is that there is a lot of wiggle room. Only explanatorily robust versions of the correspondence theory are ruled out, but arguably theories of that kind are not theories of "truth" at all, properly speaking—see Lewis (2001).

39 Field (2009) argues that there is a paradox of validity, related to the semantic paradoxes, and that this gives rise to problems for the view of validity as necessary truth preservation. I won't deal with the semantic paradoxes at any length here, but I think we should be wary of putting too much weight on liar-like puzzles when theorizing, no matter the topic. For some relevant discussion, see 12.IV.

40 See Harman (1986*b*).

41 See also the discussion of correct reasoning in Broome (2014*b*).

42 This approach treats justification as a mass term, which seems natural in this and similar contexts—see section 4 of my (forthcoming *a*), where it is applied to state the normativity involved in rule-following more generally.

43 The classic discussions are in Midgley (1959) and, less directly, Rawls (1955).

44 See Midgley (1959) and Searle (1969) for detailed developments of this distinction, as applied to linguistic rules.

45 Quoted from Boghossian (2014), page 17—Boghossian's earlier inferentialist writings are his (1996), (2000), (2001), and (2003).

46 See Zangwill (2015).

47 Peregrin (2018) also answers this argument, independently using the same chess-based analogy I use above and below. My only reservation about his discussion is that in footnote 2 Peregrin cites my (2015*b*) but seems to indicate that I reject unrestricted inferentialism, despite that paper actually being an extended defense of unrestricted inferentialism! In a personal communication, Peregrin has informed me that this mistake was due to a minor error in translating his paper into English. But the editors of *Analysis* didn't deem the mistake major enough to warrant a printed correction or retraction (Peregrin asked them to correct it).

48 Zangwill (2019) denies that the chess/language analogy is a good one, in the relevant respects, but his denial seems based on an explicit stipulation picture of the meaning-determining conventions in logic. I rejected the explicit stipulation picture in 2.II and will discuss it further in 7.III for the particular case of logic.

5

Alternative Conventions, Alternative Logics

When conventionalists say that logic is true by convention, do they thereby claim that logic is completely *arbitrary*? This has been a common charge from anti-conventionalists, and throughout history conventionalists themselves have straddled the fence on this issue. On the one hand the arbitrariness of conventions has been used to motivate an attractive brand of logical pluralism; but on the other hand the idea that something like logical validity is completely arbitrary is abhorrent—it suggests a postmodern nightmare where people can change logics as often, as easily, and with as little consequence as they change their socks.

In taking meanings to be cheap, unrestricted inferentialists say that *any* collection of rules can be used to implicitly define an expression. This seems to lead to a galloping pluralism that threatens to trample over our entire intellectual landscape.[1] The nature and extent of this pluralism must be discussed to further flesh out the philosophical theory developed in chapters 3 and 4. Saying something about this tangle of issues also serves to further defend unrestricted inferentialism against impressive putative counterexamples (like tonk).

I The Translation Mistake

Chapter 3 endorsed unrestricted logical inferentialism, as summed up in the culminating Charity to Use principle: the basic inference rules according to which a logical expression is used are automatically valid, and all logical validity flows from these basic rules. There I also briefly considered Prior's tonk connective as an apparent counterexample to unrestricted inferentialism and argued

that all of the force of this “counterexample” rests on the mistaken use of homophonic translation. This mistake is so important and so often made that I’ve given it a name:[2]

> **The Translation Mistake**: The mistake of being misled by superficial features into thinking that a homophonic translation is appropriate when it is not.

A homophonic translation of a language *L* into a language *K* translates each sentence ϕ in *L* into the syntactically identical sentence in *K*. All arguments that I’m aware of for the conclusion that tonk refutes unrestricted inferentialism commit the translation mistake. The case for this is made in detail in my paper “Talking with Tonkers”, but it is worth developing some of the same points here from a different direction. Leaving the case of tonk to one side, I think that *every* supposed counterexample to unrestricted inferentialism commits the translation mistake in one form or another.

Apparent counterexamples to unrestricted logical inferentialism conform to the following general schema: someone proposes connective $\oplus$ as implicitly defined by the rules *R*, notes that adding $\oplus$ to English and following *R* allows us to prove some obviously false English sentence, for example, “Rudolf Carnap is a fried egg”.[3] Since this sentence also occurs in English, it’s natural to translate it into English homophonically, which is what generates the putative counterexample. Yet the very fact that this sentence is provable and thus true in English+*R* while false in English already shows that the sentence *does not mean the same thing in the two languages.* This verdict will be upheld by any serious standard of translation, so the homophonic translation between English+*R* and English fails, and there is no counterexample to unrestricted inferentialism along these lines.

To think otherwise is to subscribe to a kind of word magic, where syntactic strings are super-glued to their meanings no matter how much their usage may change. Obviously this picture of language is silly and should be admitted to be so by everyone. But even sensible philosophers have been misled because they have focused on *adding* tonk and other rules to English or some other natural

language as a test case. Unfortunately, this procedure naturally pushes one into making the translation mistake. When we imagine putative English speakers accepting and following the tonk rules, we are misled by the standard pragmatic presumption of homophonic translation *between speakers in the same natural language*. Clearly though, this presumption can be defeated by radical usage changes, so it can't appropriately be assumed without argument in any of the cases that are of interest to us. The assumption of homophonic translation in these cases is an artifact of philosophically irrelevant features of the setup.

Of course, my terminology builds it in that the translation mistake is, in fact, a mistake. This doesn't mean that homophonic translation is *always* mistaken. In many cases homophonic translation is appropriate, but homophonic translation is often assumed appropriate simply because it is available, and that assumption is false. Superficial features never provide any deep evidence of the appropriateness of homophonic translation, at least in the interesting cases considered by philosophers. Misled by such features, the translation mistake occurs again and again, across all areas of philosophy, to the detriment of clarity and philosophical progress. But the mistake's impact is nowhere greater than in discussions of inferentialism and alternative logics.

At a general level I think I've already said more than enough to illustrate how putative counterexamples to unrestricted inferentialism invariably commit the translation mistake. But given the importance of this point, it's worth setting out an argument for it in detail, focusing for simplicity on the case of tonk. We can make the argument more explicit by appealing to the following uncontroversial principle:[4]

> **Save the Truth Values**: When translating language *L* into English, we should reject any translation that maps a true [false] sentence of *L* to a non-true [non-false] sentence of English, or vice versa.

To be clear, this principle is not saying that we should take all sentences speakers *take to be true*, as *really true.* Strong charity

principles of that kind arguably led Davidson astray in his discussions of conceptual pluralism.[5] My principle is merely saying that *actual truth values* should be preserved. This is something we can all agree upon, surely. Let Tonklish be a language just like English except that the tonk rules are also followed:

$$(tI)\quad \frac{\phi}{\phi \text{ tonk } \psi} \qquad (tE)\quad \frac{\phi \text{ tonk } \psi}{\psi}$$

For the moment let's ignore rejection rules and imagine we're adding these rules to a standard natural deduction-based full language modeling English; call it "*E*". Any sentence in the language is now provable, since from some theorem "*p*" of *E*, by (*tI*) we have "*p* tonk *q*" for an arbitrary sentence "*q*", and then by (*tE*) we establish "*q*". Now imagine that the arbitrary sentence is "Rudolf Carnap is a fried egg" and consider the following argument:

1. In Tonklish, the tonk rules are followed (assumption)
2. *So*: The tonk rules are valid in Tonklish (from 1 and unrestricted inferentialism)
3. *So*: "Rudolf Carnap is a fried egg" is true in Tonklish (from 2, the proof discussed above, and unrestricted inferentialism)
4. Rudolf Carnap is not a fried egg (trivial truth in English)
5. *So*: "Rudolf Carnap is a fried egg" is not true in English (from 4 and the logic of truth and negation)
6. *So*: The sentence "Rudolf Carnap is a fried egg" should not be translated homophonically between Tonklish and English (3,5, Save the Truth Values)

An analogous argument also blocks homophonic translation in every case where an alternative language can prove the (syntactic) negation of an English truth, or when we can prove, in English, the syntactic negation of one of their true sentences.

It is also possible to rule out homophonic translation when neither language's rules allow the negation of a sentence provable in the other language to be proved.[6] This is important for cases of real

concern, such as intuitionism and classical logic, and it is also relevant when considering extensions of classical logic, like some connexive logics, which don't contain as theorems any (syntactic) classical logical falsehoods.[7] One strategy for producing general arguments of this kind appeals to a global charity principle or—equivalently in my view—a global form of unrestricted inferentialism. The general idea is that the overall patterns of use in languages with different logics will always make homophonic translation insufficiently charitable: one or both sides will end up being interpreted under homophonic translation as making silly and inexplicable mistakes, contrary to charitable interpretation and the general rationality of those we are interpreting.

A related but somewhat stronger approach would require that logic itself is to be preserved under translation, just like truth values:

> **Save the Logic**: When translating language *L* into English, we should reject any translation that maps a logically provable sentence of *L* to a sentence of English that is not logically provable, or vice versa.

Where a sentence is logically provable just in case it is provable using only the logical rules of the language (and so is a logical truth, according to the first characterization in 3.I). So in an intuitionist language—Intuitionese—instances of the law of excluded middle will not be logically provable, which will allow us to rule out homophonic translation between Intuitionese and (classical) English via the following argument:

1. In English, "either Rudolf Carnap is a fried egg or he is not" is logically provable (assumption)
2. In Intuitionese, "either Rudolf Carnap is a fried egg or he is not" is not logically provable (assumption)
3. *So*: "Rudolf Carnap is a fried egg or he is not" should not be translated homophonically between Intuitionese and English (1, 2, Save the Logic)

And in a language governed by an appropriate connexive logic—Connexivese— there will be logically provable sentences corresponding to instances of Aristotle's thesis (⌜ ¬(*ϕ* →¬*ϕ*) ⌝) for instance, that are not logically provable in English (nor are their negations provable), so in these cases too homophonic translation is ruled out.

If Save the Logic is thought too strong, say because of the possibility of a language—*M*-Intuitionese—that counted excluded middle as a topic-specific *mathematical* rule, rather than a strictly *logical* rule, then we could still, as alluded to above, argue against homophonic translation for standard Intuitionese by appeal to broader charity or rationality-based metasemantic principles, such as:

> **Save the Theorems**: When translating language *L* into English, we should reject any translation that maps a provable (via basic rules) sentence of *L* to a non-provable sentence of English, or vice versa.

This principle abstracts from various distinctions we might want to make between different types of basic rules (*logical* versus *mathematical*, for example) and concerns itself only with which sentences in a given language are provable using the basic rules of that language, whether those rules be logical, mathematical, conceptual, or otherwise.

Still further refinements could be added by considering not only theorems, but provability relationships in general. In fact, this move is essential when considering languages without any theorems:

> **Save the Proofs**: When translating language *L* into English, we should reject any translation *t* that maps a sentence *ϕ*, provable using the rules of *L* from the sentences in Γ, to a sentence *ψ* of English not provable from the *t*-images of the sentences in Γ, or vice versa.

This principle is more involved than those we have previously been considering in that it is more directly holistic. But translation is widely recognized to be holistic, at least in some weak sense (the issue of

holism will be discussed at length in the next section, and general issues of equivalence between logics will be discussed in section IV). All of these principles are plausible when seen against the unrestricted inferentialist background developed in the first three chapters of this book.

I've been arguing that a proper, inferentialist understanding of meaning, particularly for logical expressions, rules out homophonic translation when the imagined alternative language has logical rules that significantly differ from our own. This dovetails with assigning a prominent role to charity and rationality in interpretation. However some may worry that my procedure will overgenerate and rule out possible disagreements about *anything*, so it is important to understand why it does not.

To dramatize the worry, imagine that an alternative language community accepted our logic but disagreed with us about scientific claims, say claims concerning electrons or whatnot. Disagreements like this are common in contemporary scientific debates and have occurred throughout the history of science. Some historically influential philosophers of science—most prominently Kuhn, Feyerabend, and Lakatos—seem to have swallowed the poison pill and argued that scientific disagreement invariably results in meaning change.[8] This is usually thought to be unduly extreme, but some may worry that my arguments generalize to the scientific case and force me into a kind of Kuhnian relativism about science. Elsewhere I have called this concern "the charity trap".[9]

The charity trap can be avoided because the metasemantics of both scientific terms like "electron" and ordinary terms like "chair" differ significantly from the metasemantics of logical terms like "and" and "not". I have already stressed this at several points (in 3.I, for example). The key difference is that the metasemantics of scientific and ordinary terms typically involves some type of causal interaction with the world. This is not an ad hoc point, and was in fact the key point of response made to Kuhn and other radical philosophers of science by Kripke, Putnam, Burge, and other early *semantic externalists*.[10] By altering our acceptance of claims about "electrons" and "chairs" in a systematic manner in response to causal feedback from the world, we all manage to talk about the very same external

world using the very same language, even when we disagree drastically. In logic, by contrast, there is no causal feedback from the world to counteract the force of charity. This is why charity is nearly unlimited in logic but severely limited in both science and everyday life.

This point is fully compatible with a global form of metasemantic inferentialism if we allow that the inference rules that determine the meanings of terms like “electron” and “chair” are linked causally to states of the world, while the meaning-determining rules for logical expressions are not.[11] With this understood, whether or not global inferentialism is accepted, while my arguments may generalize beyond logic to mathematical and conceptual matters, there is no real threat of them generalizing to *all* areas of discourse, thus the charity trap is easily avoided. I reject homophonic translation as applied to alien logics but, unlike Kuhn (apparently), I am open to homophonic translation in many actual and imagined disagreements in science and ordinary language.

Some further criticisms of my approach will be discussed in chapter 7, where specific objections to conventionalism are considered and rebutted. What is most important at present is to recognize that the translation mistake, in particular, and the mistaken assumption of homophonic translation, in general, are what drive arguments against unrestricted inferentialism and logical conventionalism. The unrestricted inferentialist can argue, in all cases that might be of interest, that homophonic translation can be conclusively ruled out. The “big picture” reason is the same in every case: languages like Tonklish, Intuitionese, and Connexivese cannot be translated homophonically into English without making speakers shockingly irrational and unaccountably foolish. And the situation is entirely symmetrical—speakers of alternative logical languages can rule out homophonic translation from our classical language into their own.

II Tonk, Translation, and Holism

While Tonklish cannot be translated *homophonically* into English, this doesn't imply that it cannot be *non-homophonically* translated. To translate Tonklish sentences into English we can simply translate every single sentence of Tonklish into a logical truth of English.[12] This translation will be completely satisfying on grounds of charity and rationality. In Tonklish, every single sentence is provable from the empty set of assumptions and so is a necessary truth provable (in Tonklish) in the manner in which we prove a logical truth of English. If we so desire, we can even set up the translation to more closely mirror the structure of proofs in Tonklish by adding various complications.[13]

Obviously, even if complicated in various ways, no such translation of Tonklish into English can be uniquely correct. But that isn't a serious problem. The point is not that this translation is unique or somehow sacrosanct; rather, it is that there are ways to translate Tonklish into English that make sense of many aspects of the Tonkers' verbal behavior.[14] The availability of such translations puts further, complementary pressure on homophonic translation in the tonk case, even apart from the direct argument against homophonic translation from the previous section.

I won't dwell on this translation, partially because the simple models of Tonklish and English that we've so far been working with have ignored the rejection rules that are built into English. Imagine that instead of this simplified model we use a bilateralist system as our full language for English (*Eb*), and note that by the assumption that English obeys classical logic, $\vdash_{Eb}$ - ($\phi \wedge \neg\phi$). Now let's add bilateralist versions of the tonk rules to *Eb*:

$$(tI+) \quad \frac{+\phi}{+\phi \text{ tonk } \psi} \qquad (tE+) \quad \frac{+\phi \text{ tonk } \psi}{+\psi}$$

Say that *Tk* = *Eb* plus these rules. We have both (*i*) $\vdash_{Tk}$ - ($\phi \wedge \neg\phi$), since *Tk* includes all of the rules of *Eb*, and (*ii*) $\vdash_{Tk}$ + ($\phi \wedge \neg\phi$) by proving some logical truth and then using (*tI*+) followed by (*tE*+). This is not a contradiction in the standard sense, it is something even

more problematic: the rules of *Tk*, in effect, tell us that we can both accept and reject any given contradiction. But as I have stressed at various points in previous chapters when discussing rejection and bilateralism (see the discussions in chapters 2 and 3), this is simply not possible. This much is agreed upon even by dialetheists. One cannot both accept and reject any given sentence, any more than one could both catch and miss a bus. The functional roles of accepting *p* and rejecting *p* cannot be jointly instantiated (2.V). This means that the rules of *Tk*, so understood, cannot be followed.

This effectively requires that we deny that *Tk* is a possible, meaningful language. But note well: unrestricted inferentialists reject *Tk* not because they are rejecting unrestricted inferentialism's neat and tidy naturalistic metasemantic picture. Instead they are rejecting it because it isn't possible for any actual beings to follow the *Tk* rules. This impossibility isn't owed to some external constraint in metasemantics; rather, it is rooted in the capacities of physical beings in a physical world. Since you can't follow the tonk rules while also following rules that enjoin you to reject some sentence in your language, no sense can be made of adding the tonk rules to such a language. This means that we can reject this language's possibility without dropping MAC and thus unrestricted inferentialism.

This isn't a revision of what I said in chapters 3 and 4. When I introduced the MAC principle in 3.II I took pains to stress that it only concerned rules that *could* be used for an expression. Since the bilateralist version of Tonklish's supposedly meaning-determining inference rules cannot be followed, they cannot be used for an expression, and thus no actual restriction on MAC is required to handle cases like this. This is not an ad hoc move. In fact, I suspect that this is really the point that bothers many anti-tonk theorists. And if this is right, then those theorists should re-open the door to unrestricted inferentialism, since adding rejection rules to their theorizing allows them to see that what was really troubling them about tonk *isn't anything that requires rejecting unrestricted inferentialism.* Carnap's castle survives Prior's panzers.

With this said, I think it's important to again stress that my approach to *Tk* doesn't commit me to an implausible vision of rule-following in which rules are always followed perfectly, without any

slip-ups or mistakes. On the contrary, as discussed at length in 2.VII–2.IX, dispositionalism about inference rule-following allows for the possibility of various errors and mistakes. This point is perfectly compatible with the fact that realizing a disposition to Φ is incompatible with realizing a disposition to Ψ in some cases, even necessarily so. It is impossible to follow the *Tk* rules because the dispositions that this would involve cannot be had by one agent or community at any given point in time. While inference rule-following doesn't invariably involve inferring according to the rule, it does involve a stable, *M*-general disposition to infer according to the rule in normal conditions when given the option, provided that certain other conditions are in place (2.VII and 2.VIII). If given the option, say by being asked about a particular contradiction, you cannot both accept it and reject it, so you cannot have the dispositions that constitute following the tonk rules while at the same time having the dispositions that constitute following the classical rules. All of this is compatible with the possibility of making mistakes and slip-ups.

That the tonk rules can't be followed while the classical English rules are also being followed is no more mysterious than that one can't both, at the same time, play the game chess and the game *R-chess*, where, in R-chess, rooks move like bishops but everything else is as it is in chess. While it is possible to slip and make a corrigible mistake, it is not possible to truly be playing both games at the same time. There is simply nothing that could count as playing both games at the same time. This is true even if you get checkmated before any rooks get moved. Your dispositions determine which rules you are following. You can't have appropriate dispositions to move your rooks that count as following both sets of rules. There are no such dispositions. Similarly for the enforcement dispositions. The tonk case is no more troubling or mysterious than this case, so there is no real threat to unrestricted inferentialism when these matters are seen correctly.

Another worry about my treatment of Tonklish concerns how to handle versions of Tonklish where speakers follow the tonk rules only in obscure religious ceremonies or other rare situations. For such languages, the vast majority of their linguistic practice is exactly like English, and so homophonic translation seems appropriate. But

this isn't a problem for my view. In fact, I think the case here is misdescribed: the imagined speakers don't *really* follow the tonk rules in their full and unrestricted form. What we've described isn't really a tonk language, in the relevant sense, but rather a slight variation of English for which the homophonic translation is appropriate when it's available. There are variations on this kind of case, but they can all be handled similarly.[15]

A related issue concerns the possibility of a "hidden" inconsistency in English (say involving the Liar paradox or Curry's paradox). Using classical logic, such an inconsistency would lead to triviality, which, by the metasemantic assumptions I've made, would lead to the result that every sentence *in English* is true. This would be an absolute disaster, but I don't think a hidden inconsistency in our language should be handled in this manner. In framing Tonklish, I've been imagining that the Tonkers exploit the tonk rules as runabout inference tickets. By contrast, English speakers are not disposed to exploit the Liar or Curry's paradox as runabout inference tickets. Nor are they disposed to do so even when given the opportunity. If I were to walk into a mathematics conference and announce that I had proved "*P* = *NP*" using the Liar, I would very quickly be thrown out (perhaps literally). No Fields medal would be forthcoming, even if they could not point out the flaw in my reasoning.[16]

Using a bilateralist formulation, we can immediately rule out the possibility of English's rules being inconsistent *in this sense*, in the same way that we ruled out *Tk*—not only do we not have the dispositions to exploit a hidden contradiction as a runabout inference ticket, we *could not* have such dispositions while still following the standard rules that we follow. Of course, there are further issues about how inferentialists should handle inconsistent collections of rules, and there are many options. I am not going to discuss any of them in detail here (though I will discuss an approach to the set-theoretic paradoxes in 11.III). It is enough to note here that the problem of a "hidden inconsistency" in our language can be faced by unrestricted inferentialists without making our language into a trivial language. This is possible because English speakers absolutely *do not* exploit the liar and other potential inconsistencies in English as runabout inference tickets.

So far I've been focused on Tonklish since the case is so extreme and is a putative problem for unrestricted inferentialism, but languages using other (non-trivial) systems of logic can also be translated into our own (classical) language in a non-homophonic fashion. This is perhaps best illustrated by translations between classical and intuitionistic logic such as that given by Glivenko.[17] I won't discuss the merits or demerits of any particular proposals here, but it is worth noting again that they can be used to put complementary pressure on the idea of homophonically translating alien languages with alien logics into our own. And although I've been assuming that English is classical for the purpose of simplicity, this general point goes through whatever the actual logic of English may happen to be (for more on non-homophonic translations between logics see section IV below).

These matters of meaning and translation are intimately connected to the issue of semantic holism. Some may think that my claims about non-homophonic translation between English and Tonklish, Intuitionese, Connexivese, and other possible alien languages involve the tacit assumption of a problematic form of semantic holism that I explicitly rejected in 3.II. To illustrate the concern: if the meaning of "Rudolf Carnap is a fried egg" differs between English and Tonklish, then somehow adding the tonk rules to English must have altered the meanings of expressions like "Rudolf Carnap", "is", and "a fried egg", and this seems to involve radical semantic holism. Yet I deliberately eschewed holism, arguing that only the basic rules for an expression need to be taken as meaning constituting for an expression, so there is some *apparent* tension here.

The way out of this tension was already presaged in chapter 3 when I distinguished between wide and narrow forms of holism (3.II). There I deliberately said that a schematic rule *R* was indirectly "for" expression $\oplus$, if $\oplus$ occurs in only *some* instances of *R*. For non-schematic rules, if an expression occurs in some instance of the rule then it will occur in all instances of the rule, but not so for schematic rules. This gave rise to the distinction between a rule being directly versus indirectly for an expression. And the expressions "Rudolf Carnap", "is", and "a fried egg" all occur in some but not all instances

of the tonk rules, and so the tonk rules are, in our extended sense, rules that are indirectly "for" these expressions as well. This might seem a bit odd at first glance, but upon reflection it is extremely natural for anyone who accepts a close tie between meaning and use. The schematic tonk rules have instances that include all of these expressions, so following the tonk rules involves drastically changing the way that "Rudolf Carnap", "is", and "a fried egg" are used. In fact, the change in use between English and Tonklish is so drastic that it would be natural to say that these expressions have changed their meanings. Similar though less drastic changes can occur in non-Tonklish cases as well.

In 3.II I recognized the holist-like ramifications of this procedure by saying it might be dubbed "narrow holism". What is important is that this is *not* the broad holist position that *every* valid rule in the language is meaning-constituting. For example, in my view the rule of (&&), from chapter 3, isn't meaning constituting in either English or Tonklish. The holism I am accepting here is only the narrow or quasi-holist position that allows that any basic rules "for" an expression, in the extended, indirect sense, are meaning-constituting for the expression. In many cases, the effect of working with the more liberal, indirect notion of a rule being "for" an expression rather than the more demanding direct notion—requiring that expressions occur in every instance of the rules for them—will make no difference at all. But in the case of extremely powerful rules, like the tonk rules, it makes all the difference in the world.[18] In these cases what really alters meaning are the instances of the tonk rules that involve the relevant expressions, so nothing occult is happening.

This is why a principle like Totality doesn't actually force our meaning-constituting rules to be separable. Recall from 3.IV that a full language is separable if every proof of sentence ϕ from set Γ uses only non-structural rules directly for connectives that occur in either ϕ or in Γ. Logical inferences are schematic. Totality required that the valid logical inferences are determined by the meaning-determining rules for the expression(s) involved in the inference. But in Tonklish, the tonk rules are meaning-constituting rules for "and" as well as every other expression. This is because they are among the basic rules of the language that involve "and". The meaning-

constituting rules for logical expressions come in a package, so no direct separability constraint is imposed by Totality, and no indirect separability constraint is violated.[19]

If we are really and truly taking seriously the close tie between linguistic meaning and language use, then we must admit that the expanded notion is the proper one to use and accept meaning differences between Tonklish and English. This isn't a trick to avoid holism. The basic rules of a language are the meaning-constituting rules; derivative rules are not meaning constituting. This is as it should be. But because our basic rules can interact in odd ways, sometimes we need to recognize that the use of one expression is drastically altered by another basic rule that doesn't explicitly contain it. In general I prefer to say that the meaning-constituting rules for a logical expression include all of the instances of basic rules in which the expression occurs. This includes instances of schematic rules for other logical expressions that don't explicitly contain it. The basic rules determine the language, and the occurrences of an expression in basic rules—all of the basic rules—determine its meaning. Usually this interaction between rules can be ignored, but not in cases where the usage of an expression is changed markedly by its occurrence in schematic rules that don't directly contain it. This is what happens with the tonk rules, and it may happen in other cases as well. For all of these reasons, this isn't an ad hoc patch; it is simply a natural consequence of unrestricted inferentialism's roots in use-based metasemantics.[20]

Properly understood, the unrestricted inferentialist's treatment of tonk and other cases is radical, but not as radical as would have been thought before various subtleties are noted and various confusions cleared up. Some extreme imagined "languages" can be ruled impossible without rejecting unrestricted inferentialism, and others are made more palatable by the existence of non-homophonic translations from them into English. The ultimate upshot of all of this is that unrestricted logical inferentialism and conventionalism leads to a type of logical pluralism, where alternative languages with alternative linguistic conventions mean different things by the various logical expressions and their cognates. More must now be said about the nature and extent of this type of "logical" pluralism.

III Pluralism and Semantic Counterparts

According to logical conventionalism, logical facts in our language are fully explained by the conventions of our language in the form of linguistic rules. But there is nothing magical about the rules we have adopted, and if we had adopted different rules, then inferentialist principles would have applied to them as well. This type of simple linguistic pluralism straightforwardly leads to pluralism about the various logical notions themselves: negation, disjunction, conjunction, and so on.

The interesting and relevant kind of pluralism here is non-trivial. Obviously linguistic signs are trivially conventional—there is nothing special about the particular sounds and symbols we have decided to use, vis-à-vis their meanings. The tie between the word "cat" and actual felines is entirely arbitrary. There could be a language exactly like English except that the words for "cat" and "dog" were swapped. This language, of course, would not amount to an interesting version of cat pluralism or dog pluralism, and we wouldn't describe the imagined language users as believing that cats were dogs or vice versa. Trivial sign pluralism of this kind is not a theoretically interesting doctrine and it surely doesn't entail any substantive form of pluralism. An interesting version of logical pluralism must go beyond this by saying that different communities could have different concepts of *negation*, and this will involve saying something about what, precisely, makes some sign in a given language a negation sign.

Before carefully characterizing the interesting type of logical pluralism, which is entailed by logical conventionalism, it's worth noting that the conventionalist's logical pluralism is distinct from a well-known contemporary version of logical pluralism endorsed by J.C. Beall and Greg Restall.[21] For Beall and Restall, logics are characterized by consequence relations, where a consequence relation holds between some set of sentences Γ and sentence ϕ if and only if there is no *case* in which every member of Γ holds but ϕ does not. This much is fairly uncontroversial, but they add to this the point that natural language allows distinct ways of precisifying "case" and these different precisifications lead to different logics. One

understanding of “case” leads to classical logic, another leads to intuitionistic logic, and so on. I won’t address the merits of Beall and Restall’s brand of logical pluralism here; I just want to note that it is distinct from the kind entailed by conventionalism.

In order to correctly characterize this notion of “pluralism”, we need some general way of saying when a symbol in a given language is a *negation symbol* or when a symbol in a given language is a *conditional symbol*, for example. The natural way to do this—for the inferentialist or anyone else—is to appeal to the inferential role of an expression in its home language. Even those who don’t accept an inferentialist metasemantics should find this approach quite natural and intuitive.[22] Just as a little wooden carving counts as a chess queen just in case it is used in a certain way, a sound or a funny mark counts as a negation symbol just in case it is used in a certain way—namely, a way that is *relevantly similar* to the way that our negation symbol is used.

So consider once again our bilateralist full language model of English, *Eb*, and assume that the symbol “¬” is the negation symbol in *Eb*. We want to focus on the structurally characterized inferential role of “¬” in *Eb*, where this is meant to be the abstract structure of “¬”’s inferential role in *Eb*. By talking about the “abstract structure” here, we signal that what’s important is not that “¬” interacts in a certain way with the symbol “∧”, but rather that “¬” interacts in a certain way with *the conjunction symbol*, whatever it might be, say by vindicating the DeMorgan laws and whatnot. We need to focus only on abstract structure here because we don’t want irrelevant features of notation or grammar to be counted as part of “¬”’s semantic role.

We need to say a bit more to fully characterize this notion. Formally, full languages are constituted by three things: (1) the symbols in the language; (2) the language’s grammar, which determines which sequences of symbols in the language are well-formed formulas; and (3) the collection of inference rules for the full language, which we can imagine as consisting of some number of basic rules along with background structural rules.

Two full languages are notational variants if they are identical save for the symbols involved (so that symbol-for-symbol swapping can change one language into the other without tinkering with the

grammar or inference rules at all). And they are grammatical variants if they are identical save for the grammars involved in the sense that, if $\alpha_1 \frown \alpha_2 ... \frown \alpha_m$ is any well-formed formula of one language, there is a systematic way to recover a unique well-formed formula of the other language, with the same symbols, and vice versa. With these two notions in hand, we can say that two languages are stylistic variants of each other if each has a notational variant that is a grammatical variant of the other. Stylistic variants differ only in the symbols and the grammatical presentation involved, so if we start in a language that uses the symbol "*j*" for Jared and the predicate "*P*" for "is a philosopher", writing "Jared is a philosopher" as "*Pj*", a stylistic variant that wrote "*k*" for Jared and "*F*" for "is a philosopher" and wrote names before predicates in simple predications would write "Jared is a philosopher" as "*kF*".

The inferential role of an expression α in a given language is characterized by the basic rules of the language in which α can occur (even indirectly). Overall, these determine the facts about provability, expressed by "$\Gamma \vdash \phi$", where either ϕ or some member of Γ contains α. The structural inferential role of α can then be given by α's inferential role *modulo* stylistic variation. In other words, if α is in a full language *L*, and a full language *K* is a stylistic variant of *L*, then there will be an expression, β, in *K*, corresponding to α, and α and β will have the same structural inferential role. A structural inferential role is the inferential role stripped of the particularities of notation and grammar that happen to have been adopted. In some interesting cases, structural inferential roles are not identical, but are relevantly similar across languages. The notion of a semantic counterpart is based on comparing structural inferential roles that are non-identical, in terms of their relevant similarity to each other.

To illustrate these ideas, let's imagine a hypothetical community, say a community of Martians. Imagine that these Martians follow all and only the rules of a paraconsistent logic like Graham Priest's *LP*; *LP* stands for the "logic of paradox" and is meant to be a logic acceptable to dialetheists.[23] According to *LP*'s semantics, both a sentence and its negation can be true. By the MVC and MAC, conventionalists will think that all of the theorems of *LP* are true in the Martian language. Let's say that "~" is the putative negation

symbol in the full language *M* representing the Martian language. If some sentence of the form ⌜ $\phi \wedge \sim \phi$ ⌝ is provable in this full language, then it will be—according to the inferentialist—true in this language, so both ϕ and ⌜ $\sim \phi$ ⌝ will be true in *M*; and if we use a standard account of falsity, where ϕ is false just in case not-ϕ is true, then according to unrestricted inferentialists, ϕ will be both true and false in *M* provided that *"~"* is a negation symbol.

Whether or not "~" is a negation symbol in *M* depends upon whether the abstract structure of the inferential role of "~" in *M* is relevantly similar to the abstract structure of the inferential role of "¬" in *Eb*. In cases where only the symbols used differ or only the grammatical conventions of the languages differ (or both), the inferential roles of the expressions in each language will be not just similar but structurally identical. Such cases aren't illustrations of negation pluralism, but only of the trivial conventionality of signs and grammar. In the "cat"/"dog"-case considered above, the imagined languages are merely trivial notational variants of each other. But the case of Martian negation is different. *Eb* and *M* aren't mere notational variants of each other, yet it is still plausible that both "¬" and "~" are negation symbols. The role that "~" plays in *M* is similar but not identical to the role that "¬" plays in *Eb*.

The notion of a structural inferential role is quite general; it can be applied to any syntactic item used in a rule-governed manner (sentences, terms, predicates, quantifiers, connectives, operators, *et cetera*). This basic framework allows us, in a perfectly general way, to compare the use and thus, the meanings, of expressions in one language with the meanings of expressions in another language. A structural inferential role could also be called a "semantic role", in recognition of this fact. If the semantic role of an expression α in a language *L* is sufficiently similar (in the relevant context) to that of an expression β in a language *K* then we say that α and β are semantic counterparts. I take the term "counterpart" from its use in counterpart theory in modal logic.[24] Counterpart theory allows for a formal treatment of modal discourse without modal operators and without transworld identity claims (these claims are replaced by claims about counterparts). It is claimed by proponents that the counterpart treatment allows for greater generality than treatments of modality

that allow for the same item to be located in different worlds. This is because the counterpart relation, unlike identity, is not an equivalence relation. Similarly, the semantic counterpart relation allows us greater flexibility and generality when talking of meaning and concepts than does a simplistic picture allowing for only semantic identity and non-identity with nothing in between.

It need not be the case that each language has a counterpart for every notion expressible in every other language; and it need not be the case that each language has only one counterpart for a given notion in another language. To be semantic counterparts is to enjoy structurally similar semantic roles, and like all similarity relations, the semantic counterpart relation is subject to the difficulties of how to weigh and tally the various similarities and dissimilarities in particular respects. The right way to tally will depend on what is relevant and salient in a given context. So what counts as a semantic counterpart for some purposes and in some situations may not so count for other purposes and in other situations. This was to be expected and isn't a problem. For some purposes, only fully classical negation will count as negation-like, but for others, intuitionistic negation or paraconsistent negation might also count as negation-like. Some imagined languages might have two or more counterparts to our negation symbol, while some may not have any.

Largely for this reason, it wouldn't be worthwhile to try to come up with a perfectly general context-insensitive account of the similarity of the structural inferential roles of two given expressions in two respective languages. There will generally be an infinite number of valid inferences involving the expressions in each language, so we can't simply count the number of inferences that "match", in each language, perhaps assigning them a weight.[25] However, since the basic rules for each language will be finite in number if the language is learnable by beings like us, we can compare these numerically, as long as we keep in mind that this comparison must be done holistically.[26] In particular, we can compare how many of the basic rules for "¬" in English come out valid for "~" in Martianese, even if they are not basic in Martianese, taking care to account for tonk-style issues of holism by comparing all basic rules "for" the expression, including those in which the expression only occurs in the schematic

portion. Problems crop up even here since some languages won't have counterparts for the primitive notions in some other language. But despite these problems and the indeterminacy they give rise to, the notion of a semantic counterpart is, for many cases that concern us, clear enough in practice. It seems to match an intuitive notion that we employ quite effortlessly.

With the notion of a semantic counterpart on the table, we can use it to define a generalization of homophonic translation that I call counterpart translation. This was already used tacitly in formulating and attempting to compare structural inferential roles across languages (we are firmly sitting in Neurath's boat). The counterpart translation of a language *L* into a language *K* will translate each sentence (predicate/name/...) ϕ in *L* into its semantic counterpart in *K*. In some cases the counterpart translation will simply be the homophonic translation, but such cases will often be relatively artificial. Using semantic counterparts and counterpart translations we can now give a plausible definition of *X*-pluralism: two full languages *L* and *K* provide an example of *X*-pluralism just in case:

1. There is an expression α in *L* and a semantic counterpart β of α in *K*
2. The counterpart translation between *L* and *K* fails with respect to α and β

For a particular translation to "fail" is for it to be ruled out on metasemantic grounds. For example, as has been discussed above, especially in section I, many metasemantic theories impose some kind of charity or rationality constraint on translation, so that any suggested translation that violates such a constraint will be ruled out on metasemantic grounds.[27] We can, of course, think that many translations are ruled out on metasemantic grounds without thinking that translation is always perfectly determinate.[28]

Clearly, we can see that "¬" in *Eb* and "~" in *M* provide an instance of negation-pluralism, in most contexts. This is because in most contexts that concern us, these two symbols are semantic counterparts, but the counterpart translation between them fails on grounds of charity, since "~" admits true contradictions, but "¬" does

not. So a version of the argument from section I against homophonic translation also rules out the counterpart translation in this case. They are both negation symbols in their respective languages—or negation-like symbols—but they can't properly be translated into each other and thus don't mean the same thing.

With this understanding of pluralism it's easy to see that mere notational variants (like switching "∧" and "∨" in a completely systematic manner) will fail to be instances of pluralism, since they will fail condition (2). And mere sameness of string for a different notion won't necessarily be an instance of pluralism since it need not satisfy condition (1). But, just as clearly, there are real examples of conceptual pluralism, as the *Eb* and *M* case illustrates. What is key is that the notion of a semantic counterpart, and the definition of *pluralism*, allow us to see how the conventionalist's claims of logical pluralism go beyond the trivial pluralism entailed by the conventionality of linguistic signs. Logical pluralism is a substantive thesis that the unrestricted logical inferentialist/conventionalist is committed to: different possible languages can have different logics, in that these languages will have distinct logical notions that are still recognizably logical notions of the same kind. This is a substantive form of logical pluralism that goes beyond trivial semantic conventionality.

At long last, this pluralism justifies my claim that focusing on classical logic, as I did in chapter 3, did not embody a deeply partisan assumption. If we had started with intuitionistic rules, as similar story could have been told. Likewise for relevant or paraconsistent rules or the like. Later in this chapter I'll discuss in more detail the different philosophical debates about logic that conventionalism leaves open. Here the point is only that, given the kind of logical pluralism endorsed by the unrestricted inferentialist, those who don't think that the logic of English is classical can still accept my overall view, despite the assumption of classical logic in chapters 3 and 4.

Sometimes claims of pluralism go beyond even this and also involve the claim that all of the alternatives are objectively equivalent in some sense. Let me introduce the term metaphysically equivalent as a placeholder for this objective equivalence notion. With this

done, we can say that *L* and *K* are an instance of egalitarian *X*-pluralism if, in addition to (1) and (2), we also have:

(3) The relevant fragments of *L* and *K* are metaphysically equivalent

Whether or not conventionalism leads to egalitarian pluralism will depend upon how we flesh out the notion of metaphysical equivalence. With a general notion of metaphysical equivalence, it also becomes possible to compare many inequivalent languages, so that we may say that *L* is metaphysically superior to *K*. In the next section, I'll discuss the issue of metaphysical equivalence between distinct logics, but I won't spend much time trying to argue that different candidate logics are or are not equivalent to each other. For the moment what is most important is that I am skeptical of attempts to understand "equivalence" in a way that relies upon any of the recently fashionable primitive hyperintensional notions of heavyweight metaphysics, such as grounding, structure, or metaphysical definition.

Historical conventionalists have almost universally rejected heavyweight metaphysical notions, but this is not strictly forced by conventionalism. It is something that must be argued for on independent grounds, though I won't attempt to do so here.[29] Here and throughout this book, I make the standard assumption of scientific naturalism (1.IV). These metaphysical notions spit in the face of naturalism, and so should be rejected by scientifically inclined philosophers. In any case, leaving aside egalitarian and broadly metaphysical claims, with a semantic counterpart-based understanding of pluralism, logical conventionalism leads to logical pluralism, both broadly, in terms of different overall logics, and narrowly, in terms of different logical notions like negation, conjunction, and universal quantification. So not only does conventionalism lead to the possibility of a plurality of distinct logics, all correct in their home languages, it also leads, with the resources developed in this section, to pluralism about the logical notions that occur within these languages.

Since the notion of a semantic counterpart involves a context-sensitive, interest-relative, and (at least) somewhat indeterminate similarity relation, it might be thought that my position is that it is

arbitrary whether or not a symbol counts as a negation symbol, but that is decidedly not the case. The entire point of the semantic counterpart treatment of pluralism is to show *how* and *why* it is *not arbitrary* whether or not something counts as a negation symbol or a negation-like concept. Some notions are determinately *not* negation symbols: in no contexts, for no interest profile, could they possibly count as relevantly similar enough to negation to count as negation-like. Our concept of chair is not a negation concept by anyone's lights.

Some principles may be so essential to negation that anything that is to count as a negation-like symbol must obey said principle. Many philosophers seem to think that this is true of the law of noncontradiction, for negation.[30] Others deny this. But it is important not to confuse this issue with claims of translation. I think that "~" in *M* counts as a negation-like symbol, but I don't think it should be translated into our negation symbol. In effect, this gives rise to two distinct notions of something like analyticity, corresponding to deep and shallow conceptual requirements on a given concept. The deep requirements hold for *all X*-like concepts, the shallow requirements only for a particular concept of *X*.

Stepping back: the crucial point here is that to admit that what counts as a negation symbol is context-sensitive and not perfectly determinate is not to admit that it is arbitrary. To adapt a bit from Poincaré to the present context—are the rules of logic nothing but arbitrary conventions? Conventions, yes; arbitrary, no![31]

IV Equivalence between Logics

I reject metaphysically heavyweight approaches to (metaphysical) equivalence because I reject the metaphysical primitives used to generate such approaches. The natural strategy for a conventionalist is to think of fine-grained equivalence, between logics at least, in terms of some type of merely syntactic transformation between the languages containing the respective logical rules, along the lines used in characterizing semantic counterparts and, thus, pluralism.

An influential syntactic approach to equivalence was proposed by Clark Glymour for comparing theories in the philosophy of science. Glymour's approach appeals to *definitional equivalence*—two theories are equivalent if the definitional extensions you get by adding definitions of the notions from the other theory are logically equivalent in a standard sense.[32] This approach isn't uncontroversial even in science, and many have felt that it doesn't generalize to allow for the comparison of different logics. Mere definitional equivalence can't be directly applied to give an interesting notion of equivalence between different logics since it appeals to a notion of logical equivalence in its very characterization.[33]

In mathematics, perhaps the most widely used notion of equivalence is mutual relative interpretability, a notion first introduced by Tarski to study undecidability.[34] Interpretability involves mappings between the languages of theories that preserve proof relations and logical form. More specifically, theory T_1 is interpretable in theory T_2 when there is a mapping between the languages of the theories, $t : L_1 \mapsto L_2$, meeting the following two conditions:

1 The mapping preserves derivability, so that when $T_1 \vdash \phi$, then $T_2 \vdash t(\phi)$

2 The mapping preserves logical form, so that $t(\neg\phi) = \neg t(\phi)$, $t(\phi \wedge \psi) = t(\phi) \wedge t(\psi)$, and so on for the other logical constants

It may also be required that the mapping t be recursive. Mutual relative interpretability is then interpretability in both directions.

As defined, this notion doesn't apply directly to comparisons between different logics, but it can do so with a few modifications. Condition (1) needs to be made relative to the syntactic consequence relation of the logic in question, either $\vdash_{L_1}$ or $\vdash_{L_2}$, where this is defined in terms of provability from the basic rules of the logic together with the logic's structural rules (3.IV). We then need t to preserve L-derivability over any pair of a set of sentences and a sentence, both drawn from the target language. Condition (2) also needs to be altered, since we may want to compare logics that

involve distinct logical constants, but the best way to weaken (2) is unclear.

Recently John Wigglesworth, following Solomon Feferman, suggested that the best syntactic approach is to require only that *t* commutes with negation, thus dropping all of the other conditions on form.[35] He then proposes, on behalf of syntactic approaches to equivalence (which he does not accept), that two logics are equivalent when, for all sets of sentences Δ, the closure of Δ under a logic L_1 is mutually relatively interpretable, in this weakened sense, with the closure of Δ under a different logic, L_2. Wigglesworth then argues that this approach is too permissive, since it implies that sentential intuitionistic logic and sentential classical logic are equivalent, using the famous translations between them offered by Glivenko, Gödel, and others, a result he takes to be intuitively wrong.[36] But the main problem here is clearly the weakened criterion of equivalence, since it already implies that the negation-free fragments of *any* two logics are equivalent for any recursive logics. And this approach doesn't actually solve the logical form issue in full generality even on its own terms, since some logics don't include a primitive negation symbol.[37]

I think the best way to compare different logics is to use the notion of a semantic counterpart that I introduced in section III. This is key because when assessing fine-grained equivalence between different logics we must ensure that connectives are being mapped to each other in the right way, otherwise you could end up classifying clear notational variants as radically distinct logics. For example, consider two formulations of classical logic in which the symbols for conjunction and disjunction are systematically swapped but everything else remains the same.[38] In this case, a comparison based on the homophonic translation would be hideously bizarre—can it be coherent for people to be talking about disjunction, really, and yet think that it (disjunction) behaves *exactly* like conjunction and vice versa? I don't think so. Aside from the symbols itself, there would be no sense in which the symbol "and" would be a conjunction symbol in this alien language rather than a disjunction symbol. A symbol counts as a version of conjunction because of how it is used,

not because of some mysterious essence attaching to the word "and".

The problem of what makes a symbol in an arbitrary language a conjunction symbol (for example) was solved in the previous section by introducing semantic counterparts. So here let us say that a mapping can only be used to demonstrate true equivalence between logics if the mapping is a counterpart translation, in the sense discussed above. Obviously, a counterpart mapping won't always be available among the primitive notions of two logics, but one will sometimes be definable in simple terms. When we define counterparts, the assessment of counterparts isn't entirely independent of the other features of the mapping, but this isn't necessarily a problem. In other cases, there will be two or more counterparts between logics, each as good as the other. And in still other cases, a logic will contain no counterpart of one or more notions in the other, because nothing in the other logic plays a similar enough structural inferential role to count as a counterpart, whether primitive or defined. All of this is to be expected and was discussed in the previous section.

Let $cp(x)$ be the counterpart mapping from the connectives of $L_1$to the connectives of L_2, and say that the logic of L_1 is logically interpretable in the logic of L_2, just in case there is a recursive mapping $t : L_1 \mapsto L_2$, meeting the following two conditions:

(1) The mapping preserves derivability across logics, so that for set of sentence of L_1, Δ, and sentence of L_1, ϕ, if $\Delta \vdash_{L_1} \phi$ then $t(\Delta) \vdash_{L_2} t(\phi)$

(2) The mapping preserves logical form, in the sense that $t(\neg\phi) = cp(\neg)^\frown t(\phi)$, $t(\phi \wedge \psi) = t(\phi)^\frown cp(\wedge)^\frown t(\psi)$, and so on for the other logical constants[39]

Mutual logical interpretability is then logical interpretability in both directions. Let us then say that two logics are cognitively equivalent just in case they are mutually logical interpretable with each other.[40]

Assuming, as is plausible, that the negation and disjunction symbols in classical (C) and intuitionistic (I) logic are semantic

counterparts to each other, then we can see that C and I are not cognitively equivalent, since if they were, then there would be a translation f from L_C to L_I, meeting the above conditions. Since for all sentences in L_C, $\vdash_C \phi \vee_C \neg_C \phi$, then it must be that $\vdash_I f(\phi) \vee_I \neg_I f(\phi)$, but this isn't the case, since no instance of excluded middle is provable from the empty set using the basic rules of I. The more involved syntactic approach adopted here wards off the problematic overgeneration that critics have claimed afflict all syntactic approaches.

Is cognitive equivalence sufficient for metaphysical equivalence? There are general worries about using purely syntactic approaches to analyze worldly, fact-based notions of equivalence. For example, for standard interpretability approaches to equivalence between non-logical theories, Hilary Putnam noted that it is in principle possible for the structure of a theory to be the same while the theories are connecting up the world in entirely different ways.[41] We might finesse this by requiring that the mapping demonstrating logical equivalence also preserves coarse-grained truth conditions for all declarative sentences in the relevant languages, where a sentence's coarse-grained truth conditions can be represented by the set of possible worlds in which the sentence is true.[42]

Maybe there are contexts in which it makes sense to say that two logics are metaphysically equivalent when they are interpretable in something like the standard sense and the translation involved preserves coarse-grained truth conditions. There may also be cases where equivalence claims are themselves partisan, so that we might also want to require that both logics sanction the relevant equivalence claim, from their own perspective.[43] Since I won't be resting on any robust equivalence claims, I won't take any firm stand on these matters here. The central goal of this section has just been to demonstrate that, contrary to what some have thought, conventionalists and inferentialists have fine-grained, robust approaches to equivalence—even equivalence between logics—available to them. This overall point is relevant when assessing the general issue of if and how conventionalists can engage in standard debates in the philosophy of logic.

V Logical Disagreements and Disagreements about Logic

Where does the logical pluralism entailed by conventionalism leave disputes in the philosophy of logic? This question has no simple answer, since there isn't just one kind of logical dispute or disagreement. To start clarifying things, let's first distinguish between (*a*) logical disagreements and (*b*) disagreements *about* logic.

First take (*a*), logical disagreements. Say that two speakers are having a logical disagreement if they disagree about which sentences logically entail other sentences. There are a couple of ways that this can happen trivially. One is if the speakers defer to some common source or are making corrigible computational errors. In either of these cases, the speakers follow the same logical rules and so at least one of them is making a mistake. Such disputes happen between beginning students in the logic classroom, but they aren't the kind of logical disagreements that concern philosophers. The second way that logical disagreements can be trivial is when the speakers speak languages that are notational variants of each other, for example, if the symbols "∧" and "∨" have their meanings swapped in their respective languages. In this case, the speakers speak different languages, but their logical conventions don't substantively differ.

A *real* logical disagreement is one that isn't trivialized in either of these ways. So in a real logical disagreement, speakers disagree about the logical facts without either making some corrigible mistake or speaking a notational variant of the other's language. For inferentialists, such disagreements would have to bottom out in the two disagreeing parties adopting different linguistic conventions, but this makes them non-disagreements, as we've repeatedly seen. Both parties end up being correct in their own languages, against the backdrop of their own conventions. Accordingly, when you "really" logically disagree with someone, then the two of you don't actually disagree at all. Instead, you end up talking past each other. If one speaker is a classical logician and the other is an intuitionist, a disagreement about whether an instance of the law of excluded

middle is a logical truth is, for the most part, a waste of time for both parties. The debate is wholly and fully resolved by saying that excluded middle is a logical truth in the language of one speaker but not the other. There is no further independent matter of fact to argue about.

But conventionalism's deflationary treatment of logical disagreements doesn't mean that all debates in the philosophy of logic are pointless. To see this, we have to turn to (*b*) disagreements *about* logic. Disagreements about logic are most naturally engaged in by metalinguistically ascending to discuss particular logics *L* and L^* in relation to each other, but they often occur wholly in the object language and thus look like logical disagreements. Disagreements about logic come in at least three distinct forms that are often confused with each other. The resources of conventionalism allow them to be neatly distinguished:

(*b*1)Disagreements about which logic is correct in some particular natural language (call these descriptive disputes) (*b*2) Disagreements about which logic should be used (call these normative disputes)

(*b*3)Disagreements about which logic is objectively correct in some language-independent sense (call these metaphysical disputes)

All of these disputes can, at least in principle, be engaged in by conventionalists, but with perhaps more clarity than is usually involved. Let's take each type of dispute in turn.[44]

Conventionalists are well placed to engage in (*b*1)-style descriptive disputes, but adjudicating such disputes is, in part, a matter of empirical linguistics. To resolve descriptive disputes, in addition to adopting particular metasemantic and interpretive principles, such as the inferentialist principles of chapter 3, we would need to know whether or not English speakers (for example) really do reason according to modus ponens in order to decide whether modus ponens is valid in English. The dispute in the philosophy of logic between classical and relevant logicians often seems to be of the (*b*1) form: the relevantists point to the oddity of classically valid arguments when presented in English as evidence that the logic of

English is not classical.[45] Logical conventionalists can coherently engage in this debate (and others like it) on either or neither side.

Conventionalists are also well-placed to engage in (*b*2)-style normative disputes concerning which logic we *should* use. This is because conventionalists think that if we adopt some particular logic *L* then the rules and principles of *L* will be valid in our language. Deciding whether we should adopt some logic *L* at the expense of our current logic depends on many factors. We need to know both our goals and interests and also the properties of *L* in order to make a fair assessment of whether *L* will better serve our interests than our current logic. We must also factor in whether we would be able to use *L* and the "cost" of switching from our current logic to *L*. And additionally, since we use reasoning in these evaluations, there is no guarantee that there is any neutral place to "stand" when making these evaluations. Your endpoint may depend, in part, upon your starting point, logically speaking.[46]

The dispute in mathematics between classical logicians and intuitionists often seems to be of the (*b*2) form: the intuitionists admit that most mathematicians use classically valid but intuitionistically invalid forms of inference, but urge that they stop doing so for various reasons (invalidity, the threat of inconsistency, the uninformativeness of non-constructive proofs, *et cetera*).[47] Conventionalists can coherently engage in this debate (and others like it) on either or neither side.

(*b*3)-style metaphysical disputes are perhaps what disputants in the philosophy of logic most often take themselves to be engaged in, but conventionalists are unlikely to find this type of dispute particularly compelling. Unrestricted inferentialism ties the correctness of our logical principles directly to our use of language. Against this background, what could it possibly mean to say that the law of excluded middle is a logical truth in our language but that it isn't *really* a logical truth?

There are several recent theories in metaontology and substantive metaphysics aimed at making sense of ideas like this, but the heavily metaphysical flavor of metaphysical disputes and the metaphysical notions needed to make sense of them are unlikely to be endorsed by conventionalists.[48] As I have already mentioned several times, I

have elsewhere argued against these notions. I find them mysterious and think acceptance of them leads to both epistemological and metaphysical puzzles. Conventionalism is much better paired with a generally naturalistic view of the world, not a revival of medieval metaphysics (1.IV).[49]

The general approach to the philosophy of logic that I endorse and think that conventionalists should endorse is similar to the one endorsed by the arch-conventionalist Rudolf Carnap and summed up in his (in)famous *principle of tolerance*:

> The standpoint which we have suggested—we will call it the *Principle of Tolerance* ... relates ... to all questions of logic. ...
> The first attempts to cast the ship of logic off from the *terra firma* of the classical forms were certainly bold ones, considered from the historical point of view. But they were hampered by the striving after 'correctness'. Now, however, that impediment has been overcome, and before us lies the boundless ocean of unlimited possibilities.[50]
> ... the *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, i.e., his own form of language, as he wishes. All that is required of him is that, if he wishes to discuss it, he must state his methods clearly, and give syntactical rules instead of philosophical arguments.
> The tolerant attitude here suggested is ... the attitude which is tacitly shared by the majority of mathematicians.[51]

The principle of tolerance enjoins us to react to proposed alternative logics not with metaphysical condemnations or metaphysical arguments, but by assessing their practical usefulness as tools with which to communicate and investigate the world. In other words, philosophers of all stripes are well placed to engage in (*b*2) disputes about logic, but these are best carried out by clearly distinguishing them from the metaphysically suspect (*b*3) disputes about logic, lest serious confusions impede our investigations. And even for those inclined to engage in (*b*3) disputes, clearly distinguishing these

disputes from others is best done using something like the framework developed here.

I suspect that only logical conventionalism—and the logical pluralism that it entails—provides us with the conceptual resources and the metaphysical clarity to see these and many other matters aright. Conventionalism does not make logic arbitrary, nor does it completely wash away all of the varied practical and historical reasons we might have for preferring one logic over another. Logical truths and logical notions are reflections of the rules of our language, and language is the most important and cherished of our cultural and cognitive tools. Decisions about languages and thus, logics, are among the most important that we face. With conventionalism accepted, we can face these decisions without confusion.

≈≈≈≈≈≈

UNRESTRICTED LOGICAL INFERENTIALISM leads to logical conventionalism and logical pluralism. The nature of this pluralism is somewhat elusive, but the effort of clarifying it, undertaken in this chapter, pays off by giving us a clear view of the nature of both logical disagreements and the wide variety of disagreements about logic. This chapter completes my three-chapter presentation of logical conventionalism. The next chapter applies my conventionalist theory to the epistemology of logic as a further illustration of the view's power and significance, and the following chapter answers various objections. Part III's account of mathematics will provide a more complicated version of conventionalism that naturally extends the theory of logic developed in this part.

---

1 I take the lovely phrase "galloping pluralism" from Crispin Wright, who used it in conversation when describing my view several years ago.

2 In my (2015*b*).

3 This example is based on one from Priest (1998).

4 This principle figures in the argument against homophonic translation where logics differ from my (2018); my argument there is a variation on one from Quine (1970*b*).

5 Most famously in Davidson (1974).

6 This is one reason that my argument improves upon Quine's (1970*b*) infamous "change of logic, change of subject" argument, see again my (2018) for discussion of Quine's argument and a version of my own without explicit reliance on inferentialism. And see Morton (1973) for an important early critical discussion of Quine's argument.

7 See Humberstone (2000*b*), where such extensions are called "contra-classical" logics.

8 See Kuhn (1962), Feyerabend (1975), and Lakatos (1970).

9 See section 4 of my (2018); see also the discussion of charity in my (2016*c*).

10 Kripke (1980), Putnam (1975), and Burge (1979) are the canonical semantic externalist writings. See Morris (2018) for a popular account of how all of this leads to an objection to Kuhnian philosophy of science.

11 Once again see Brandom (1994) and Sellars (1953) on this point.

12 I think that "tonk" is itself a logical notion, since it plausibly meets my criteria for being logical. This might be denied since in a trivial tonk language no sentence expresses contingent truths, thus perhaps undercutting topic neutrality.

13 See the discussion in my (2015*b*).

14 For much more on all of this, see my (2015*b*).

15 For discussion of some of these variations see section 5 of my (2015*b*).

16 Among others, Burgess and Burgess (2011), Chihara (1979), and Wittgenstein (in Diamond (1976)) have flirted with the idea that our actual truth principles are inconsistent. Some of these authors have noted that speakers are not actually disposed to exploit inconsistencies. I think this is of crucial importance for understanding the liar and related paradoxes. I plan to investigate this matter in future work. The distinction between complex and composite dispositions introduced in my (forthcoming *a*) provides a potentially useful tool with which to analyze this approach (see also 2.VII). I don't mean to suggest that these authors all have similar views, see Chihara (1977) for criticism of Wittgenstein's approach.

17 In Glivenko (1929).

18 Of course, the tonk rules are only as powerful as they are given certain background assumptions; see Cook (2005) for extended discussion.

19 As noted in 3.IV, the bilateralist system of rules I have used is separable. This is worth noting since standard classical systems are not separable.

20 This narrow holism also allows us to get around the problems Kroedel (2012) tries to raise for implicit definition accounts of logical knowledge, though I think even rigidly non-holist views can probably avoid these problems.

21 See Beall and Restall (2006); see Restall (2002) for arguments that this kind of pluralism doesn't lead to a conventionalist-style pluralism of the kind I endorse.

22 I think this is important and under-recognized; the semantic counterpart approach to pluralism I develop below is also discussed in my (2015*a*) and

(2015*b*).

23 See Priest (1979) and (1987) for LP and a defense of dialetheism.

24 See Lewis (1968).

25 Something like this is done in Paseau (2019).

26 In the terminology of Paseau (2019), this, in effect, leads to a distinguished supervenience base, avoiding some (but not all) problems involved in "counting".

27 Harman (1973) argued that since translation is based on similarity in use, it is not transitive, so any assumptions of *meanings* being shared across translation is mistaken. I deny that correct translation is simply based on similarity in use—the counterpart translation is not automatically correct, even when it is available.

28 As an illustration of this, arch-translation indeterminist Quine (1960*b*) infamously argued that paraconsistent negation can't plausibly be translated into our classical negation for reasons of charity; see the discussion of deviant logics in chapter 6 of Quine (1970*b*). For a full discussion of Quine's arguments, see my (2018).

29 I have argued against these notions elsewhere: see my (2016*a*), (2016*b*), and (2016*c*).

30 Relatedly, in set theory, many—Boolos (1971), for example—have claimed that the axiom of extensionality must govern any concept of set. See Quine (1970*b*) and Burgess (2004) for related notions of "analyticity".

31 See the famous discussion of conventionalism in Poincaré (1902).

32 See Glymour (1970), apparently anticipated by de Bouvère (1965).

33 This point is made in Wigglesworth (2017). For some other writings on logical equivalence, see French (2019), Segerberg (1982), and Woods (2018).

34 See Tarski, Mostowski, and Robinson (1953).

35 See Wigglesworth (2017) and Feferman (2000). In the appendix to my (2015*c*) I weaken the quantifier condition on relative interpretability while keeping the form conditions for sentential logic in place, thus allowing the notion to be applied to the different metaphysical theories considered by ontologists.

36 See Glivenko (1929), Gödel (1933), and Gentzen (1936)—the Gödel/Gentzen mapping works for first-order logic, not just sentential logic.

37 As was implicit above, Wigglesworth (2017) requires that for each pair of a set of sentences and a sentence there be an appropriate translation t meeting the equivalence condition. Clearly though, it is more appropriate to require (as I will here) that there be a translation t, such that all pairs of sets of sentences and sentences meet the equivalence condition. This point is made by Woods (2018).

38 An example like this is given in Haack (1974).

39 This condition assumes for simplicity that the logics share structurally identical grammars—conjunction symbols written between their conjuncts, for example

—but this is not essential. For generality, we require that g be a mapping that takes each sequence of symbols of L_1 to a sequence of symbols of L_2 that is the "same" *modulo* grammatical formation rules and subsentential translations. So that, for instance if p, q are grammatical sentences of L_1 which correspond to a, b in L_1, L_2 's counterpart of "∧" is "&", and L_2, unlike L_1 (which we assume to be grammatically standard), writes conjunction expressions before their conjuncts in grammatical sentences, then $g(p \wedge q)$ = "&*ab*". This is the same move made above when defining "semantic roles" as *structural* inferential roles.

40 In some ways my approach here is analogous to approaches to equivalence and related notions in computer science; see Felleisen (1991).

41 Putnam (1983) proposes preservation of goodness of scientific explanation as his worldly supplement to relative interpretability. A realistic case where relative interpretability breaks down as a plausible account of equivalence is in basic arithmetic, where Nelson (1986) has shown that Robinson arithmetic is mutually relatively interpretable with some of its proper extensions.

42 See Stalnaker (1984) for coarse-grained approaches to propositions as sets of worlds and Kaplan (1989) for the complication of this story using context. This requirement is similar to a condition on equivalence used by Eli Hirsch (2011) in his discussions of metaontology. Another option is Putnam's (1983) requirement that the mapping preserves explanatory goodness. Another option is to use *strong interpretability*, requiring not only that theorems map to theorems, but also that non-theorems map to non-theorems. See chapter 4 of Bonevac (1982) for some discussion. A variant of this option might cohere well with bilateralism.

43 Something like this approach, on its own, without most of the other modifications above, is pursued in Woods (2018).

44 Something like the contrast between (b1) and (b2) is drawn in Burgess (1981) and Hanson (1989).

45 For relevant logic (sometimes called "relevance logic") see Anderson and Belnap (1975) and Read (1988).

46 See Colyvan (2006) and Woods (2019) for relevant discussions.

47 See Dummett (1973*a*) for a philosophical defense of intuitionistic logic.

48 See Fine (2001) and Sider (2011) for examples of the kind of heavyweight metaphysics needed to make sense of (b3) disputes.

49 Up until this point this section's content partly overlaps with section 5 of my (2018).

50 Carnap (1937), page xv.

51 Carnap (1937), pages 51–52.

6

The Epistemology of Logic

Conventionalism has long been seen as holding the promise of naturalizing the epistemology of logic. Even critics of conventionalism typically admit that if the view were correct it would completely demystify the hitherto mystifying epistemology of logic. But the exact details of this promised conventionalist epistemology were never worked out by historical conventionalists. At best, conventionalists recognized and stressed that, unlike many traditional theories of logic, conventionalism would allow us to eschew scientifically disreputable and borderline magical faculties like *intuition* or *pure reason* in logical epistemology.

Surprisingly, such faculties have made something of a comeback in contemporary philosophy.[1] A major reason for this is that inferentialists and naturalists have failed to provide a fully satisfying *a priori* epistemology even for basic logic. But I think that unrestricted inferentialists, and thus, conventionalists, can do so quite simply. Unsurprisingly, the major epistemological stumbling block has been stubborn adherence to restricted inferentialism at all costs. In light of this, the simple conventionalist epistemology of logic in this chapter lends further theoretical support to unrestricted logical inferentialism.

I The General Strategy

How do we know that contradictions are never true or that modus ponens in valid? The natural inferentialist approach, with which I agree, is based on a simple but powerful idea: we come to know various logical truths, that certain rules are valid, and so on, by proving to ourselves that they are using the basic rules of our language. Proofs are a standard source of knowledge and

justification, especially in mathematics. But in the special case of logical knowledge various circularity worries arise, so we must tread carefully. This section lays out the background required for this careful tread.

I will focus on knowledge, but similar epistemological questions arise for justification. Fortunately, questions about justification can be answered quite similarly, and I'll make a note of any points where my approach to logical knowledge would require modification in order to apply to justification.

Proof-based knowledge of *logical truths* seems fairly uncontroversial, even among non-conventionalists, but conventionalists have a particularly tidy account. How can we know any instance of the law of excluded middle (*LEM*)? We can know it because we can prove it. And the exact same thing goes for any other logical truth as well (4.I and 4.III). And proofs of instances of excluded middle won't use excluded middle as an axiom, and in fact won't contain any undischarged axioms at all. For these reasons, *a priori* knowledge of logical truths is satisfyingly explained via proof without any serious complications or residual mystery. The real difficulties arise in applying this "prove it" strategy to our knowledge of validity claims. For this reason, my focus in the remainder of this chapter will be on the epistemology of validity.

In order to prove claims about validity, we first need to be able to express claims about validity in our object language. Perhaps the simplest way to do this is to construe validity claims as claims about necessary truth preservation. As discussed in chapter 3, in order for this to be coextensive with the notion of logical validity the necessity here must be broad enough to cover all coherently describable scenarios. I called this "strong validity" in 3.I. For simplicity, throughout this chapter, to say that a rule of inference *R* is *valid* is to say that it is necessarily truth-preserving in this sense.[2] In fact, we can simplify further: since it won't make any essential difference, I'll drop the necessity claim and construe validity as simple truth preservation. More complex and nuanced statements of validity could also be offered, but the conclusions I arrive at for this simplified statement carry over to more complex statements of validity as well, including those using necessary truth preservation,

those using higher-order quantification, those using the semantic turnstile, those explicitly quantifying over interpretations of our language, those using a primitive validity-predicate, and so on.

The simplified formulation still requires the use of a truth predicate, "*T*", in the object language. As is standard, we can take this truth predicate to apply to formulas of our language by way of a name-forming device, such as quotation, which I will again represent using square brackets: "[...]".[3] There is a great deal of philosophical and technical debate about the nature and logic of truth, but we can bypass most of these debates here.

I will continue to use sentences as truthbearers, but as discussed in 2.V and 3.I, I don't think this is a truly partisan assumption. The sentences of logic and math that concern us are neither context-sensitive nor indexical, so there is little difference here between ascribing truth to sentences as opposed to utterances or mental tokens. And since we are dealing with interpreted sentences there is little difference between assigning truth to sentences themselves as opposed to the propositions or contents they express. But keeping our focus on sentences aids in clarity when discussing alien languages and translation (as in chapter 5).

In terms of the logic of truth we need make only the relatively uncontroversial assumption that the truth predicate is implicitly defined, at least partially, by the following introduction and elimination rules:

$$(T[...]I) \quad \frac{\phi}{T[\phi]} \qquad\qquad (T[...]E) \quad \frac{T[\phi]}{\phi}$$

Of course, when we have these rules together with classical logic in a language that can talk about its own syntax, paradox soon descends upon us. But while the issues concerning the paradoxes are interesting, they are orthogonal to the epistemological issues that concern me in this chapter, so those who are bothered by the liar and related paradoxes should tacitly restrict these rules to sentences in the truth-free fragment of the language.[4]

All of our logical rules have only finitely many premises, and when a rule *R* has only finitely many (possibly schematic) premises, validity claims about *R* can be expressed by saying that if the conjunction of truth claims for each of *R*'s premises holds, then *R*'s conclusion is also true. So the validity of (*MP*) is expressed by the following schema:

$$\ulcorner((T[\phi] \wedge T[\phi \rightarrow \psi]) \rightarrow T[\psi])\urcorner$$

In words, this schema tells us that if the premises of an (*MP*) argument are true, then the conclusion is also true. It will be useful to focus on the case of standard and familiar inference rules like (*MP*), but roughly the same general approach also applies to bilateralist rules that use rejection, by adapting the bilateralist notion of validity of chapters 3 and 4. For example, the extended validity of the rule (¬*E*) can be expressed using a defined "falsity" predicate, "*F*", as follows:

$$\ulcorner(T[\neg\phi] \rightarrow F[\phi])\urcorner$$

In this particular case validity follows from the mere definition of "falsity" as truth of negation, but the overall point is quite general.

Some may worry about statements of validity requiring all of these various conceptual resources (truth, the conditional) but I don't think this should concern us. Validity is a complex notion, and reasoning and thinking about it requires a bit of conceptual sophistication. Even while admitting this, some may also object that schematic letters already do the work of generalization, thus making the truth predicate redundant. But any such simplifications would only streamline my approach and make it easier, so I'll stick with the more natural and intuitive but slightly more complicated truth-involving statement of validity throughout this chapter.[5]

Can we extend the proof-based account of logical knowledge to validity claims? Unfortunately and in stark contrast to the epistemology of logical truths (or falsehoods), extending this approach to the epistemology of validity involves serious

complications. The central issue involved is well known: proofs of validity for basic rules of our language are *rule-circular*—the proofs use the very same rules that they are attempting to prove valid. To illustrate, here is a rule-circular proof that (*MP*) is valid:[6]

1.	$T[\phi] \wedge T[\phi \rightarrow \psi]$	{1}	*assumption*
2.	$T[\phi]$	{1}	$\wedge E(1)$
3.	$T[\phi \rightarrow \psi]$	{1}	$\wedge E(1)$
4.	ϕ	{1}	$T[...]E(2)$
5.	$(\phi \rightarrow \psi)$	{1}	$T[...]E(3)$
6.	ψ	{1}	$MP(4,5)$
7.	$T[\psi]$	{1}	$T[...]I(6)$
8.	$((T[\phi] \wedge T[\phi \rightarrow \psi]) \rightarrow T[\psi])$	Ø	$CP(7)$ *discharge*(1)

As this proof shows, once we have our truth rules, it is quite easy to establish the validity of (*MP*), provided that we use the (*MP*) rule, itself, in the proof.[7]

Rule circularity contrasts with the more traditional notion of *premise circularity*: a premise-circular argument's conclusion appears as one of its premises. While premise-circular arguments have been almost universally rejected in epistemology, the status of rule-circular arguments isn't so clear. The obvious problem with premise-circular arguments is that they can be used to prove *anything*; but something analogous holds for rule-circular arguments too. *Any* rule can be used to establish its own validity via a rule-circular argument, at least against the background of the standard rules for the connectives along with the structural rules. This has been tacitly noted in the literature, but it hasn't been explicitly stated, as far as I'm aware.[8]

In other words: all natural deduction rules in standard systems can be used to prove their own validity using rule-circular arguments. This means that simple valid rules, complex valid rules, inductively good but invalid rules, and crazy invalid rules can all be used to give rule-circular proofs of their own validity. A rule-circular approach to logical epistemology keeps bad company. And the problem is so

general that rule-circular arguments for validity clearly cannot be good *as such*. Because of the wide heterogeneity of bad company, there is no simple trick for screening out problem cases. But unrestricted inferentialism cuts through this tangle, allowing a comprehensive proof-based account of knowledge of validity to be developed. I'll make a start by considering the general conditions that arguments need to meet in order to potentially lead to knowledge of their conclusions.

II Arguments and Knowledge

My concern here is not whether rule-circular arguments can ever be persuasive, I'm concerned only with their *epistemic* worth, if any. Can you ever attain knowledge of the conclusion of a rule-circular argument by coming to believe the conclusion on the basis of the argument? If not, the "prove it" approach to knowledge of validity is a non-starter.

But it isn't ad hoc to think rule circularity is not automatically bad, epistemically speaking. In fact, the idea that rule-circular arguments might have *some* epistemic worth is an old and fairly popular one.[9] Most saliently here, Paul Boghossian explored this idea in the context of inferentialism in an influential series of papers.[10] Here I will begin by following Boghossian's trail, though using my own setup and terminology, before branching off at the point where *unrestricted* inferentialism overcomes the difficulties that plagued his approach.

To start let's consider, in general, when an argument for "*p*" from (possibly empty) premise set Γ, using rules $R_1, \ldots, R_n$ can possibly be used to found knowledge of "*p*".[11] I use the term "argument" here rather than "proof" to flag that this discussion is not meant to apply only to deductive or mathematical proofs, but instead to any kind of argument for a conclusion. Two basic necessary conditions are straightforward:

(1) The argument's conclusion must be true
(2) The argument's premises must support its conclusion

For (1): obviously, you can't know something that isn't true—in some inductively strong arguments you can do everything that is epistemically required of you and still fail to gain knowledge of the conclusion simply because it isn't true. For (2): support doesn't have to be as strong as validity, since unless we are inductive skeptics, we will think that strong inductive arguments can sometimes lead to knowledge of their conclusions. Three directly epistemic necessary conditions are also fairly straightforward, though they are easier to state if we call the argument-using subject "*S*" and the argument itself "*A*":

(3) *S* knows all of *A*'s premises
(4) *S* does not know *A*'s conclusion before inferring it using *A*
(5) *S* infers *A*'s conclusion on the basis of *A* while retaining knowledge of all of *A*'s premises throughout the inference (usually but not always this will involve coming to believe *A*'s conclusion on the basis of *A* itself)

For (3): obviously you can't come to know the conclusion of an argument via inference without knowing the premises of the argument. For (4): you can't *gain* knowledge of something you already know. For (5): we're interested in when arguments can be used to gain knowledge of their conclusions, so we have to consider only cases where an argument is *actually used* in coming to a conclusion. All of these principles are relatively uncontroversial.

So we have five necessary conditions on the use of an argument to gain knowledge of its conclusion, but are they jointly sufficient? There are two major sources of worry about sufficiency, one that I will set aside, one that will be discussed in detail. The worry I'll be setting aside concerns puzzles about failure of warrant transmission.[12] There seems to be something problematic about reasoning from "I see a katana in front of me" and "there is a katana in front of me" to the conclusion that "my vision is working correctly in this case"—or, at least, it seems problematic when my justification for the second premise comes *entirely* from the first premise. This kind of bootstrapping challenges approaches like mine, but while related issues will concern us below, traditional bootstrapping

worries won't directly arise. This is because the rule-circular arguments we're considering lack undischarged premises, and so worries that our justification for acceptance of the premises is problematically intertwined in a way that defeats knowledge will be idle. Partly for this reason I'm going to ignore these issues here; if they nag at you, simply imagine that we've added a condition to the above list saying that the justification on which *S*'s knowledge of each of *A*'s premises is based is *suitably independent* of the justification on which *S*'s knowledge of *A*'s other premises is based.

The worry about joint sufficiency that is relevant concerns the role of *A*'s inference rules. Do we need to add some epistemic requirement that *S* must satisfy with respect to *A*'s inference rules if *S* is to use *A* to gain knowledge? It seems clear that we do when we consider cases like the following: recall that the fundamental theorem of algebra states that any *n*th degree polynomial with complex coefficients has *n* (perhaps repeated) solutions in the complex numbers. The theorem was first proved by the great German mathematician Karl Friedrich Gauss in his 1799 doctoral dissertation. Let us now imagine a similar situation to our own, except Gauss is *lazy*:

> *Lazy Gauss.* Tired of working on a dissertation attempting to prove the fundamental theorem of algebra while at the same time revolutionizing number theory in the *Disquistiones Arithmeticae* and inventing non-Euclidean Geometry, the young Gauss decides to save himself some trouble by adopting the following inference rule:
>
> $$(Algebra) \quad \frac{\alpha \text{ is an n}th \text{ degree complex polynomial}}{\alpha \text{ has } n \text{ complex solutions}}$$
>
> After turning in his one-page proof of the fundamental theorem of algebra using this rule, Gauss is shocked to find his work rejected and himself strongly chastised by his examiners.

It's hard not to sympathize with lazy Gauss's examiners here. Clearly there is something *epistemically* inappropriate about Lazy Gauss's

procedure despite the fact that (*Algebra*) is a valid rule. It seems quite natural to suppose that in order to use a complex rule like (*Algebra*) to gain knowledge, we first need to demonstrate that the rule is truth-preserving. In other words, in the case at hand we'd first need to *prove the fundamental theorem of algebra*.[13]

Of course, some will not be convinced; they will argue that lazy Gauss did *know* the fundamental theorem of algebra, he just didn't *know that he knew*, and neither did his examiners know that he knew, *et cetera*. But these moves aren't very appealing. More plausibly, cases like Lazy Gauss's demonstrate that we need to add some epistemic condition on inference rules to our five conditions above in order to correctly analyze our actual epistemic norms and practices.

The natural condition to add demands that *S* know that every inference rule in *A* is, at least, generally reliable (either valid or inductively strong), but unfortunately adding such a condition delivers intuitively incorrect results. Although the exact analysis of knowledge claims is controversial and problematic, it is generally agreed that knowledge that *p* requires belief that *p*.[14] This means that if *S* is to *know* that some rule *R* is reliable, then *S* must surely *believe* that *R* is reliable, and this means that *S* must have beliefs about the reliability of inference rules like *R*. *S* must think that *R* is truth-preserving *most of the time*, or something similar. Now consider the following case:

> *Simple Sam.* Simple Sam is happy, but conceptually he's in a sorry state. He can perform basic logical reasoning using modus ponens (*MP*) and other simple rules, but he completely lacks the concept of truth (in the sense that he has no beliefs involving the truth predicate and, *a fortiori*, no beliefs about the reliability of (*MP*)). One day Simple Sam is on the way to visit his girlfriend, Simple Sally. He knows that if her window is open, then Sally is home; and as he gets near the house he can see and thus comes to know that her window is open. Then Sam uses a bit of (*MP*) reasoning to conclude that Simple Sally is home.

There is tremendous intuitive pressure in the *Simple Sam* case to say that Sam comes to know that Sally is home on the basis of a simple (*MP*) argument despite not believing and thus not knowing that (*MP*) is reliable.[15] This might be denied by saying that even simple agents somehow *tacitly* know that the rules they're using are valid, but again, I find this implausible.

Another response claims that as cognitive sophistication increases, so too do epistemic demands; so poor Simple Sam can come to know things using inferences without knowing anything about the rules used in his arguments, but we cannot. Alas, the burdens of the brainy. I don't find this at all plausible either; the differing demands here seem to be *dialectical* rather than *epistemic*. Children lack the cognitive ability to explain and defend their reasoning and so too does Simple Sam, but we don't, so it's legitimate to ask us, but not them, to offer defenses in particular cases. Yet an inability to defend your practice dialectically is not at all the same thing as the practice being in epistemic goodstanding.

Boghossian in effect argues, and I agree, that the situation here shows that in order to have a necessary and sufficient set of conditions on *S*'s ability to use argument *A* to found knowledge of *A*'s conclusion, we need to add some requirement that *S* must meet vis-à-vis *A*'s inference rules. But, as the Simple Sam cases demonstrates, this requirement cannot, in general, demand that *S* *know* that the rules used in *A* are valid or generally reliable.

What we need is a more subtle way to evaluate rules of inference epistemically. Fortunately, there is a notion in the epistemological literature that seems to provide exactly what is called for here: *entitlement*. The terminology was introduced by Tyler Burge but has been subsequently used in slightly different ways by a number of philosophers.[16] In the present context, Boghossian uses the term in a wide sense allowing that not just mental states like belief but also inferences and transitions between mental states can be the subject of entitlements. I will follow him in this, taking care to note that entitlements are context-sensitive and that they can in principle apply to transitions between mental states that are causal or non-causal, voluntary or involuntary, and so on. Perhaps most importantly here,

the epistemic entitlement to use an inference rule *R* does not require prior knowledge of *R*'s validity.

With all of this in mind, we can add a principle concerning entitlements to use certain inference rules to conditions (1)–(5) from above:

(6) *S* is epistemically entitled to use all of the inference rules in *A* and retains those entitlements throughout the process of inferring *A*'s conclusion

It seems plausible that, modulo concerns about bootstrapping (which are easily handled, in any case), conditions (1)–(6), schematic though they may be, provide a necessary and sufficient condition for *S* using *A* to gain knowledge of *A*'s conclusion. With this account in hand, we can easily see why premise-circular arguments are epistemically worthless—in a premise-circular argument the conclusion appears as one of the premises, so to gain knowledge using the argument requires that there is a point in time at which you both knew the conclusion (by (3)) and also didn't know the conclusion (by (4)), but this is obviously impossible.

One point of clarification is worth noting. In this discussion I have been acting as if there is always a single argument that a subject is using. But if arguments are individuated in a purely syntactic manner, this is implausible. To think otherwise is to think that there must be a fact about whether *S* used argument A_1 versus A_2 when these arguments differ only in their notation and not at all in their structure. Is there always a unique argument being used modulo this kind of notational variation?[17] Probably not, though I think in most cases of interest there is very little indeterminacy here. The argument being used is determined by *S*'s dispositions, even if it is not apparent in the moment which argument *S* is using. When there is indeterminacy about this, we can still take conditions (1)–(6) as necessary and sufficient by *supervaluating* over all of the admissible arguments, where admissibility here is understood as *admissible as* the argument here used by *S*. *S* gains knowledge of the conclusion *p* just in case any admissible argument *A* is such that conditions (1)–(6) are satisfied.[18]

Before applying this account directly to the "prove it" approach to knowledge of validity, it is worth pausing to make an important meta-epistemological observation about something that may be worrying some readers: the notion of entitlement is, as I am understanding it, wholly compatible with basic scientific naturalism. I am not at all thinking of entitlements as inalienable and laid down by the gods. It involves no metaphysical weirdness or cognitive spookiness (1.IV). Our epistemological practices are important and useful, and I see the task of this chapter as explaining how they extend to knowledge of logical validity. As such, my approach is somewhere *between* an analysis and an explication. We are not merely taking our practices as they are on the surface, but nor are we significantly revising and altering them. The notion of an epistemic entitlement is an inchoate part of our epistemic theorizing, and by recognizing it, we can gain a better grip on our more standard and explicit epistemic notions, such as knowledge and justification. I cannot speak for others who have used the notion, but that is the role that the notion of entitlement is playing in my theorizing and it should be so understood throughout this discussion.

III Inferentialism and Epistemic Entitlement

Given section II's discussion, can rule-circular arguments, such as section I's rule-circular argument for the validity of modus ponens, ever lead to knowledge of their conclusions? The question turns on whether we are epistemically entitled to use (*MP*) in this context, but we don't yet know what this requires. As the discussions of Lazy Gauss and Simple Sam show, what we need is a *mixed* account—it cannot paint every rule, nor even every valid rule, with the very same epistemic brush when it comes to entitlements.

Happily, inferentialism is perfectly placed to provide the needed mixed account. The natural idea, first developed explicitly and in detail by Boghossian, but seemingly endorsed implicitly by many other inferentialists, is that we are automatically entitled to use meaning-determining rules of inference. Boghossian call this the

meaning entitlement connection, which, transported into my framework, can be stated as follows:

> **Meaning Entitlement Connection (MEC)**: Speakers are automatically epistemically entitled to use the basic rules of their language in all contexts.

The **MEC** claims that we don't need to do anything to earn the epistemic right to use the meaning-constituting rules of our language, since we *automatically* have an unearned entitlement to these rules.[19] Boghossian argues for the MEC as follows:

> Suppose it's true that my taking A to be a warrant for believing B is constitutive of my being able to have B-thoughts (or A-thoughts, or both, it doesn't matter) in the first place. Then doesn't it follow that I could not have been epistemically blameworthy in taking A to be a reason for believing B, even in the absence of any reason for taking A to be a reason for believing B? For how could I have had *antecedent* information to the effect that A is a good reason for believing B, if I could not so much as have had a B-thought without taking A to be reason for believing B in the first place? If inferring from A to B is required, if I am to be able to think the ingredient propositions, then it looks as if so inferring cannot be held against me, even if the inference is blind.[20]

Translated into my terminology this suggests something like the following argument: assume that (*MP*) is meaning-constituting for the arrow ("→"); then in order to have knowledge that (*MP*) is valid one must have a belief involving the conditional (involving the arrow symbol with its usual meaning) but it is impossible to have such a belief if one doesn't use the arrow symbol according to (*MP*) and (*CP*), so it would be impossible to come to know that (*MP*) is valid without reasoning according to (*MP*), ergo one is always entitled to reason according to a meaning-constituting rule like (*MP*).

Of course, in chapter 3 I first treated the conditional as a defined expression, and then introduced bilateralist basic rules for the

conditional in 3.VI. But there I also showed that those rules could prove the standard versions of (*MP*) and (*CP*), so nothing is lost and some familiarity is gained by proceeding with the details of Boghossian's given example.

Boghossian's argument tacitly assumes a plausible "ought implies can" principle—if one ought to *F* then one must be able to *F*. Contraposing: if one cannot *F*, then it isn't the case that one ought to *F*. So in the case at hand one cannot try to establish the validity of (*MP*) without reasoning according to (*MP*), so it is not the case that one ought to try to establish the validity of (*MP*) without reasoning according to (*MP*). Ergo: if (*MP*) is a meaning-determining rule for the arrow symbol as you use it, you are automatically entitled to use (*MP*) *even when trying to establish* (*MP*)*'s validity*.

Boghossian's argument is intimately related to the following line of thought: if one is trying to establish what is expressed by some sentence "*p*" where $c_1, \ldots, c_n$ are various ingredient meanings or concepts that are part of what is expressed by "*p*", then there can be nothing epistemically inappropriate or problematic about simply using the fundamental rules and procedures for operating with $c_1, \ldots, c_n$. And this is true even if there are other ways of trying to establish what is expressed by "*p*" available to us, for example, consulting a local oracle or flipping a coin.

Though related to his argument, I think that this way of putting things advances somewhat on Boghossian's, since his quoted argument almost seems to assume that following the rule (*MP*) is so fragile that it is destroyed by merely failing to employ (*MP*) in *some particular instance* where one is using the conditional, but that is implausible. In my discussion of inference rule-following in 2.VIII I endorsed a more robust and realistic version of dispositionalism. My type of dispositionalism is compatible with Boghossian's line of thought if understood in the way I've framed it here. So we don't need to accept an implausible fragility about rule-following in order to argue for the MEC.

Related to these variations on Boghossian's theme, there is a very general Carnapian reason for inferentialists to accept the MEC. According to inferentialism (chapter 3), the meaning-constituting rules are basic rules for the logical constants. The basic rules are

those that we follow simply because we do, not because we have established them in some other manner. In effect, choosing basic rules is choosing a language, and this is a non-epistemic action to which we clearly have an entitlement, vacuously. Just as we have a vacuous *epistemic* entitlement to play chess or go to the gym. Choosing a language is a precondition for substantive epistemic achievement, not something that is itself a substantive epistemic achievement. And thus we clearly have the epistemic right to play any language game we choose, even if there are better language games available. This involves having an entitlement to the basic rules of our language, whichever rules they may be. So the unrestricted inferentialist has available a general argument for entitlement to the basic logical rules.[21]

To use some terminology from chapter 4: if *R* is a constitutive rule of a practice, then one is always *epistemically* entitled to use *R*. This point makes good on a claim I made in 4.VIII, that the constitutive role of meaning-determining inference rules actually gives rise to a normative/regulative function for those rules rather than being in tension with it. This line of thought further strengthens the inferentialist case for the MEC and seems like something that many historical conventionalists tacitly endorsed.[22]

Now, with these points in mind, let us again consider section I's rule-circular argument for the validity of (*MP*): the only inference rules used in the argument were conditional proof (*CP*), truth introduction ($T[\ldots]I$), truth elimination ($T[\ldots]E$), and elimination ($\wedge E$), and modus ponens itself. It is plausible that all of these rules are basic, meaning-constituting rules of our language, and so we are entitled to use them even in this context. So if the other conditions ((1)–(5) from section II) are satisfied our rule-circular argument can found knowledge of (*MP*)'s validity. But in rule-circular arguments like ours these other conditions will be satisfied trivially, and so the "prove it" strategy can be applied to explain our *a priori* knowledge of validity, despite rule- circularity.

Two clarifications about this must be stressed: (*i*) nothing in the inferentialist picture allows us to use rule-circular arguments to explain our knowledge of the validity of non-meaning constituting but still valid rules of inference like *Algebra*. A rule like *Algebra* is not a

basic rule of our language. It is a derivative rule, and so we don't have an automatic entitlement to it, rather, we must *earn* such an entitlement by proving it to be valid using rules we are already entitled to and, in fact, this process neatly reconstructs actual mathematical practice.[23] And (*ii*) the inferentialist story I'm endorsing has a crucial externalist element: knowing or believing that (*MP*) is meaning-constituting plays no role whatsoever in the account of knowledge of (*MP*)'s validity. It is simply the external fact that (*MP*) is meaning-constituting that gives us the right to use it without having to do any collateral epistemic work. This also means that we can only use (*MP*) to successfully establish the validity of (*MP*) if we are, as a matter of external fact, using a language in which (*MP*) is a meaning-constituting rule. This externalist element is crucial. It is the rules you follow, in actual fact, rather than your beliefs about which rules you follow, that matter. This is exactly what we want and expect, I think.

Bearing point (*ii*) in mind is also a helpful corrective against overstating what has been established here—being in the state of knowing that (*MP*) is valid is not obviously the same thing as being in a state of being able to claim, with justice, that you know that (*MP*) is valid. Much less is it being in the even more demanding state of knowing that you know that (*MP*) is valid. In order to reach such a position, one might also have to convincingly argue that all of the background conditions required by our account of logical knowledge are met. Chief among them, one would need to be in a position to argue that (*MP*) is meaning constituting for the conditional symbol(s) in natural language; and this involves entering into a (*b*1) descriptive dispute about logic (5.V). Losing sight of this crucial externalist component distorts the epistemology of logic beyond recognition.

IV Clearing Out the Bad Company

The just-developed inferentialist approach to our *a priori* knowledge of validity is so natural and simple that you might wonder why anyone has ever felt compelled to instead posit spooky mental abilities to account for our knowledge of logic. The answer, as it so

often is when inferentialism is involved, is the problem of bad company.[24]

Cases like tonk inspired Boghossian to alter the MEC to be friendly to restricted inferentialism. The hard part is doing this in a way that preserves our reasons for accepting the MEC in the first place. Without discussing the various epicycles, I don't think this endeavor has met with much success. Fortunately, unrestricted inferentialism once again cuts through the tangle with a single sword-stroke. In particular, chapter 5's discussion of alternative logics and the translation mistake can be adapted to show that there is no need at all to restrict the MEC in order to deal with bad company.

Some of the putative problems can be dealt with by all inferentialists— including restricted inferentialists—simply by appealing to the externalism that was already built into the account. For example, why can't English speakers obtain knowledge of the validity of the invalid rule of affirming the consequent, via a rule-circular proof? Well, since the argument's conclusion isn't true in our language, then by (1) of section II, this possibility is already trivially ruled out. And even if we are considering only *justification* for acceptance of the validity of affirming the consequent, a rule-circular proof doesn't help, since we English speakers don't actually follow affirming the consequent as a basic rule of our language, so the MEC does not entitle us to the rule.

This is another point where a realistic dispositional account of rule-following (chapter 2) is crucial for avoiding confusions. According to such accounts we can follow a rule, globally, without always following it, locally (due to error, inattention, *et cetera*). I think that these points show that the problems of bad company are pretty easily avoided if we limit our attention to our own language.[25] But the worries here extend to alternative languages.

However, we've already seen at length, in chapters 3, 4, and 5, that when the translation mistake is avoided, these worries too, are dissolved. Let me illustrate this in the present context with the case of the slur term "boche", which, like "tonk", has long been thought a problem for unrestricted inferentialism. Following Dummett, say that

“boche”’s meaning is given by the following introduction and elimination rules:

$$(bI)\quad \frac{\alpha \text{ is German}}{\alpha \text{ is boche}} \qquad\qquad (bE)\quad \frac{\alpha \text{ is boche}}{\alpha \text{ is cruel}}$$

With these rules, we can provide a rule-circular argument for the validity of the boche rules. This seems problematic, for not all Germans are cruel. Consider the following case:

> *Frau Frieda.* Frieda is a kindly, elderly German woman who spends most of her time helping starving children, playing with puppies, and volunteering at soup kitchens. She cries at the merest hint of violence and never has an unkind word or action for anybody or anything.

Since unrestricted inferentialists are committed to the validity of (*bI*) and (*bE*), it seems they must admit that Frieda, being German, is boche, and thus Frieda, being boche, is cruel. But Frieda is clearly not cruel. What has gone wrong? Again, the translation mistake has been committed. In a consistent language much like English except including the boche rules, the sentence “Frieda is cruel” really would be true, but the predicate “is cruel” in that language would be best translated into English as “is cruel or German”, rather than homophonically.

What about adding the boche rules to English? If we can already establish “Frieda is not cruel” then adding the boche rules results in inconsistency. So we would have to alter the English rules slightly if we added boche. Whether or not it engenders an inconsistency though, adding the boche rules to English changes the meaning of our word “cruel”. Nothing more mysterious than that is going on here. It is senseless to protest to the Bochers that Frau Frieda simply *is not cruel* (with a loud stomp of the foot). Frau Frieda surely isn’t cruel in *our* sense, but she just as surely is cruel in *their* sense. We can imagine further complications where the speakers of the extended language retain the attitudes that we ourselves have toward those that exhibit cruelty. This is problematic, since Frieda shouldn’t be

shunned or treated poorly simply because she's German. But note that this criticism of the speakers of the new language is *moral* and not *semantic*. So again, there is no problem for unrestricted inferentialism. A plausible account of translation frees the account of every blemish.[26]

If you were to add the boche rules to English (and avoid inconsistency), the boche rules would be valid, and you could demonstrate this with a rule-circular argument. Similarly, in a language that already includes the boche rules as basic rules, as long as the language has the required expressive resources, speakers of that language too could prove the boche rules valid and thus come to know that they are valid using rule-circular arguments. But this is fine, since the boche rules *are* valid in *that* language. And we can admit this without admitting that all Germans are cruel (they aren't) or that the boche rules are valid *in our language* (they aren't).

The situation surrounding boche should be very familiar after chapter 5's discussion, so I won't belabor it. Avoiding the translation mistake clears away all of the troubles when dealing with bad company. Although there are some differences of detail among the various cases, all putative bad company problems are handled smoothly and easily provided that we pay attention to differences between languages and take care not to make the translation mistake when translating between them.

Of course, in some alternative languages, like Tonklish, the conclusion of the rule-circular argument won't even state the validity of the tonk rules, despite superficial appearances. This is because the Tonkers can express no coherent information about the world in their trivial language (5.II). In most other cases though, like boche, the boche-speakers will be able to successfully establish the validity of their supposedly problematic rules just as we establish the validity of our unproblematic rules. This is not a problem provided that we make the requisite distinctions. Boche-speakers proving to themselves and gaining knowledge of the validity of the boche rules is *nothing to us*, since our language doesn't include rules for boche. Good for them, I say.

Stepping back: for the meaning-constituting rules, a rule-circular proof of validity will always be available, construed syntactically.

However, whether the conclusion of this proof in a given language actually expresses the validity of the relevant rule will vary from case-to-case. For valid but non-meaning-constituting rules a rule-circular argument for validity won't automatically be available, since we don't have a guaranteed entitlement to make epistemic gains with these rules across all contexts. A case-by-case examination of bad company problems would reveal various interesting and important example-specific details, but I think it is already clear from what has been said that my approach can handle all possible cases smoothly. In particular, our knowledge of the validity of all of the basic rules *of our language* can be given a rule-circular explanation, exactly as we desired. Here as elsewhere, unrestricted inferentialists maintain that for systematic and well-motivated reasons, bad company is no problem at all.

Unrestricted inferentialism, but not restricted inferentialism, allows for a streamlined account of our *a priori* knowledge of logical truth and validity via rule-circular proofs without the various complications and epicycles that Boghossian and others were forced to add to their own similar approaches in order to ward off bad company. This is why unrestricted inferentialism alone, and thus logical conventionalism, unlike other similar views, fulfills conventionalism's historical promise to smoothly integrate the *a priori* epistemology of logic into our overall scientific picture of the world.

THE INFERENTIALIST/CONVENTIONALIST THEORY of logic developed and extended in chapters 3 through 5 allows for a plausible, naturalistically acceptable, proof-based epistemology for logic. Only unrestricted inferentialism can be applied in this manner, since restricted inferentialist versions of the strategy falter on bad company. This is yet another theoretical virtue in the view's favor. Of course, the account of logical epistemology offered here provides less than some might have hoped. Some philosophers might have hoped that our logical epistemology would provide a tool to use in (*b*3) metaphysical disputes about the validity of our logical principles and laws, but I've already rejected the foundations of such disputes in 5.V. For myself and other conventionalists, a proper epistemology of logic doesn't provide an Archimedean point from which to move the logical world—the intuitionist, the relevantist, the dialetheist, and

the classical logician can all explain their own parochial logical knowledge along the lines suggested here. What then has been accomplished in accounting for our knowledge of validity? Mainly this: knowledge of logical laws like modus ponens has now been integrated into our standard epistemic practices. This means that we have a clear template for applying our standards of epistemic evaluation to them and mixing them with our non-logical epistemic evaluations, so that what could have been special and mysterious has been rendered standard and explicable. This is certainly not everything, but nor is it nothing. Yet even this much is chimerical if the notion of truth by convention itself is incoherent or otherwise unacceptable for general philosophical reasons.

---

1 For but one example, see BonJour (1998).

2 As noted in chapter 3, this approach to validity has sometimes been rejected; for example, Hartry Field says, “Our notion of good argument is an essentially normative notion, not capturable even extensionally in terms of truth-preservation” on page 268 of his (2009). One of the worries about the truth-preservation approach is the threat of paradox; for related discussion, see 12.IV.

3 In theories that can relatively interpret Robinson Arithmetic we don’t need to add any special devices for dealing with syntax but can instead use Gödel coding; Robinson Arithmetic was introduced in Robinson (1952) and popularized by Tarski, Mostowksi, and Robinson (1953); Gödel coding was first presented in Gödel (1931). Technical issues related to Gödel’s theorem will be discussed extensively throughout part III.

4 If greater generality is wanted, note that the arguments I give can be repeated at each level of the Tarski hierarchy; Tarski (1933*a*)—and please ignore the revenge problem that I’m flirting with in quantifying over the entire Tarski hierarchy in this footnote. See also 11.III and 12.IV. For some worries about proving soundness within the Tarski hierarchy, see the discussion in Field (2008). It is also worth noting that by working with schematic letters I am bypassing important issues concerning compositional theories of truth and the relative power of different axiomatizations.

5 At one point in section II this choice seems like it could make a substantive difference—see the discussion of Simple Sam.

6 For ease of reading, I will present the proof in a linear format rather than the two-dimensional form in which the rules were presented above. On the right of each line I have explicitly noted the assumptions (by line number) that the line

rests on along with annotation telling how the line was arrived at and, if appropriate, which premises were discharged in applying the relevant rule. This type of format for natural deduction proofs was introduced by Lemmon (1965) and Mates (1965).

7 In addition to (MP), this proof also uses the familiar →-introduction or conditional proof (CP) rule, as well as a standard ∧-elimination rule.

8 Tacit awareness is evinced when Haack (1976) and Wright (2001) point out that rules like affirming the consequent and the tonk rules can be used to give rule-circular justifications of themselves. The generality also holds for rules like (CP) that allow for premises to be discharged, though the statement of validity is a bit more complex in such cases. If there are devices for forming infinite conjunctions in our language, the triviality result can also be extended to rules with an infinite number of premises, such as the omega rule (which will be discussed in chapter 10).

9 Black (1954) gives a "self-supporting" (that is: rule-circular) argument for the goodness of inductive rules and argues that this type of selfsupport is not epistemically problematic; Dummett (1973*b*) argues that circular arguments of some kind have a role to play in *explaining* the role of deductively valid arguments in our language.

10 See Boghossian (1996), (2000), (2001), and (2003); however, Boghossian has since repudiated this approach, see his (2014).

11 In this context, it is safe to ignore *some* of the structure an argument has, for example, that the rules are applied in a particular order in a given instance of an argument.

12 See Wright (1985) and (2000).

13 This case is similar to Boghossian's (2003) "flurg" example.

14 Lewis (1996) denies this, but few have followed his lead.

15 This is one place where it might seem like I've relied upon the fact—which I claimed to be inessential—that my formulation of validity used a truth predicate in addition to schematic letters. But a structurally identical case could be constructed easily enough even if validity were formulated merely schematically and without a truth predicate, for example by stipulating that while Sam reasons according to (MP) he lacks belief in the complex but non-truth-involving schematic conditional that states (MP)'s validity.

16 See Burge (1993), Wright (2004), Boghossian (2003), and Peacocke (2004).

17 For some relevant discussion, see Malmgren (2019).

18 Supervaluations were first introduced in Mehlberg (1958) and named in Van Fraassen (1966).

19 There may be ways to explicitly suspend entitlement to a rule like (MP)—to a student, "try to prove this without using (MP)"—but in all contexts without any such explicit suspension, the MEC claims that we are epistemically entitled to use (MP) provided that it is a basic rule of our language. Alternatively, and more plausibly, the suspension here is merely *pragmatic* and the student

remains epistemically entitled to use (MP) even while operating under the pragmatic suspension.

20 From Boghossian (2003); quoted from page 279 of his (2008). See also the related discussion in his (2000).

21 The only worries about the MEC that seem to me to have any bite concern cases of reasoning under uncertainty, as in the paradox of the preface or the lottery paradox. Even if handling such cases required a restriction on the MEC it won't interfere with anything I am saying here. The arguments we are considering here are short and have no undischarged premises, so there is no possibility of small uncertainties summing to a significant uncertainty (in a valid argument the uncertainty of the conclusion cannot exceed the sum of the uncertainty of the premises—see Edgington (1997) for discussion). This isn't to admit that a restriction is required, just to say that, even if it is, it won't interfere with the argument of this chapter.

22 For example, Carnap (1950); see also Coffa (1986).

23 There are inductively strong rules in our language that we presumably earn the epistemic right to use in some less demanding sense. I won't discuss the epistemology of inductive rules here nor the way that non-automatic entitlements are earned more generally.

24 Boghossian himself abandoned the view because of Kripkenstein-style concerns about rule-following; see his (2014). But the complications he added to deal with bad company had already considerably diminished the simplicity and appeal of his theory.

25 This applies even to rules that we have entitlement to use in some contexts, even though they aren't meaning constituting or automatically entitling—see Dogramaci's (2010) statistical inference rule, for an example of such a rule.

26 See also Brandom (2000) and the discussion of conceptual rejection in Warren (2016*c*) and in 9.V.

7

The Very Idea of Truth by Convention

Along the path of developing and applying my conventionalist theory, I've already dealt with a number of important objections, including objections to the unrestricted inferentialism that underwrites conventionalism. But I have yet to say anything about some of the most important and influential objections to both logical conventionalism, in particular, and to truth by convention, in general. The time for that has now come—this chapter is devoted to demolishing these arguments.

Yes, *demolishing.* I think that all of these arguments can be conclusively answered and without any philosophical razzle dazzle or hand-waving. Maybe I'm wrong. Maybe I have understated the force of these arguments, many of which are *extremely* influential in contemporary philosophy. Maybe, but maybe not. At the very least, I hope that my discussion will make it clear that the case against conventionalism has been drastically overstated in contemporary philosophy.

I The Contingency of Conventions

Logical truths are necessary. This much is uncontroversial, but it provides grist for a mistaken but folkloric objection to logical conventionalism. Let's focus on the law of excluded middle (*LEM*) in order to concretely pose this objection as an argument against logical conventionalism:

1. The *LEM* is necessary
1. Our conventions are contingent, that is, our linguistic conventions, including those responsible for the truth of the *LEM*, could easily have been otherwise
1. *So*: The *LEM* could have been false (from 2 and the assumption of conventionalism)
1. *So*: The *LEM* is not necessary (from 3, by definition)
1. *So*: ⊥ (1, 4)

Call this the contingency objection to logical conventionalism. The contingency objection tries to generate an absurdity by combining the incontestable fact that our conventions are contingent with the generally accepted necessity of the logical truths. I have admitted both the necessity of logic (4.I) and the contingency of our conventions (chapter 5), so my view seems particularly vulnerable to this objection. However, the objection is deeply confused.

The central confusion in the contingency objection comes out in the move from (2) to (3): from the fact that our linguistic conventions, including those that we use to explain the truth of the *LEM*, could have been otherwise, it does not follow that the *LEM* could have been false, but only that in some possible languages a sentence of the syntactic form ⌜ $\phi \vee \neg\phi$ ⌝ is false. But rather than (3), this only establishes the possibility of languages where sentences of the syntactic form ⌜ $\phi \vee \neg\phi$ ⌝ are false, a matter which is entirely irrelevant to whether or not a sentence of this form could be false *in our language*. In fact, this argument is really just a disguised form of the trivial conventionality of linguistic signs that was explored at length in chapter 5.

A more sophisticated version of the argument might use semantic counterparts (5.III). This revised argument wouldn't rely on the trivial conventionality of linguistic signs, but would instead aim to show that there is a possible language where the *counterpart* of the *LEM* is false, not merely a syntactically identical item. But this conclusion is not established by (2) alone. To establish it we would need to show that there are alternative conventions of a very particular kind. Even if it were established, it wouldn't be a problem for conventionalism. An alternative language where a semantic counterpart of the *LEM* is

false does *nothing at all* to show that the *LEM* could have been false in *our* language. So the more sophisticated version of the argument fails for the same reason that the simple version failed—the contingency of our conventions is irrelevant to the modal status of the *LEM*.[1]

When properly understood, the contingency objection is no more serious than the claim that since the words "dog" and "galaxy" could be systematically swapped in an imagined language, it is possible for a dog to be many light years across. The standard ways to read these counterfactuals evaluates them with respect to our existing linguistic conventions, rather than with respect to the conventions of our counterparts in the given counterfactual. We could, of course, define such a reading of the counterfactual.[2] But these so-called counter*conventional* conditionals don't undercut the standard necessity of the logical truths. Similar confusions beset *temporal* variants of the contingency objection.

I won't bother spelling this point out further, since the discussion of the trivial conventionality of linguistic signs in chapter 5 should have already illuminated these issues (especially the discussion of the translation mistake in 5.I). Additionally, previous conventionalists have discussed and rebutted this and related objections in some detail (though not always in a manner that I would endorse).[3] The use/mention error at the heart of the contingency objection is not at all subtle, but it is related to the more subtle and alluring confusions that plague a much more popular argument against truth by convention.

II The Master Argument

The most influential—and perhaps most important—argument against conventionalism is the one I have elsewhere called the master argument against conventionalism.[4] The master argument is based on the idea that while conventions can determine which proposition or fact a given sentence expresses, they (conventions) are powerless to make propositions or facts true or false, so the very

possibility of truth by convention is foreclosed. The basic idea remains the same whether the argument talks of “propositions”, “facts”, or “states of affairs”.

The argument has been given over and over again in the philosophical literature. Here is Stephen Yablo giving an influential statement in a review of Alan Sidelle’s book on conventionalism about the necessary *a posteriori*, attributing the idea of Casimir Lewy:[5]

> Near the middle of his book *Meaning and Modality,* Casimir Lewy takes up the theory that “necessary propositions... ‘owe their truth to’ linguistic conventions.” All that conventions can do, he protests, is help determine what a sentence *says*, or what proposition it expresses; whether the proposition holds true is another question, to which the rules of usage are quite irrelevant. Such a view does not rule out conventional truth entirely, since our fiat might take the form: sentence *S* shall express a proposition that is (among other things) *true*. Far though from showing how *p*, the proposition expressed, could be conventionally true, this approach will be circular unless *p* has its truth value *independently* of the convention.[6]

Paul Boghossian makes similar points in answering an imagined proponent of “metaphysical” analyticity (the linguistic theory of necessary truth):

> How could the *mere* fact that S means that p make it the case that S is true? Doesn’t it also have to be the case that p?... he ... will want to say ... that, in some appropriate sense, our meaning p by S *makes it the case that* p. But this line is itself fraught with difficulty. For how can we make sense of the idea that something is made true by our meaning something by a sentence?[7]

Theodore Sider makes similar points more colorfully:

> There are a number of ways I can cause the proposition that my computer monitor has been thrown out the window to be true. I could throw the monitor out myself, pay or incite someone else to do it, and so on. I cannot, however, cause the proposition to be true simply by pronouncing. I can pronounce until I am blue in the face, and the computer will remain on my desk; my pronouncements do not affect the truth values of statements about computer monitors. Statements about conventions are different. These we, or at least our linguistic community, *can* make true by pronouncement. A convention consists of the activities of language users; that is why we can so easily make it the case that conventions exist ... Only statements *about pronouncements*, for example statements about conventions, seem to be made true by our pronouncements. Statements about monitors, or bachelors, or rain, are about a part of the world we cannot affect simply by pronouncement. That it is either raining or not raining is about rain; I cannot affect the world in the matter of rain simply by pronouncement; therefore I cannot make it the case that either it will rain or it will not rain simply by pronouncement.[8]

All of these authors are sketching versions of the master argument, but none of them develop the argument in detail. So it is worth going through things a bit more carefully.

According to the argument, the conventionalist claims that for any sentence *S* expressing a certain kind of truth (for example: logical, mathematical, *a priori*, necessary, *et cetera*), our linguistic conventions $\mathcal{C}$ *make it the case that S* is true. But *S* is true just in case *S* expresses proposition *p* and *p*, and our conventions only have the power to determine *which* proposition *S* expresses; they have no power to make it the case that *p*. So conventionalism is false. The structure of this argument is basically the same if framed in terms of “facts”, but for now I will stick with the “proposition” version, since it seems favored by most proponents of the argument.

We can make things even more explicit by being a bit more formal. Assume that *S* is a sentence that is not itself about our linguistic conventions. Then the argument runs as follows:[9]

1. □ (sentence *S* is true ↔ ∃*p*(*p* is a proposition ∧ *S* means that *p* ∧ *p*))
2. ∀*p*¬linguistic conventions make it the case that *p*
3. *So*: ¬linguistic conventions make it the case that sentence *S* is true (1, 2)

The conclusion is close to a direct denial of conventionalism, so the argument seems to show that the very idea of truth by convention is ill conceived.

The first two premises are supposed to be obvious. Premise (1) is nearly definitional, at least when talk of "propositions" is accepted. Clearly a sentence like "ravens are black" is true just in case it says something and what it says is the case. Premise (1) merely applies this basic approach to all sentences, which seems unobjectionable (Boghossian even calls his version of this premise a truism).

There are a number of things to be said in favor of premise (2). Boghossian and Sider take it as obvious and suggest that it's hard to see how it could possibly be false, while Yablo seems to think that denying it leads to vicious circularity, at least in the context where we have stipulated that a sentence is to express a truth. Boghossian also sketches an argument for premise (2) elsewhere in his discussion: if our conventions made *p* the case, then it wouldn't have been that *p* before those conventions were in place. But surely it was the case 2 + 2 = 4 (for example) long before humans, languages, or linguistic conventions existed. This sounds a lot like the contingency objection I criticized in section I, but instead of involving sentential truth, it involves the notion of a proposition *being the case*. Because of this, if a use/mention distinction is to avoid the objection, it must be done more subtly. Despite the points made in its favor, I don't think that premise (2) is unassailable. Its status will depend both on what exactly is meant by "proposition" in the argument, and on what is meant by "makes it the case that".

Taking the last point first, I think there are many ways of understanding the "makes it the case that" locution used in the master argument. But as far as I can tell, the reading favored by proponents of the argument is different than the reading favored by conventionalists. This isn't a point about the *words*—instead of

"makes it the case that" some other expression could easily have been used. Whichever expression we fix on will have a similar range of readings. Yablo, Boghossian, and Sider seem to read the expression causally. Since they think that our conventions cannot *cause* it to be the case that *p*, they think that premise (2) is true. But as I discussed in 1.II, the correct conventionalist thesis is *explanatory*. If the conclusion of the argument is truly to conflict with conventionalism, "makes it the case that" in premise (2) must be given an explanatory reading.

So understood, there isn't any obvious reason to think that (2) is true. Whether it is true or not will depend upon our favored theory of *propositions* (or *facts* or *states of affairs* for variant forms of the argument) and how it interacts with our account of *explanation*. This brings us to the second point of clarification about premise (2): what exactly is meant by "proposition"? Historical conventionalists very rarely went in for "proposition" talk in any form. But I think that conventionalists should say about "propositions" what I said about "facts" in 4.V—on metaphysically lightweight readings, of course there are propositions, but on metaphysically heavyweight readings, of course there aren't propositions.

Obviously there isn't just one "lightweight" way of understanding "proposition" or "fact" talk. If our account of propositions ties them closely enough to linguistic expressions, then premise (2) becomes implausible. And Boghossian's argument for (2) can then be answered in the same way I answered the contingency objection in section I. So conventionalist proponents of language-like "propositions" or "facts" can deny (2).[10] But I personally don't think a language-like approach is the most natural or appealing picture of lightweight propositions.

Understood using a more natural notion of lightweight propositions, premise (2) might still be true. This is brought out in Boghossian's discussion of whether conventionalists can evade the issue by appealing to a deflationary conception of truth:

> Even on a deflationary view of truth, there is presumably a distinction between the *sentence* 'Snow is white' and that which makes the sentence true, namely, snow's being white. And the essential point for my purposes is that it is one thing to say that 'Snow is white' comes to express the claim that snow is white as a result of being conventionally assigned the truth value True; and quite another to say that snow comes to be white as a result of our conventions.[11]

So understood, premise (2) may well be true. But I don't think this sinks conventionalism, since when (2) and (3) are understood properly, and read as expressing explanatory claims, the master argument is invalid. The basic problem is simple: as noted in 1.III, explanatory contexts are hyperintensional. Yet the explanatory version of the master argument implicitly denies this by moving from premises to conclusion using something like the following principle:

1. $\Box(\phi \leftrightarrow \psi)$
2. Δ makes it the case that ϕ
3. *So*: Δ makes it the case that ψ

I have simplified a bit to make the point clear.[12] In essence though, the master argument assumes that explanatory "makes it the case that" contexts are merely intensional. They are not. They are hyperintensional, so we have been given no reason to accept the argument's conclusion even when both premises are accepted.

Given this, conventionalists can coherently accept both premises of the master argument while at the same time rejecting the conclusion. For lightweight propositions like we've been considering I don't think that this approach is ad hoc or in tension with conventionalism. Our conventions explain the truth of sentences like "it will rain or it will not", and—I will argue in part III—they also explain the truth of sentences like "2 + 2 = 4". But if someone were to ask why it is that 2 + 2 = 4, they would plausibly be looking for a *mathematical* answer. Similarly if someone were to ask why it was that either it will rain or it will not, we would probably answer by appealing to a general version of the law of excluded middle, in the

object language. Conventionalists can coherently explain lightweight propositions using other lightweight propositions, if they are so inclined.

Some may worry that "explaining" lightweight propositions (or facts) with other lightweight propositions (or facts) undercuts conventionalism.[13] But I don't think this is right. The crucial epistemological role of conventions remains, when we talk in the material mode, using *our* language (chapter 6). Obviously conventionalists can't coherently think that we need some non-conventionalist theory of logic or mathematics to pick up the slack. But these admissions leave no gap for a non-conventionalist philosophical theory to fill. Our explanatory practices are varied, but conventionalists can leave nothing unexplained without denying any truisms.

What is happening in this type of case is not an appeal to an alternative philosophical theory, but rather to other logical or mathematical claims, themselves, in the material mode, in the object language, unadorned by philosophical commentary. According to this approach, conventions explain the truth of sentences that are true by convention, and the propositions expressed by said sentences are explained by distinct propositions expressed by other such sentences. There are various ways that our explanatory practices can handle this kind of level-jumping without sinning against the basic tenets of naturalism (1.IV). These explanations are object-level reflections of the explanatory relationships that hold between our linguistic conventions, themselves expressible directly via linguistic ascent. Yet further shadows of syntax, as it were.

Maybe these "explanations" aren't very deep, but I don't think that is important. Conventionalists can plausibly deny that there is any deeper sense in which it can be explained that it will rain or it will not. There is no gap left here for some anti-conventionalist philosophical theory to fill. Many instances of the law of excluded middle are obviously *about* the world, but that doesn't mean that they communicate any actual information about the world. Such propositions don't specify a way that the world must be. They are utterly empty.[14]

Another reading of the “makes it the case that” phrase concerns metaphysical truthmaking relations between facts and sentences or propositions. I don’t think this is what historical conventionalists had in mind, for a related reason. The fact that it is now raining can plausibly be said to “make it true that” it is either raining or it is not, but that doesn’t undermine conventionalism. It is simply a necessary connection of the kind discussed in 4.VI. I am happy with this way of talking in some informal contexts, but I am not too tempted to say on this basis that the fact that it is raining *fully explains* the truth of “it is raining or it is not”. The conventions still must play some role in this “explanation”, and they can account for the truth of the sentence all on their own. And even if truthmaker explanations were counted as fully explanatory for some particular logical truths, metaphysical extravagences like negative facts would likely be required to cover all such truths (in a less metaphysically heavyweight sense, I have no problem at all with talk of “making true”, see 11.I). Even if successful, this approach still leaves mathematical truths and *general* logical truths like the *LEM*, as well as facts about logical truth and necessity, completely untouched. Perhaps truthmaker theorists would say that the world as a whole trivially makes logical laws and mathematical claims true. So understood, it is not a philosophical competitor to conventionalist explanations. With trivial truthmaking, there is no way the world must be in order for these sentences to be true (the discussion of “heavyweight” facts below is also relevant here, perhaps).

So far we have seen two plausible conventionalist responses to the master argument. The first rejects premise (2) because lightweight “propositions” or “facts” are closely tied to language, and so are plausibly explained by linguistic conventions. The second accepts premise (2) for lightweight propositions (or facts) but rejects the conclusion, pointing out that the argument is invalid, since explanation is hyperintensional. This second option admits that while we might sometimes “explain” logically true propositions with other trivialities, there is no deeper sense of “explanation” that applies to them. This is a practice-internal sense of explanation, mentioned in 1.IV. An analogy could be drawn to fictional truths or other branches of discourse where cross-level explanations sound a bit odd.

These are responses to the argument when read as concerning a lightweight notion of "proposition" or "fact". On "heavyweight" readings of the master argument, the conventionalist response is much simpler. On this reading, *there are no logical facts*. So on the robust reading, our conventions do not make it the case that it is a fact that it is either raining or it is not, but nothing else does either. There are no such facts. So understood, both premises of the argument are false or senseless. Of course, saying this doesn't require treating all necessarily true claims with the same brush. Perhaps the necessary truth that "water is H_2O" states a fact that requires an explanation, and similarly for a supposed necessary truth like "there are infinitely many concrete spatiotemporally isolated worlds" or "there is a necessary being who created reality". So far, conventionalists are only committed to rejecting heavyweight propositions or facts for *logical* truths. In part III I will extend conventionalism to mathematics, but here I make no claim about all putative necessary truths. Conventional necessity is trivial, but perhaps not all necessity is conventional.[15]

I think I've said enough to conclusively show that the master argument fails to undermine conventionalism. But perhaps there is a way to patch the argument with some slight alterations. One patch is to simply stipulate as a general metasemantic principle that sentential truth must be explained in terms of propositions. But obviously this type of reformulation is no better than simply gainsaying conventionalism from the start.[16]

Another attempted patch replaces premise (1) of the original argument with a stronger equivalence claim that guarantees full and complete hyperintensional equivalence.[17] As a representative example of this move consider the following modification of the master argument:

1. For it to be the case that sentence *S* is true just is for it to be the case that $\exists p$(*p* is a proposition $\wedge$ *S* means that *p* $\wedge$ *p*)
2. $\forall p \neg$linguistic conventions $\mathcal{C}$ make it the case that *p*
3. *So*: $\neg$linguistic conventions $\mathcal{C}$ make it the case that sentence *S* is true (1, 2)

The context ⌜ for it to be the case that ϕ just is for it to be the case that ψ ⌝ used here is simply stipulated to, if true, guarantee full hyperintensional equivalence. This means that if premise (1) of the modified master argument is true, then the problems with hyperintensionality are avoided and the argument as a whole is valid. If right, this limits the conventionalist's options.

While it seems a promising variation at first glance, the idea that the new version of premise (1) guarantees full hyperintensional equivalence is deeply problematic, even leaving aside 2.II's worries about stipulations in metasemantics. Whenever two embedded sentences involve different *concepts*, they will come apart in at least *some* hyperintensional contexts, such as belief reports or contexts involving quotation. So either sentences the form of (1) are always false when the embedded sentences syntactically differ, in which case premise (1) is false and the modified argument is unsound, or sentences of the form of (1) can be true without guaranteeing full hyperintensional equivalence. In this latter case, it is at best an open question whether premise (1) is true in the relevant cases. And what is worse, even if it is true, we no longer have any reason to think that a true "just is" equivalence preserves explanatory goodness. So either "just is" equivalence is too demanding for premise (1) to be true, or too undemanding to give us a reason for thinking the argument valid.

My response to both the original and the modified master argument leans heavily on the hyperintensionality of explanation. As discussed in 1.III, the hyperintensionality of explanations isn't controversial. As also noted briefly in 1.III, I think it is plausible that explanatory contexts, like belief reports, quotation, and other hyperintensional contexts, are sensitive to the syntax used to express the relevant facts. This wouldn't amount to making our notion of explanation entirely anti-realist or non-factual, it would just mean that explanations are sensitive to the guises in which facts are expressed. I think this is quite intuitive. Why the quintic is insoluble by radicals is a mathematical matter, best explained by Galois theory.[18] But why a mathematical sentence, expressing this fact, is true is—I will argue in part III—a matter of convention. The intuitive force of explanatory claims often depends more on the

communicative reading of “explanation” than the metaphysical reading (1.III). All told, I think that the conventionalist offers a philosophical explanation of the logical and mathematical facts, without being forced to deny any evident truths or say anything intuitively controversial. What exactly should and shouldn’t be said depends on which strand in our explanatory practices is picked up. The logical and mathematical facts that are relevant to our practices are fully explained by our conventions. Beyond this, there is nothing left unexplained, and so no explanatory task is left to a non-conventionalist theory.

III Quine, 1936

The master argument attacks the *very idea* of truth by convention, but the most famous anti-conventionalist argument aimed directly at *logical* conventionalism comes from “Truth by Convention”, an extremely influential 1936 paper by W.V. Quine.[19] To this day, many prominent philosophers think Quine’s arguments are decisive. Recently Theodore Sider advised those with a “hangover from logical conventionalism” to:

> …pop a couple of aspirins, re-read your Quine (1936) … and report back.[20]

Despite being one of the most lauded critical pieces in contemporary philosophy, detailed examination shows that Quine’s central argument is only a problem for an implausible version of conventionalism very different from the kind endorsed in this book. I have already examined Quine’s argument in detail in print, but an analysis must be given here as well.[21]

In “Truth by Convention”, Quine takes linguistic conventions to be explicit in basically the manner I discussed and rejected in 2.II:

> ***Explicitness***: A linguistic convention concerning a sentence takes the form of an explicit stipulation concerning the sentence.

Quine then argues against any version of logical conventionalism that accepts ***Explicitness*** in two stages.[22] First, he argues that according to explicit conventionalism, for each logical truth ϕ, we—as a linguistic community—have stipulated that ϕ is to express a truth. But since there are an infinite number of logical truths expressible in our language, there is no sense in which any human being or finite collection of human beings could have made this infinite number of explicit stipulations. Let's lay this out more carefully as Quine's super-task argument:

1. If explicit conventionalism is true, then for each logical truth in our language ϕ, we have explicitly stipulated that ϕ is to express a truth (from the definition of explicit conventionalism)
2. There are an infinite number of logical truths in our language
3. *So*: If explicit conventionalism is true, then we have made an infinite number of explicit stipulations (1,2)
4. We have not made an infinite number of explicit stipulations (given the finite number of humans and the nature of explicit stipulations)
5. *So*: Explicit conventionalism is not true (3,4)

We might try to challenge (4). A task whose completion requires completion of infinitely many sub-tasks is called a *super-task*.[23] And some super-tasks, such as the Zeno-ian super-task of walking across a room, can be completed in a finite time. But in this case, each explicit stipulation requires some minimum finite amount of time to complete, and so the super-task as a whole would also take an infinite amount of time to complete.

Given this, the only way for an explicit conventionalist to resist the argument is to amend their view so that stipulations concern only finitely many general claims, from which the infinite number of particular logical truths can then be derived (this involves rejecting premise (1) of Quine's super-task argument). For example: we stipulate that any sentence of the form $\ulcorner \phi \vee \neg\phi \urcorner$ is to be true, and from this we derive that the infinitely many particular instances of this schema are true. If this approach is applied to other logical truths as well, it solves the infinitude problem, but only by appealing to

background logical principles that allow us to derive particular instances from our general claims. In other words, this procedure explains logical truth in terms of explicit conventions *together with background logical notions*.

Let's illustrate this with an example. Imagine we have made the following general stipulation:

$$(meta\ LEM) \quad \text{Any sentence of the form } \ulcorner \phi \vee \neg \phi \urcorner \text{ is true}$$

And we want to derive that the particular sentence "$p \vee \neg p$" is true. To infer this, we need the premise (*i*) "$p \vee \neg p$" is of form $\ulcorner\ \phi \vee \neg \phi\ \urcorner$, and this is fine, since it expresses an uncontested matter of syntactic fact. But this is not enough, for we also need (*ii*) if (*meta LEM*) and (*i*) are both true, then "$p \vee \neg p$" is true. But now this premise, (*ii*), also needs to be added, and we are moved to (*iii*) if (*meta LEM*) and (*i*) and (*ii*) are true, then "$p \vee \neg p$" is true. At this point, we are off on a regress—every premise we add will just be more grist for the mill to go into our ever-expanding stock of premises. The regress could be avoided by noting that these background principles were already true *prior* to our stipulations, but that sins against explicit conventionalism. So even the "general stipulation" version of explicit conventionalism is forced to claim that an infinite number of stipulations have been made in order for any one of the proposed derivations to go through. It appears that the general form of explicit conventionalism has done no better than the particular form we started with.[24]

Let's also lay this argument out explicitly as Quine's regress argument:

1. If explicit conventionalism is true, then we have stipulated the validity of some finite number of general logical principles (this is the proposal under consideration)
2. There are an infinite number of logical truths in our language (uncontested fact)
3. *So*: If explicit conventionalism is true, then we have to derive an infinite number of logical truths from our finite number of stipulated principles (from 1,2, and the conventionalist proposal)
4. This derivation requires that either an infinite number of explicit stipulations have been made or that we appeal to some logical principles that are true but were not explicitly stipulated to be true (this was just argued for using the Carroll-style regress)
5. *So*: Explicit conventionalism is false (from 3,4, the super-task argument, and our definitions)

Explicit conventionalists have a choice: appeal either to explicit stipulations of particular sentences or to explicit stipulations of general principles. If they choose the first, Quine's super-task argument refutes them; if they choose the second, Quine's regress argument refutes them. Either way, explicit conventionalism stands refuted. Let's call this combination argument Quine's master argument.

I think it's plausible that Quine's master argument refutes explicit logical conventionalism. But linguistic conventions aren't explicit stipulations (2.II), they are implicit rules (2.III). With this move from the explicit to the implicit, Quine's master argument is powerless. This has been recognized in the literature before. Here is Boghossian clearly stating the central point:

> This argument of Quine's has been very influential; and I think that there is no doubt that it works against its target as specified. However, it is arguable that its target as specified isn't the view that needs defeating. For, surely, it isn't compulsory to think of someone's following a rule R with respect to an expression e as consisting in his *explicitly stating* that rule in so many words in the way that Quine's argument presupposes. On the contrary, it seems far more plausible to construe x's following rule R with respect to e as consisting in some sort of fact about x's *behavior* with e.[25]

Sider and other contemporary fans of Quine's "Truth by Convention" seem to have missed this point. But Quine himself did not. He clearly recognized that his arguments only worked against explicit versions of conventionalism. He had a very different complaint for implicit conventionalism:

> In dropping the attributes of deliberateness and explicitness from the notion of linguistic convention we risk depriving the latter of any explanatory force and reducing it to an idle label. We may wonder what one adds to the bare statement that the truths of logic and mathematics are *a priori*, or to the still barer behavioristic statement that they are firmly accepted, when he characterizes them as true by convention in such a sense.[26]

Despite Sider's dismissive rhetoric, Quine's *entire case* against plausible versions of logical conventionalism rests on the strength of this point. A widespread failure to understand this key point has led to a drastic overestimation of the relevance of Quine's arguments for the prospects of logical conventionalism.

And Quine's complaint fails. This is already shown by the discussion of conventions in chapter 2. Saying that a sentence is provable from our rules alone is very different than saying it is firmly accepted. A sentence like "it has rained in the past" is firmly accepted if anything is, but is not provable on the basis of our rules alone. The case for positing inference rules from 2.III already heads off this simple version of Quine's worry.

In later work Quine discussed linguistic rules more in more detail. There he distinguished between rules *fitting* and *guiding* our behavior, and argued that for a rule to "guide" our behavior, rule-followers must be able to explicitly state the rule. In order to uphold this, Quine was forced to argue against the kinds of linguistic rules posited by post-Chomsky cognitive science. I have already addressed Quine's discussion of this point directly in 2.IX, and indirectly in 4.VII and 4.VIII and throughout chapter 6, so I won't rehearse things here. It suffices to say that there is a perfectly good sense in which implicit rules of inference can guide our behavior, even if most of us are unable to explicitly formulate these rules on command.

Quine's worries about linguistic rules typically took the form of underdetermination arguments. If two distinct collections of rules lead to the very same valid arguments, then how can we say that we follow one collection rather than the other? But this worry, too, has already been answered. According to the inferentialist theory developed in chapter 3, only the basic rules of a language are meaning constituting. And the distinction between these basic rules and derivative rules can be drawn in terms that Quine himself should accept. We follow the basic rules directly, in terms of simple dispositions linking premises and conclusion. But we follow the derivative rules only indirectly, in terms of complex or merely composite dispositions linking the premises and the conclusion. Even arch-behaviorist Quine should accept this distinction, by his own methodological lights. We can distinguish the basic and derivative rules from each other on general and largely behavioral grounds (and thus so can Quine and his modern-day followers).

IV Variations on a Theme by Quine

The previous section addressed Quine's anti-conventionalist arguments as they occur in "Truth by Convention". But my impression is that many philosophers think that some variant of Quine's argument against explicit conventionalism is the real problem for implicit conventionalism. There are a number of ways of

extending this inchoate idea into a full-blown objection to logical conventionalism, but I think they all exhibit one or two central failings.

Before establishing this we need to get clearer on the exact nature of the concern. In Michael Dummett's writings on Wittgenstein's later work on the foundations of mathematics we find a powerful statement of the kind of challenge that I have in mind:

> The conventionalism that is so widespread is, however, a modified conventionalism. On this view, although all necessity derives from linguistic conventions that we have adopted, the derivation is not always direct. Some necessary statements are straightforwardly registers of conventions we have laid down; others are more or less remote *consequences* of conventions. … This account is entirely superficial and throws away all the advantages of conventionalism, since it leaves unexplained the status of the assertion that certain conventions have certain consequences.[27]

When applied to the case at hand, Dummett challenges the conventionalist to provide some satisfying account of how some of our logically valid rules can be *consequences* of our directly adopted conventions. This challenge is much like Quine's except that it doesn't obviously rely on treating conventions as explicit stipulations or declarations. At least at first glance, this challenge seems to face the implicit conventionalist just as surely as it does the explicit conventionalist.

Some authors have interpreted Dummett's challenge as concerning our *recognition* of the necessity of the conventionally given necessary truths.[28] But I think this is a mistake. The challenge is better seen as concerning the metaphysical relationship, understood broadly, between the direct registers of our conventions and their consequences. This doesn't directly concern our recognition of the necessity of the conventional truths. In fact, in 4.I I sketched an account of the necessity of conventional truths that can be extended to explain our recognition of these truths (and of their necessity) without any significant issues by adapting the "prove it"

epistemology of chapter 6. So I don't think that the necessity of conventional truths raises psychological or epistemological puzzles, as such.

Dummett's challenge, properly understood, threatens to trap the conventionalist between two unpalatable alternatives in trying to explain the relationship between our conventions and their "consequences". On the one hand, it seems to sin against the spirit of conventionalism to take the relationship to itself be a non-conventionally explained necessary connection, but on the other, it seems problematically circular to take the relationship to be a matter of convention. Before addressing this dilemma directly, I want to say something about Wittgenstein's characteristically extreme response.

The later Wittgenstein seemingly accepted the dilemma. His solution was to entirely bypass the need for a distinction between our direct conventions and their indirect consequences by adopting a thoroughgoing and extreme brand of conventionalism. This move is what gives rise to the late Wittgenstein's notorious account of mathematical proof where each step along the way is, in effect, a matter of *choice*. In essence, on Wittgenstein's view, we are not *compelled* by the steps in a mathematical argument to proceed on to the conclusion.[29] This position certainly avoids the need for any non-conventionally explained necessary connections and it does so without any problematic circularity, but most philosophers have found Wittgenstein's account unacceptably radical, for a reason succinctly voiced by Dummett:

> This account is very difficult to accept, since it appears that the mathematical proof drives us along willy-nilly until we arrive at the theorem.[30]

There are things that Wittgenstein and others have said that attempt to draw the sting from this criticism. Perhaps the phenomenology of feeling compelled can be accounted for even when we are, in fact, making a free choice.[31] But even if true, I don't think this manages to make Wittgenstein's position acceptable. Fortunately though, we don't need to adopt Wittgenstein's radical version of conventionalism in order to meet Dummett's challenge.

To make sure the challenge for my theory is posed in the strongest possible terms, note that my brand of conventionalism seems to play right into Dummett's hands by making a distinction between the basic, meaning-constituting rules of our language—like (*&I*) and (*&E*)—and the derivative but still valid rules of our language—like (&&) from 3.II. Much like Dummett's "modified" conventionalist, I seem to be maintaining that our conventions, in the form of basic rules, have certain "consequences", in the form of derivative rules. Wittgenstein and Dummett aren't the only philosophers who worried about this issue. Something like it has often been used as an objection to logical conventionalism.[32]

This challenge is beguiling, but talk of "consequences" here is highly misleading. It suggests an account whereby the metaphysical point of note is that the derivative rules are *explicitly derived from* the basic rules. This would make these proofs themselves essential to the conventionalist's explanation of the validity of the derivative rules. The same picture is lurking, even more directly, behind Patricia Blanchette's similar challenge for any meaning-based explanation of logical truth:

> To say that the meanings of a collection of terms 'suffices for' or 'guarantees' the truth of a sentence seems to mean little more that that the sentence's truth *follows logically from* facts about those meanings … And if this is right, then one cannot, without vicious circularity, give a characterization of the logical properties and relations in terms of meanings. … [T]he circularity of the proposed analysis, would seem likely to pose difficulties for virtually any attempted analysis of … logical properties and relations … [T]o say that certain features *F* of a sentence or claim *make* that claim a logical truth is to say something dangerously close to saying that the sentence's having *F entails* that the sentence is a logical truth.[33]

Let's leave aside for the moment the question of whether this really would be viciously circular.[34] The same basic challenge is also sometimes framed in terms of "consistency" rather than "following

from" or "consequence". Here is Frederick Copleston putting this spin on the challenge during a BBC debate with A.J. Ayer from 1949:

> …does it not seem that there is at least one proto-proposition which governs all possible systems of logic? … [C]onsistency, or observation of law, within a game, whatever these laws may be, is itself, it seems to me, a kind of proto-principle. … [I]t does seem to me that there is, at any rate, a principle of consistency, which seems to me to be a kind of proto-proposition governing all reasoning.[35]

But reading Copleston's "consistency" here as *logical* consistency is misleading, just as reading Dummett and Blanchette's claims of "forcing" or "making" as claims about *logical* entailment is deeply misleading.[36] So misleading, in fact, that it amounts to a distortion of the conventionalist picture.

The conventionalist is not claiming that derivative rules need to be explicitly derived from basic rules, at least not exactly. Concerning derivative rules like (&&), the essential conventionalist claim is not that (&&) is valid because it logically follows from the validity of our basic rules that (&&) is valid. But rather that (&&) is valid because it is a rule of use of our language that must be valid according to our inferentialist metasemantics. This is related to a point noted in 3.I's discussion of the Totality principle, but let's go over it in more detail.

In 3.II I provided a proof of the derivative rule (&&) using the basic rules (*&I*) and (*&E*) together with the structural rule of transitivity, but the status of this proof is easy to misunderstand. The existence of this proof *shows* that anyone who follows the standard structural rules, together with the (*&I*) and (*&E*) rules for "and", also, thereby, follows the rule of (&&). Derivative rules *follow from* or are *consequences of* the basic rules in the sense that anyone who follows the basic rules also, thereby, follows the derivative rules. To directly follow the basic rules for "and" involves indirectly following (&&). The same dispositions underwrite both things. This is a point about *constitution*, not logical entailment. More particularly, it is a point about what is involved in following a linguistic rule of inference, not a point about the status of logic or other logical principles. It is

true that we demonstrate the validity of derivative rules to ourselves by formally proving them, as when we formally prove a rule like (&&) in a system with the rules (*&I*), (*&E*), and transitivity. But our construction of this proof isn't playing any more immediate role in the conventionalist explanation of the validity of (&&) or any other derivative rule. Rather, it is *that* the rule can be derived from basic rules that is important.

Similarly, that you must reason in a fashion globally "consistent" with the law of noncontradiction in order to mean what we mean by "not" is not a "proto" or ur-principle of contradiction that governs all languages. It is instead a point about constitution. Using "not" in such-and-such ways is what constitutes meaning what we mean by "not". You are free to use "not" in completely different ways, just as you are free to move your wooden horse in any way you like. It is just that only certain ways of moving the horse count as playing chess, and only certain ways of using "not" count as meaning negation. If your use of a symbol isn't "consistent" with the rules for the symbol, in a broader sense, then you will have changed the symbol's meaning.

These points conclusively answer crude versions of Dummett's criticism, but there are other, perhaps more subtle, variations on Quinean themes in the general vicinity. For example, the Dummett passage quoted earlier in this section goes on to formulate the problem in a more normative fashion, somewhat suggestive of Kripkenstein's normativity problem for dispositionalism about rule-following (2.VII):

> It appears that if we adopt the conventions registered by the axioms, together with those registered by the principles of inference, then we *must* adhere to the way of talking embodied in the theorem; and *this* necessity must be one imposed upon us, one that we meet with. It cannot itself express the adoption of a convention; the account leaves no room for any further such convention.[37]

Despite the change in focus, I think my discussion already goes a long way toward answering this challenge. It isn't that you *must*, in

some external normative sense, follow (&&) if you follow (*&I*), (*&E*), and transitivity; it is rather that if you *do not* follow (&&), then you *are not* following at least one of (*&I*), (*&E*), and transitivity. You must follow (&&) if you follow the other rules, because following the other rules suffices for following (&&). To again use Tractarian language, the proof from 3.II doesn't establish this, *it shows it to be the case*. You can stop following (&&) if you like, but that will involve ceasing to follow one of (*&I*), (*&E*), and transitivity, since directly following these rules suffices for indirectly following (&&). This is no more mysterious than the fact that you can't fly in a red airplane without flying in an airplane.

Taking everything together, the explanation of the validity, in our language, of a derivative rule like (&&) is simply that we follow the rule (&&) by way of following various other rules—including (*&I*), (*&E*), and transitivity—and by the MVC and MAC these basic rules are valid and thus so too is (&&). In this way, the validity of the basic rules accounts for the validity of (&&) as well. Of course, the explaining facts here, suitably supplemented, logically imply the validity of (&&) (the fact to be explained), but that they do so is not itself any part of the explanation of (&&)'s validity. This much is already secured by my discussion of the role of entailment claims in explanations in 4.II. There I noted that even if it is a necessary condition on *E*'s explaining *p* that *E* (suitably supplemented) logically entails *p*, it doesn't mean that the fact that *E* logically entails *p* is itself part of the explanation of *p*. No logical (or metalogical) claim of this kind is part of my conventionalist explanation of the validity of derived rules like (&&).

But *even if* these entailment claims were included in our explanations of validity, by itself that wouldn't automatically undercut the conventionalist credentials of the explanation.[38] For conventionalists, these entailment claims are themselves given a conventionalist explanation, and so no deeply non-conventional element needs to seep into our explanations, even when including logical entailment claims among the explaining facts. So while I think the use of entailment claims in explaining the validity of derivative rules can and should be avoided, I don't think the avoided circularity

itself would be as obviously problematic as some anti-conventionalists claim.

It is also important to realize that even if a conventionalist accepts the circularity, that doesn't commit them to the radical Wittgensteinian position sketched above. To think otherwise is to confuse different senses of "conventional". As I've been using the term, something is conventional if it is explained with only substantive appeal to our linguistic conventions (basic rules of our language), but this doesn't mean that what is conventional is somehow *arbitrary* or *subjective* or metaphorically *soft* (throughout chapter 5 the distinction between the conventional and the truly arbitrary was drawn at great length).

Getting clear on this last point is essential when considering one last variation on Quinean themes. What is the status of our conventions themselves and the features of those conventions, according to the conventionalist? For instance, some collections of linguistic rules—like many containing the tonk rules—are trivial, while others are not. But what is the status of this distinction itself? It seems problematic to say that certain tonk languages are trivial *by convention*, but denying this seems to undercut the conventionalist credentials of the theory I've offered, or so our imagined objector would argue. Clearly it is not subjective or non-factual that a given trivial language is trivial. But this isn't obviously the same thing as saying it isn't a matter of convention. In the relevant sense, syntax is a matter of fact.

Let me try to explain this in a bit more detail. Given our linguistic conventions, "trivial" and "non-trivial" pick out features that any given language may have or lack as a matter of objective fact. There is nothing problematic about this claim. Conventionalists can and should be, in a manner of speaking, *factualists* about our linguistic conventions. As an analogy, let's consider our game of chess together with similar games with variant rules, for example, a version of chess without the castling rule, a version without the en passant rule, *et cetera*. Presumably it is a fact that some chess-like games are a theoretical win for white while some chess-like games are a theoretical draw. What explains the fact that chess-like games fall into these two categories? In this case, it seems patently obvious

that different chess-like games have their features objectively against the fixed background of what we mean by “win” and “draw”. And I would say the very same thing about the fact that languages can be either trivial or non-trivial.[39] Otherwise, that a given language *L* is trivial, for example, would itself have to be provable using the rules of our language alone, and this surely isn’t the case—our rules of language do not allow us to conclude that “English is non-trivial” without any undischarged premises, for example.

In short: syntax is a matter of fact, not convention, and saying this in no way undercuts logical conventionalism.[40] There are technical challenges to this claim, and technical versions of the arguments of this section, but they will be discussed in part III, in light of the arithmetization of syntax (see especially 12.II, 12.III, and 12.IV). For now though, this concludes my discussion of variations on Quinean themes. So far as I can see, there is no way to use some such variation to make serious trouble for my brand of logical conventionalism. The feeling to the contrary stems from either a badly mistaken view of the relationship between basic and derivative rules, a harmful confusion of the conventional with the entirely arbitrary, or an implausibly negative view of a relatively benign form of circularity.

V Williamson and Expert Disagreement

Timothy Williamson has influentially argued against notions of analyticity, conceptual truth and even the *a priori*/*a posteriori* distinction itself.[41] Some of the arguments take direct aim at key inferentialist assumptions, so they warrant some discussion here.

His central argument is based on the idea that experts can disagree about any basic logical principle while retaining full understanding of the logical notions involved:

> It is often claimed that assent to arguments by modus ponens ... is a precondition for understanding the word "if." … Indeed, this is a standard example in the literature. However, Vann McGee, a distinguished logician, has published purported counterexamples to modus ponens for the indicative conditional in English. … Does McGee not understand the English word "if"? In conversation, he appears to understand it perfectly well. By ordinary standard, he *does* understand it. Before he had theoretical doubts about modus ponens, he understood the word "if" if anyone has ever understood it. Surely his theoretical doubts did not make him cease to remember what it means.[42]

Against the natural response of appealing to the division of linguistic labor, where speakers all mean the same thing by "if" because they defer to the experts and it is expert usage that fixes the meaning of "if", Williamson says:

> The trouble is that McGee *is* an expert on conditionals. He publishes on them in the best journals. He does not defer in his use of "if" to any higher authorities.[43]

If correct, Williamson's argument undermines the key inferentialist assumption of meaning-constituting rules of inference.

But the entire force of Williamson's argument rests on homophonically translating in his imagined cases. Williamson applies homophonic translations when I and other conventionalists claim it is inappropriate to do so. This is the crucial step in his arguments, but all that Williamson says in defense of homophonic translation is that since we can imagine that those we are translating would (apparently) insist that they mean the very same thing that we do, we should translate them homophonically:

> To insist on applying … a non-homophonic translation scheme to them in the teeth of their protests would be to treat them less than fully seriously as human beings, like patients in need of old-fashioned psychiatric treatment, whose words are merely symptoms.[44]

Despite the rhetoric here, there is no sense in which translating someone non-homophonically in the face of their protests is treating them "less than fully seriously as human beings". Neither we nor they are perfect self-interpreters. What is more, even if they were, it wouldn't follow that they would be correct about all of their translation claims. Even if I knew perfectly well, in some magical sense, what I meant by Δ, I wouldn't be able to tell that Δ in my idiolect should or should not be translated into Δ in your idiolect unless I also knew perfectly well what *you* meant by Δ. And on no reasonable view of knowledge of meaning do we automatically have such knowledge.

So even if we assume, implausibly, that we have perfect access to what our own sentences mean, perfect access to what the sentences of others mean would also be required for Williamson's response to work. That is obviously beyond our ken. Even if I *intend* to mean by Δ exactly what you mean by Δ, intentions are not metasemantically sacrosanct. I can intend to speak Swahili all that I want, but if I simply prattle on as I normally do I will not succeed in carrying out my intentions (compare 2.II on stipulations). For conventionalists and most others, what our words mean is determined by how we use them.[45]

Given the general discussion of conceptual pluralism in chapter 5, we can easily diagnosis the initial plausibility of Williamson's examples—they all involve moving from, for example, the conditional to a closely related conditional-like concept (this can be spelled out more precisely using the semantic counterpart machinery from 5.III). When this similarity is absent, the homophonic translation isn't even tempting. Imagine instead that McGee thought the conditional was his office chair. Since Williamson aims to undermine *all* understanding-assent links and *all* meaning-determining rules, he needs to undermine these cases as well as more realistic cases.

We could construct a chain of individually more plausible theoretical assumptions leading to this point. Perhaps McGee becomes convinced by certain physicalist arguments that all things are physical, and thus that logical notions are physical too, since they exist and we talk about them. Following this, he becomes convinced, through serious study of Parmenides and Zeno, that all is one. Thus, in a sense, the conditional, being real, is one with his

office chair, also being real. We can imagine this or even more involved causal paths to this conclusion, but of course none of this would prove that homophonic translation is appropriate in this case. Just because it isn't intuitively clear at which precise point in a sequence of steps meaning has changed, it does not mean that meaning hasn't been changed somewhere along the way. We can start with a clear case of blue and, through a series of pairwise indistinguishable changes, end up with a clear case of green.

As far as I can tell, in his discussions Williamson rejects a transparent tie between meaning and use in favor of a semantically primitive public language that we all speak.[46] But I don't think that this can possibly be the correct view of language. As a matter of undisputed linguistic fact, new languages spin off from old ones and new senses for old words are coined all of the time. When speakers have at least the disposition to defer to each other and coordinate, a shared content can often be straightforwardly assumed, but not otherwise. In the cases we have imagined, where experts seem to be in a dispute about the valid laws for the conditional or whatnot, this condition is not satisfied. Experts, as Williamson noted, do not generally defer to each other's judgments. And there is no semantic glue that sticks words, whatever their use, to their old meanings. The reality of linguistic change and drift demonstrates this.

THERE IS NO quick and dirty refutation of the very idea of truth by convention. Sometimes in philosophy, a convincing argument is answered with a sense of bravura—you feel like you just watched a magician perform a trick. Maybe you were convinced in the moment, but the intuitive force of the argument returns the second you stop reading. By contrast, I hope, it is clear that the responses I have made to these prima facie convincing arguments don't rely on any tricks. I have aimed not just to show *that* these arguments fail, but *why* they fail. Yet just because there is no refutation of logical truth by convention doesn't mean that conventionalism can be extended without limit. Even an extension from logic to mathematics involves myriad obstacles and non-trivial difficulties.

[1] For the same reason, conventionalists do not need to respond to this argument by denying that necessary truths are themselves necessary. A fully standard account of necessity is open to conventionalists.

[2] This is done in Einheuser (2006).

[3] See Malcolm (1940) and Sidelle (2009).

[4] In my (2015*d*), which this section significantly updates.

[5] Sidelle (1989) and Lewy (1976); arguably versions of the argument are also found in Ewing (1940), Lewis (1946), Pap (1958), and Sober (2000), among other places.

[6] Yablo (1992), page 878; in the original there is a footnote giving page 53 of Lewy (1976) as the source of the internal quote.

[7] Boghossian (1996), pages 364–365.

[8] From Sider (2003), section 4.1.

[9] In my (2015*d*) this same basic argument is presented a bit more informally. Some may worry about the use of "*p*" in the first premise, but I think no clarity would be gained by adding formal complications (likewise for making 2 more explicit).

[10] This type of move is suggested in Haze's (2017) discussion of my (2015*d*). Relatedly, Topey (2019)—which I discovered while this book was going to press—argues that conventionalists should take conventions to explain the fact that all vixens are foxes, and other worldly facts. I have no strong objection to these if only "lightweight" facts are involved, but as I indicate in this section, I think conventionalists have a lot more flexibility here.

[11] Boghossian (1996), page 380.

[12] The argument actually used the modus tollens direction. And I suppressed the role of conjunction (the argument also plausibly assumes that "makes it the case that" distributes over conjunction, that is: if Δ makes it the case that p and q then Δ makes it the case that p and Δ makes it the case that q). Of course, *partial* explanatory contexts do not plausibly distribute over conjunctions—Δ can partially explain p and q by partially explaining p and having nothing at all to do with q. This is why proponents of the master argument aren't committed to absurd conclusions by premise (1) of the argument, like (for example) that since our conventions partially explain the truth of our sentence "JFK was assassinated" they must also partially explain the truth of the proposition expressed by this sentence, when they clearly do not. That argument fails by implicitly assuming that partial explanatory contexts distribute over conjunctions.

[13] I have sometimes encountered this charge in conversation, and Mark Balaguer raised this worry after reading an earlier version of this section.

[14] Compare the distinction between "aboutness" and "emptiness" in Unger (2017*b*).

[15] Compare Unger's (2014) notion of a concretely empty idea, which, despite what some commentators have thought, is not to be identified with a

necessarily true idea (see Unger (2017*a*)).

16 See my (2015*d*) for fuller discussion of this type of patch.

17 The idea of this modification was suggested to me by Bradford Skow.

18 See Stewart (1989).

19 Quine (1936).

20 Sider (2011), page 216. Benacerraf (1973) and Soames (2003) are similarly laudatory.

21 In my (2017*a*).

22 I don't endorse the principles whose names are enclosed in asterisks.

23 This terminology, in roughly this sense, was introduced in Thomson (1954).

24 As Quine himself noted, this regress is structurally analogous to a general regress for logical inference presented in the famous dialogue in Carroll (1895).

25 Boghossian (1996), page 381.

26 Quoted from the close of Quine (1936); see also Quine (1960*a*) for related discussion.

27 From Dummett (1959), which is Dummett's review of Wittgenstein (1956), quoted from pages 169–170 of Dummett (1978).

28 Wright (1980), chapter 18, interprets Dummett's challenge in this fashion.

29 This is perhaps developed most clearly at greatest length in Diamond (1976); see also the canonical Wittgenstein (1956). Dummett (1959) accepts this reading of Wittgenstein; Stroud (1965) argues against it.

30 Dummett (1959), quoted from page 170 of Dummett (1978).

31 See Wittgenstein (1953) and (1956); see also Wilson (1982).

32 For perhaps the earliest clear statements of this type of argument, see Whiteley (1936) and Black (1936); see also the discussion on pages 61–63 of Putnam (1971).

33 Blanchette (2001), pages 132–133.

34 Bennett (1961) rejects the vicious circularity in response to Dummett but also claims that while this move avoids the metaphysical worry, it doesn't answer the deeper, normative challenge. I have already addressed related normative challenges in 2.VII, 4.VII, 4.VIII, and chapter 6, so here I am focused on the metaphysical issues.

35 The debate took place on June 13, 1949. I have quoted from pages 600–601 of the transcript published as Ayer and Copleston (1957).

36 Ayer in fact made the appropriate response to Copleston in their debate—"No, because in that sense the notion of contradiction as you understand it wouldn't apply. … No, there are no laws that govern all games, but each game has a certain set of laws governing it. … What's common to all of them is that if the game is conducted in accordance with certain rules, then if you don't observe those rules, you're not playing that game, though possibly some other. … [S]urely all that you are saying is that in a language, namely the one we are now using, where one of the principles of correct reasoning is the

observance of the law of non-contradiction, anyone who violates this law isn't reasoning correctly. That is certainly a valid statement, but it *is* conventional." Quoted from pages 600–602 of Ayer and Copleston (1957).

37 This immediately continues the earlier Dummett (1959) quote.

38 Once again, see Bennett (1961).

39 Obviously we are not trying to explain the logical truth that for any language L, either L is trivial or L is not trivial—this logical truth, like all others, is given a wholly conventional explanation in the manner of 4.I.

40 Note that these points fall short of saying that syntax is *determinate*, that will be taken up in part III; see also my (2015*a*).

41 See Williamson (2006*b*), (2007), and (2013).

42 Williamson (2007), pages 92, 94; see McGee (1985) for the objection to modus ponens.

43 Williamson (2007), page 94.

44 Williamson (2007), page 115.

45 So does McGee mean something different by "if" than we do? I think it's a possibility, but there are various features of this particular case that muddy the waters; this matter is discussed at greater length in one of my unpublished papers.

46 This is apparent at several places in Williamson (2007), but Williamson's (1994) epistemic account of vagueness is also relevant.

PART III

REVITALIZING MATHEMATICAL CONVENTIONALISM

8

From Logic to Mathematics

Part II developed and defended an inferentialist-conventionalist theory of logic based on the background set out in part I. Here in part III I'll extend the same general approach to provide a conventionalist theory of mathematics. The extension is non-trivial, for mathematics has many features that logic lacks. These features cannot be handled by a direct application of the theory of part II; further substantive philosophical work is required.

To set the stage for that substantive philosophical work, it's worth clearly explaining why historical conventionalist accounts of mathematics are unsatisfying, and why a straightforward extension of the inferentialist theory of part II doesn't provide a satisfying philosophical account of mathematics. This involves setting out the two central problems with this extension: (1) the ontological commitments of mathematics; (2) the richness of mathematical truth.

I Mathematical Conventionalism in History

The parade of quotations in 1.I demonstrates that many historical conventionalists applied their conventionalism to mathematics as well as logic. But too often they didn't put very much thought into this extension. As usual, Ayer's bracing treatment in *Language, Truth, and Logic* provides a clear illustration. There he treated all necessary and *a priori* truths—including the truths of mathematics—in the same manner: as tautologies, true by definition.

Ayer's simplistic account doesn't show any awareness of the special difficulties involved in extending conventionalism from logic to mathematics. Some of these difficulties stem from Gödel's incompleteness theorems and related results. Since Gödel's

theorems appeared only five years before the first edition of Ayer's book, perhaps we can cut him some slack for not mentioning them, especially in a popular work.[1] As Paul Benacerraf wrote in a late retrospective:

> Of course, everyone knew nothing was ever as simple as Ayer made it out to be—*Language, Truth, and Logic* was the *Manifesto*, not *Das Kapital*. But didn't he have it approximately right? Wasn't the basic point correct, even if he had to fudge the details a bit for popular presentation? Or so almost everyone seemed to think.[2]

But even leaving Gödel aside, we shouldn't cut Ayer too much slack. He also ignored other problems with the extension to mathematics, including the difficulties posed by mathematical ontology. In fairness to Ayer though, these difficulties were ignored by virtually every other conventionalist as well.

But they weren't ignored by Carnap. When it first appeared in 1934, Carnap's *Logical Syntax of Language* was immediately the most important and interesting conventionalist work in existence. Despite being more technical than philosophical, *Logical Syntax* was unsurpassed in philosophical sophistication during the heyday of logical empiricism. Some of the key ideas needed for extending conventionalism to mathematics first appear in embryonic form in *Logical Syntax*. I will discuss *Logical Syntax* directly in 13.II. For now it suffices to say that the problems weren't fully solved and that few other serious conventionalist theories were developed.

Eventually mathematical conventionalism died out, in part because of the influence of the general anti-conventionalist arguments discussed in chapter 7. Additionally though, conventionalists never succeeded in meeting mathematics-specific challenges with positive theoretical developments, and people like John Pollock and Gödel himself advanced powerful arguments aimed directly at *mathematical* conventionalism (these arguments will be discussed in chapter 12).[3]

After the death of mathematical conventionalism proper, conventionalist-adjacent approaches slowly began to reappear, most prominently the Neo-Fregean approach to mathematics inaugurated by Crispin Wright.[4] Neo-Fregeans reject the dreaded label of

“conventionalism”, but there is a decent amount of overlap between their approach and conventionalism. Despite these similarities, the conventionalist theory I’ll be developing has several advantages over Neo-Fregeanism and other approaches that have been propounded in the literature. Or so I will attempt to demonstrate.

As a preliminary step toward developing my approach, we need to clearly understand why a direct application of part II’s inferentialist approach to logic will not do for mathematics proper.

II Mathematical Inferentialism

There is no immediate problem in extending unrestricted inferentialism from logic to mathematics, but when using only the resources discussed in part II you won’t get a theoretically adequate theory of mathematics. Still, the simple theory that results from a direct extension provides a base upon which a satisfying inferentialist-conventionalist theory can ultimately be built, so I’m going to outline that simple theory here.

With logic, I argued that bilateralist natural deduction rules provide a reasonable reconstruction of the actual logical rules in our language. For mathematics the analogous approach is to take the axioms of our mathematical theories, in natural deduction form, as the meaning-constituting rules of our mathematical language as it currently exists. In standard mathematics there isn’t an obvious difference between working with bilateralist rules instead of standard rules (see the appendix to this chapter), so I’ll work in a standard natural deduction system for the sake of familiarity. Of course, any standard natural deduction approach can be recast in bilateralist terms, so this shouldn’t be thought of as abandoning bilateralism. Let’s illustrate the simple extension for the crucial case of arithmetic.

Given unrestricted inferentialism, whatever the basic rules are for our arithmetical language, we can take them to be meaning constituting for our primitive arithmetical terminology by MAC, and automatically valid by the MVC. As unrestricted inferentialists, we can be confident going in that, whatever the rules for arithmetic turn out to

be, the unrestricted inferentialist can see them as implicitly defining our arithmetical vocabulary.

The standard formal model of the pure language of arithmetic is known as (first-order) *Peano arithmetic* (PA).[5] PA is an axiom system, and inferentialism doesn't apply directly to axiom systems, but we can easily get around this by reformulating PA as a natural deduction system with five non-logical rules, one for each of the standard Peano axioms. This will then give us a full language to which unrestricted inferentialism can be applied.[6]

Imagine to start that we're working in a standard first-order language $\mathcal{L}$ and we add to it a monadic predicate "N" (is a number), a constant "0", and a one-place function symbol "S*x*" ("successor of") with the following syntactic inference rules governing their deployment:

$$(P1)\ \frac{}{\mathrm{N}0} \qquad (P2)\ \frac{\mathrm{N}\alpha}{\mathrm{NS}\alpha} \qquad (P3)\ \frac{\mathrm{N}\alpha}{0 \neq \mathrm{S}\alpha} \qquad (P4)\ \frac{\mathrm{N}\alpha \quad \mathrm{N}\beta \quad \mathrm{S}\alpha = \mathrm{S}\beta}{\alpha = \beta}$$

$$(P5)\ \frac{\varphi(0) \quad \forall\xi(\varphi(\xi) \rightarrow \varphi(\mathrm{S}\xi)) \quad \mathrm{N}\beta}{\varphi(\beta)}$$

$$(P5)\ \frac{\varphi(0) \quad \forall\xi(\varphi(\xi) \rightarrow \varphi(\mathrm{S}\xi)) \quad \mathrm{N}\beta}{\varphi(\beta)}$$

I'll call these the *Peano rules* or the PA-rules.

Considered as English rules of inference, the first four of these are relatively straightforward. They concern basic facts about the number predicate, the zero constant, and the successor function. The fifth and final rule is a schematic rule-form of mathematical induction. As a rule schema, it is really infinitely many new inference rules. This is familiar from the logical case, as standard rules like modus ponens are also schematic. As discussed at various points, starting in 2.IV, these schematic rules should be understood in a fully open-ended fashion if they are to plausibly model our natural language practices.

This will be discussed at length in chapters 9 and 10, so I won't belabor it here.

In a full language including the Peano rules, each of the standard Peano axioms can be proved (for induction, this means that each instance of the schematic induction principle can be proved using the Peano rules). Conversely, if we were to start with the Peano axioms in a standard natural deduction system for logic we can prove each of the Peano rules as derived rules. If we add to our language the standard axioms for the addition ("+") and multiplication ("×") functions we get standard first-order Peano arithmetic.[7] Formally there is no real difference between PA and the PA-rules, at least with classical logic in the background. A simple extension of the theory of logic in part II to arithmetic says that the Peano rules are the meaning-constituting rules for our arithmetical notions, and that truth in arithmetic is wholly explained in terms of provability from the Peano rules. And like logical rules, the Peano rules are *syntactic*.

The Peano rules are supposed to model the actual meaning-determining rules of our arithmetical language. Some may balk at this. But I think they provide a plausible starting point reconstruction of our *current* language of *pure* arithmetic. The "current" and "pure" qualifiers are both important. Contemporary arithmetic, like other branches of mathematics, has developed and evolved to the point that it is now fairly well segmented and self-contained, purged of all quasi-empirical claims. But this process only came to fruition relatively recently. Because of this, my claim about the Peano rules and their relation to our practice should not be understood as a *historical* hypothesis. Rather than being born in the forbiddingly abstract form of the Peano rules, arithmetic—like most other branches of mathematics—initially grew out of direct applications to the world, probably in the form of generalizations of our practices of counting things and making cardinality comparisons (applied math will be discussed in 11.V).

To talk about arithmetic changing over time is to say that our arithmetical *concepts* changed and developed, not that arithmetical truth itself has changed and developed. Obviously, our mathematical concepts were created. Cognitive scientists can give us some insight into how this happened, perhaps.[8] But to reason from the creation of

our mathematical concepts to the creation of mathematical truth or mathematical objects is a use-mention confusion. Unfortunately, some mathematicians have been tempted by this fallacy; philosophers are typically, though not always, immune.[9]

The philosophically important point here is just that inferentialism about arithmetic, even when taking only the Peano rules as meaning-constituting, isn't committed to denying that pure arithmetic grew out of applied arithmetic. If the actual meaning-determining rules of arithmetic were initially rules that directly built in applications, unrestricted inferentialism can still be applied to offer a rational reconstruction. Take the rule-form of Hume's principle:

$$(\#I)\quad \frac{F \approx G}{\#F = \#G} \qquad (\#E)\quad \frac{\#F = \#G}{F \approx G}$$

Where "F" and "G" are predicate terms, "#" is a term-forming "number of"-operator, and "≈" is an equinumerosity relation (definable in either set theory or second-order logic). Hume's principle is central to the approach of Neo-Fregeans, but unrestricted inferentialism could just as easily apply to these or any other rules. Once again, the flexibility and purity of unrestricted inferentialism pays dividends.

Though starting with (#*I*) and (#*E*) is an option, I will continue to use the Peano rules as a base. I do this both to more closely cleave to current models of pure arithmetic and also to better illustrate how my approach generalizes to non-arithmetical mathematical theories. Any mathematical axioms whatsoever can be cast in natural deduction form and used to determine meanings. But it is not obvious which abstraction principles should be used for "set" terminology, for example (the obvious choice, Frege's Basic Law V, was famously shown to be inconsistent). Unfortunately, the simple approach of taking the Peano rules alone as meaning-determining arithmetical rules faces two major problems.

III Major Problem #1: Ontology from the Armchair

According to mathematical conventionalism, mathematical truths are *true by convention*. Like logical truths, they are true as trivial byproducts of the conventions governing our mathematical language.

How can this be right, given that mathematics is replete with existence claims? Mathematics contains theorems saying that solutions to particular equations *exist*. And mathematical theories directly assume the *existence* of particular kinds of abstract objects—natural numbers, rational numbers, real numbers, sets, points, lines, and so on. But the idea that existence claims could be *conventional*, *trivially true*, *true by definition*, *analytic* or the like has been very widely rejected, for seemingly good reasons. The canonical discussion concerns the infamous ontological argument for the existence of God.

The ontological argument is infamous precisely because it purports to prove that God exits using only armchair reasoning. The argument was invented by Sr. Anselm in his *Proslogion* and was given in a modified version by Descartes in the *Meditations*; modern versions of the argument have been offered by Kurt Gödel and Alvin Plantinga, among others.[10] A simple version runs as follows:

1. God's essence includes all perfections (*definition*)
2. Existence is a perfection (*definition*)
3. *So*: God's essence includes existence (1,2)
4. *So*: God exists (3)

The canonical objection to this type of argument was presaged in Gassendi's comments on Descartes's *Meditations.*[11] It was also powerfully voiced by Cleanthes in Hume's *Dialogues Concerning Natural Religion*:

> Whatever we conceive as existent, we can also conceive as non-existent. There is no being, therefore, whose non-existence implies a contradiction. Consequently there is no being whose existence is demonstrable.[12]

And it was given its most influential expression by Kant in the first *Critique*:

> …if we reject subject and predicate alike, there is no contradiction; for nothing is then left that can be contradicted. To posit a triangle, and yet reject its three angles, is self-contradictory; but there is no contradiction in rejecting the triangle together with its three angles. … I cannot form the least concept of a thing which, should it be rejected with all its predicates, leaves behind a contradiction…[13]

The standard reading of Kant's objection is that it concerns the *logical form* of existence statements. According to this reading, existence is not a predicate, but the ontological argument treats it as if it were (in premise 2) and so that premise is false and the argument fails.

But logical form is a red herring: one can accept that existence is a predicate without *thereby* accepting the ontological argument for the existence of God.[14] So the argument's problems can't merely be based on a misunderstanding of the logical form of existence claims. I say this not because I think that existence is a predicate (I don't), but because the standard reading misses the key insight behind Kant's criticism of the argument.[15]

Rather than logical form, the central issue is that an existence statement—whatever its logical form—seemingly cannot be analytic or otherwise *trivially* true. Kant's criticism is based on the thought that, for any kind of thing F, if you reject the existence of Fs, you can't in so doing entangle yourself in incoherence or contradiction.[16] This implies that you can *always* reject the existence of Fs without incoherence, for *any* F. And this point, even if not read into Kant's critique of the ontological argument, is quite plausible. For this reason, it is very widely accepted by philosophers.

If it is correct, then given the evident existential commitments of nearly all branches of mathematics, it seems that mathematical conventionalism cannot be maintained. It is largely for this reason that conventionalism has historically been associated with *nominalism*—the rejection of the existence of abstract objects.[17] I've already indicated that Ayer and most other conventionalists ignored this issue. Carnap is really the only prominent conventionalist to seriously deal with the ontology of mathematics. He did so in different ways in his earlier syntax period and his later semantic period.[18] Those with conventionalist-adjacent views, such as Neo-Fregeans,

have had a lot to say in attempting to defuse this issue, but I don't think that any previous approach, including Neo-Fregean approaches, has been completely successful at alleviating all of the worries.

It is important to see that the problem here is very different than the familiar problem of implicitly defining theoretical terms in science. There is a standard account of defining theoretical terms, the so-called Ramsey/Carnap/Lewis method.[19] Applying this method to arithmetic would tell us that the theoretical role of the "number" predicate is given by the existential closure of our arithmetical theory (say, PA). So then the "numbers" would be the things, whatever they may be, that satisfy the theory. But this still leaves us with the problem of *whether there are any numbers*. In science this is exactly what we want; the world determines whether or not there are physical things that satisfy our scientific theories. But offloading the mathematical existence facts to the world is not in the spirit of conventionalism. If a mathematical theory requires some extra-linguistic posits in order to be true, then mathematical conventionalism is false.

In chapter 9 I will provide a *metaontology*—a theory of existence, a metatheory for ontology—that shows how conventionalism about mathematics can be maintained without collapsing into nominalism and while continuing to take mathematics at face value. My approach to metaontology builds upon the approaches of Carnap and Dummett as well as the work of the Neo-Fregeans and metaontological writings by Hilary Putnam and Eli Hirsch.[20] This is an important step for mathematical conventionalists, but unfortunately even if the ontological issues are resolved, simple inferentialist theories of mathematics still face another major problem.

IV Major Problem #2: Determinate Truth

Standard mathematical claims—at least in basic mathematics—seem to be fully factual, either true or false, even if poor creatures like us aren't ever able to discover which. This is to say that there is a *fact of the matter* even about unproved mathematical claims likes

Goldbach's conjecture. Obviously, this doesn't apply to quasi-mathematical claims using vague terminology, such as "small". It may be neither true nor false that 97 is a *small* number, but this is entirely due to the indeterminacy contributed by "small", not to any indeterminacy in our arithmetical terminology.

The determinacy of standard mathematical claims, especially the claims of basic arithmetic, is intuitively plausible. It is hard to even imagine what it would be like for arithmetic to be indeterminate—how could it be indeterminate whether there is a largest twin prime pair? What kind of borderline case could there possibly be? The difficulty of understanding how arithmetical claims could possibly be vague or indeterminate is one reason for thinking they aren't.

But conventionalists seem unable, in principle, to give an account of determinacy even for basic arithmetic. In fact, almost *everyone* seems unable to give an account of determinacy even for basic arithmetic. A recent joint paper of mine with Daniel Waxman gives a general argument for this, building upon an earlier challenge by Hillary Putnam as developed by Hartry Field.[21] The central reason is based on Gödel's incompleteness theorems. Part II's account of logic applied perfectly only when something like the completeness theorem held (4.III). Though Gödel's "incompleteness" theorems concern *completeness* in a different sense—so-called *negation completeness*—the theorems and related results show that a determinate notion of arithmetical truth can't coincide with provability.

The argument for this is fairly simple: according to incompleteness, any consistent, recursive mathematical theory T that interprets even a fairly weak amount of arithmetic is negation incomplete—for some sentence ϕ in the language of T, T proves neither ϕ nor ϕ's negation.[22] There are non-recursive theories that don't have this drawback, but non-recursive theories cannot be generated by any computational process (assuming Church's thesis). So characterizing arithmetical truth in terms of our grasp of such a theory flirts with anti-naturalist cognitive spookiness. At least, this is a standard way of thinking, but matters are not quite so straightforward. Occasionally, in some respects at least, non-recursiveness can be finessed.[23] But without some plausible strategy for finessing things, learning a non-

recursive theory seems to violate the Cognitive Naturalism constraint from 1.IV.

This portion of the challenge plagues all philosophical approaches to arithmetic, not just conventionalism. But Platonist approaches have the option of appealing to theory-transcendent arithmetical facts in order to explain determinacy. The idea would be that our theories take us so far, but not all of the way. Still, arithmetic is determinate because, as with science, the worldly facts about numbers and sets and the like outstrip our practice.[24] But this type of approach violates both the Totality principle and the Metaphysical Naturalism constraint from 1.IV. Conventionalists and other naturalists don't accept practice-transcendent mathematical facts. Accepting such metaphysical extravagances would be a case of the cure being worse than the disease.

The determinacy of mathematical truth is a general puzzle. Conventionalists are forced, by their commitment to naturalism, to face the puzzle honestly and without any outsourcing of determinacy to the physical or non-physical world. While the puzzle arises for arithmetic, more involved theories of mathematics with computationally richer notions of truth pose even more of a problem. But there is more wiggle room here. It is not intuitively clear that set-theoretic truth (for example) is fully determinate. I suspect that the majority opinion is that there is substantial indeterminacy in set theory. In any case, I'm happy to leave set-theoretic truth as a spoil to the victors, but if conventionalists can't provide an account of the determinacy of arithmetic then arithmetical conventionalism is unsatisfying. This doesn't mean the fortunes of conventionalism come down to this and only this. It would still be better to be a conventionalist without determinacy than to be a Platonist with determinacy. But we want a *satisfying* conventionalist theory, and that requires explaining arithmetical determinacy.

Some readers might be wondering why this problem is at all serious for the conventionalist. Can't we solve it at once, with a single axiom, presented in rule-form? Letting "*T*" and "*F*" be truth and falsity predicates, respectively, and using brackets as a name-forming device, consider the following axiom schema:

$$\text{(Bivalence)} \qquad T[\phi] \vee F[\phi]$$

As an open-ended axiom schema applied to our arithmetical language (Bivalence) says that any sentence in the language is either true or false. This doesn't quite capture the right intuitive content though, since the informal bivalence principle uses exclusive disjunction, while (Bivalence) uses inclusive disjunction. This could be rectified by explicitly excluding cases of overlap:

$$\text{(Bivalence+)} \qquad (T[\phi] \vee F[\phi]) \wedge \neg(T[\phi] \wedge F[\phi])$$

Neither of these is very helpful though, since if we have the standard truth rules and the usual definition of falsity as truth of negation, these are already classical theorems. So it seems that both forms of bivalence are *already* necessary truths in our language. Does this mean that we have already solved the problem without knowing it?

No, we have not. The problem is that even the truth of (every instance of) (Bivalence+) doesn't secure determinacy in the required sense. This type of point is familiar from discussions of supervaluationism and vagueness; it is also related to the problem I raised for deflationist approaches to metasemantics in 4.IV.[25] Say that a sentence is *determinately true* just in case it is true in each admissible interpretation of our language, and similarly for falsity. The principle of (Bivalence+) doesn't ensure that every sentence in the language of arithmetic (or set theory) is either determinately true or determinately false. The (Bivalence) schema can itself be determinately true even when ϕ is true in some interpretations, false in others, and so neither determinately true nor determinately false.

So consider adding a "determinately" operator to our language, written as "$\triangle$", and trying to stipulate away indeterminacy with the following principles:

$$\text{(}\triangle\text{ Bivalence)} \qquad \triangle\, T[\phi] \vee \triangle\, F[\phi]$$

$$\text{(}\triangle\text{ Bivalence+)} \qquad (\triangle\, T[\phi] \vee \triangle\, F[\phi]) \wedge \neg(\triangle\, T[\phi] \wedge \triangle\, F[\phi])$$

Unfortunately, building these claims into the logic of determinacy undermines its meaning. What I mean is that in order for "$\triangle$" to express *determinacy*, in the intended sense, it needs to align with the intuitive content of our notion. If we define "$\triangle$" in some other way, there is no reason to think that it expresses the relevant notion of determinacy. The same points go for determinacy predicates as well.

The overall problem here is entirely general, in any case. Moving from our starting theory T to T+(Bivalence) or even T + ($\triangle$ Bivalence+) only provides more grist for the mill. If this new theory is itself recursive, then it is negation incomplete if it is consistent and interprets a minimal amount of arithmetic. So call this theory T′—T′ will have its own Gödel-Rosser sentence. So a determinate notion of truth cannot be explained in terms of provability from T′. To whatever extent determinacy arises, it must be explained.[26] There is no royal road to accounting for arithmetical determinacy on conventionalist grounds. We have to go up the hard way. I will attempt the ascent in chapter 10.

≈≈≈≈≈≈≈≈

THE INFERENTIALIST APPROACH to logic developed in part II can be applied directly to mathematics simply by reformulating the axioms and axiom schemas in any branch of pure mathematics as meaning-determining inference rules for the special vocabulary in said branch. However, that approach leaves lingering mysteries about how mathematical existence claims can be true by convention and doesn't lead to a satisfying account of truth even for basic arithmetic. As such, a satisfying conventionalist theory of mathematics requires more work. This work is the task of part III.

Appendix: Bilateralism and Arithmetical Incompleteness

Say that a bilateralist system *b* is *attitude complete* just in case for every sentence ϕ in the language of the system either $\vdash_b + \phi$ or $\vdash_b - \phi$ and a*ttitude incomplete* otherwise, and *attitude consistent* just in case

for no sentence ϕ in the language of the system do we have both $\vdash_b +\phi$ and $\vdash_b -\phi$ and *attitude inconsistent* otherwise. Let's also reintroduce the following definitions:

(t) Let t be a translation from the signed language to the unsigned language such that $t(+\phi) = \phi$ and $t(-\phi) = \neg\phi$

(Γ^t) If Γ is a set of signed sentences, let $\Gamma^t = \{t(\alpha) | \alpha \in \Gamma\}$

Let Γ be a set of signed sentences and α a signed sentence and recall that when α is a signed sentence, then α^* is just like α except having opposite sign.

Lemma. *if* $\Gamma \vdash_b \alpha$ *then* $\Gamma^t \vdash t(\alpha)$

Proof. By inspection of chapter 2's *b*-rules (including the added structural rules) and the definition of t, a *b*-proof of α from Γ transforms into a standard proof of $t(\alpha)$ from Γ^t; the only case that isn't immediately obvious involves negative signs, but since $+\neg\phi \dashv\vdash_b -\phi$, our result follows. __

Call a bilateralist theory BA a *theory of arithmetic* just in case BA^t satisfies the conditions for Gödel's first incompleteness theorem.

Theorem. *If* BA *is a recursive, attitude-consistent system of bilateralist arithmetic, then* BA *is attitude incomplete*

Proof. Assume otherwise, so that BA is an attitude-consistent, attitude-complete bilateralist theory of arithmetic and consider the Gödel/Rosser sentence G for BA^t. By assumption, either $BA \vdash_b +G$ or $BA \vdash_b -G$, if the first, then by our lemma $BA^t \vdash G$, contradicting Gödel's theorem; if the second, then by our lemma, $BA^t \vdash \neg G$, contradicting Rosser's theorem. Either way, the assumption that BA is a recursive, attitude-consistent, attitude-complete theory of arithmetic is false. __

1 Ayer briefly discusses Gödelian issues in his (1992).

2 Benacerraf (1998), page 43.

3 See Gödel (1953/9) and Pollock (1967).

4 See Wright (1983), Hale (1987), and the essays in Hale and Wright (2001). Also relevant is Dummett (1956), as a precursor to NeoFregeanism, and more recent approaches influenced by the Neo-Fregeans, including Fine (2005), Rayo (2013), and Linnebo (2018).

5 PA derives from Dedekind (1888) and Peano (1889).

6 I first did this in print in my (2015*a*), but arithmetical axioms were formulated as natural deduction rules previously in Tennant (1978) and Parsons (2008), for example. And induction is typically formulated in rule-form in primitive recursive arithmetic (PRA), which derives from Skolem (1923).

7 These axioms are: (i) $\forall x(x + 0 = x)$, (ii) $\forall x\forall y(x + s(y) = s(x + y))$, (iii) $\forall y(y \times 0 = 0)$, and (iv) $\forall y\forall z(y \times s(z) = y \times z + y)$. They could also be added in rule-form.

8 See Lakoff and Núñez (2000) for a book-length treatment.

9 Mathematicians bite from the apple in Hersh (1997) and Davis and Hersh (1981). For an overview of why post–Frege (1884) philosophers haven't been tempted, see Burgess (2003). Davis and Hersh were greatly influenced by the quasi-historical treatment of mathematics in Lakatos (1976).

10 See Charlseworth (1965) for an English translation of Anselm's *Proslogion*, Cottingham, Stoothoff, and Murdoch (1984) for Descartes's *Meditations*, Feferman, Dawson Jr., Goldfarb, Parsons, and Solovay (1995) for Gödel's version, and Plantinga (1974).

11 Gassendi's comments on Descartes are in Cottingham, Stoothoff, and Murdoch (1984).

12 Quoted from Hume (1779), part IX.

13 Kant (1787), A595–A596; I have quoted from pages 502–503 of Norman Kemp Smith's translation.

14 See, for example: Fine (2009), McGinn (2000), and Parsons (1980) and (1982).

15 I respond to Fine's (2009) argument for an existence predicate in my (forthcoming *b*).

16 The opening of Field (1984*b*) offers a similar reading of Kant's critique.

17 This is the standard modern usage, stemming from Goodman and Quine (1947), but the term "nominalism" originally concerned only the rejection of universals.

18 His most influential work on ontology is Carnap (1950).

19 See Ramsey (1978), Carnap (1958), and Lewis (1970).

20 See Carnap (1950), Dummett (1956), Wright (1983), the essays in Hale and Wright (2001), Putnam (2004), and the essays in Hirsch (2011). In addition, see the important critical discussions of the metaontology of Neo-Fregeanism (or "Abstractionism", as Hale and Wright have more recently been terming it) in

Eklund (2006), Hale (2007), Hawley (2007), Sider (2007), and Hale and Wright (2009).

21 See Warren and Waxman (forthcoming *a*), Putnam (1980), and Field (1994*b*), (1998*a*), and (1998*b*).

22 It is enough that T relatively interprets Robinson arithmetic. My statement of Gödel's theorem also incorporates the strengthening due to Rosser (1936).

23 See Benacerraf (1996) for an example.

24 Arguably this is what Gödel (1964), McGee (2000), and Sider (2011), among others, think.

25 See Mehlberg's (1958) distinction between "logical" and "meta-logical" versions of the law of excluded middle; for influential formal developments of supervaluationism see Van Fraassen (1966) and Fine (1975); and see Williamson (1994), chapter 5, for a general overview.

26 See the metasemantic challenge of Warren and Waxman (forthcoming *a*).

9

Metaontology, Existence, and Reference

There exist infinitely many prime numbers; this is a theorem in elementary number theory. It is also old—a version of it is proposition 20 in book 9 of Euclid's *Elements*.[1] But if infinitely many prime numbers exist, then prime numbers exist, and so numbers exist. Any non-revisionary philosophical theory of mathematics needs to secure the truth of this and the many other existence claims that permeate mathematical discourse. But mathematical conventionalism must make these and other existence claims *true by convention*, and it is very widely thought that existence claims, at least, cannot be trivially true, analytic, or true by convention (8.III).

Both Carnap and the later Neo-Fregeans discussed and made progress on this issue, but I don't think that a completely general and fully satisfying account of existence in mathematics has yet appeared. The key to providing such an account is to integrate an account of *mathematical* existence with a general and well-motivated account of existence more generally—a *metaontology*. Both Hilary Putnam's conceptual relativity and Eli Hirsch's quantifier variance will be of use here. But both Putnam and Hirsch have made some unhelpful terminological choices, and Putnam (but not Hirsch) has often mixed his metaontological views with a dubious form of anti-realism. So in this chapter I will set things up in my own way to develop and defend a general metaontological theory along lines inspired by Putnam, Hirsch, Carnap, and the Neo-Fregeans. When applied to mathematics this metaontology will vindicate mathematical existence and reference to mathematical objects *by convention*.

I Trivial Ontological Realism

Quine influentially analyzed ontological commitment using the existential quantifier.[2] In slogan form: "to be is to be the value of a variable". Following Quine's account, let's say that a theory T is ontologically committed to the existence of Fs, for sortal predicate "F", just in case $T \models \exists y Fy$, and that a language L is ontologically committed to the existence of Fs just in case there is a true sentence in the language that would entail the sentence "$\exists y$Fy", provided the needed vocabulary had been introduced.[3]

In adopting this "Quinean" approach I am not endorsing any of the more controversial metaontological doctrines that Quine himself attached to his approach, such as: rejection of higher-order quantification, rejection of non-extensional contexts, and the fundamental ontological relevance of eliminability and indispensability considerations.[4] My central point is just to analyze existence claims using existential quantification instead of an existence predicate. If anything, this makes the problems discussed in this chapter even more difficult to deal with, so no cheating is taking place by fixing upon this entirely standard approach to existence.[5]

So using this Quinean approach, our arithmetical theory A is ontologically committed to numbers (using the number predicate "N"—8.II) just in case A entails "$\exists x$Nx", and similarly for the ontological commitments of our arithmetical language. Clearly the following is a correct proof that there are numbers, using only 8.II's Peano rules:

$$\frac{\mathrm{N0}}{\exists x \mathrm{N}x}$$

So our arithmetical language and thus our "theory" of arithmetic, as (partially) reconstructed by the Peano rules, entails "$\exists x$Nx" and so is ontologically committed to numbers. The above proof uses (*P*1) together with a standard ∃-introduction rule. For inferentialists, these rules are automatically valid (by the MVC), so the sentence "$\exists x$Nx" is necessarily true, automatically. In other words: *numbers exist*, and the above proof quickly and easily establishes this on conventionalist grounds.

One prima facie worry about this simple procedure also takes off from another Quinean slogan—"no entity without identity".[6] The idea behind this slogan is that we shouldn't accept putative types of objects without clear identity conditions for said type.

> Take, for instance, the possible fat man in that doorway; and, again, the possible bald man in that doorway. Are they the same possible man, or two possible men? How do we decide? How many possible men are there in that doorway? Are there more possible thin ones than fat ones? How many of them are alike? Or would their being alike make them one? Are no *two* possible things alike? Is this the same as saying that it is impossible for two things to be alike? Or, finally, is the concept of identity simply inapplicable to unactualized possibles? But what sense can be found in talking of entities which cannot meaningfully be said to be identical with themselves and distinct from one another?[7]

This is both plausible and sensible, in big-picture terms. But when we dig deeper, we can see that the crucial notion of "identity conditions" can be read in either a strong or a weak manner.[8] The strong manner requires that before we accept *X*s and associated vocabulary we have identity conditions for *X*s that are stated in terms of the notions we already accept. This seems to require something like an *abstraction principle*:

$$\text{(ABS)} \qquad \forall\alpha\forall\beta(t(\alpha) = t(\beta) \leftrightarrow Eq(\alpha,\beta))$$

Where "*t*(*x*)" is a term-forming operator for the new type of object, *α* and *β* are variables of an already understood type, and "*Eq*(*x*,*y*)" is an equivalence relation (a symmetric, reflexive, and transitive relation) on objects of the type of *α* and *β*. In theory, a principle of the form (ABS) is supposed to enable us to understand truth conditions for the new objects in terms that we already understand. This is the approach taken by the Neo-Fregeans, but the purity of the approach is complicated by the fact that their favored applications require abstraction principles that are both impredicative and second-order.

Requiring that identity conditions for *every* kind of object be stated in more fundamental terms seems to lead to an infinite regress. For any kind X_1, we only understand X_1 in terms of another kind X_2, and so on and so forth. Maybe some would quibble with the idea that we must understand one *kind* in terms of another. But whether we call them "kinds" or not, (ABS) will give identity conditions for X_1 in terms of a relation over some other "type" or "kind" or *whatever* of things. To avoid going in a circle, there must be a point at which we reach the conceptually fundamental entities of our conceptual scheme. For these entities we cannot hope to explain their identity conditions in other terms. Spacetime points, material objects, and some mathematical objects are, plausibly, among the entities that we understand on their own terms, rather than in terms of some other type of thing.

This point can be accommodated by understanding the Quinean demand for identity in a weak manner. The weak manner requires not that we give identity conditions for the *X*s in independent terms, but only that we give some condition of identity, of whatever kind. It is fine if the identity conditions are themselves stated using *X*-vocabulary. This matches what we often do, in practice. When the beginner learns set theory, they are taught the axiom of extensionality:

$$\text{(Extensionality)} \qquad \forall\alpha\forall\beta(\text{Set}(\alpha)\wedge\text{Set}(\beta)\rightarrow(\alpha=\beta\leftrightarrow\forall z(z\in\alpha\leftrightarrow z\in\beta)))$$

Extensionality tells us that two sets are identical just in case they share all and only the same members. But the notion of "membership" is itself a set-theoretic notion that is introduced along with our "set" predicate and not understood in prior, non-set-theoretic terms. Yet (Extensionality) still tells us all that we need to know about the identity of sets. With it we are able to compare sets and assess their identity and non-identity in a wide range of cases. Understood properly, I think that this is all that Quine's slogan demands.

Similarly, the Peano rules seem to tell us all that we need to know about the identity of numbers. There are issues about cross-sortal identifications (Frege's famous Caesar problem) but these are issues for the strong reading as well (see section VI below for discussion). Moving to an abstraction principle doesn't provide a magical

philosophical elixir. What it does do is increase the *dialectical* strength of ontological realism about Fs. If somebody already understands and accepts an equivalence relation over Fs and Gs, they can't coherently claim not to understand identity conditions for Fs if those identity conditions are characterized directly in terms of that equivalence relation.[9] But there is nothing improper about introducing some new vocabulary and stating identity conditions for the new things directly in terms of the new vocabulary. Not everything can be understood in terms of something else.

This metaontological demand for identity conditions isn't in tension with unrestricted inferentialism. Talk any way you like, any way at all. Use any rules you can dream up and follow. But not all ways of talking or using terms like "exists" or "there are" will count as ways of talking about *objects* or *things*. But if we have quantification and identity conditions, that is all that is needed. I don't think there is any deeper essence to our concept of existence than this.

What goes for *existence*, also goes for *reference*. Following general suggestions by Frege (for example, the infamous discussion of the "context principle") that have been developed in the literature on Neo-Fregeanism, say that a theory T is referentially committed to an object α, for singular term "α", just in case $T \models \exists x(x = \alpha)$.[10] And say that language is referentially committed to an object α just in case there is a true sentence in the language that entails the sentence "$\exists x(x = \alpha)$" upon introduction of the needed vocabulary.

Using the Peano rules and our standard rules for quantification and identity, we can easily prove that our arithmetical theory A is referentially committed to the existence of zero, of one, of two, and so on and so forth. Neo-Fregeans usually prefer to approach the ontology of mathematics through *reference* rather than *existence*, but the fundamental problem space is the same in either case. For conventionalists, both the existence of numbers and reference to numbers via numerical terms are easy to establish using the Peano rules.

According to mathematical inferentialism/conventionalism, the existence of numbers is not a substantive and controversial fact about our world, but is instead a *trivial byproduct* of our using our arithmetical language in the way that we do. With a fully standard

conception of ontological and referential commitments, we can see that the historical association of conventionalism with nominalism was a mistake. Conventionalism does not entail nominalism, instead it entails ontological realism about numbers and other mathematical objects, which is the negation of nominalism.[11] It can't be overstressed that the ontological realism here is *trivial*—it doesn't involve metaphysically suspect *Platonic* assumptions about the nature of numbers and other abstract objects. This is a conventionalist-friendly, and thus, scientifically friendly, ontological realism about mathematics.

There are infinitely many natural numbers, so for conventionalists it is also trivial that the world contains infinitely many things. I don't think that more things existing trivially adds any additional philosophical baggage. If one thing can trivially exist, why not two, or $\aleph_0$, or $\aleph_{217}$?[12] There trivially being infinitely many things is not a substantive feature of reality. The source of this claim's truth, as with a claim about a single number, is the linguistic conventions of our arithmetical language. For this reason, claiming that there are infinitely many numbers is very different than claiming that there are infinitely many material objects. *That* is a claim about independently existing aspects of the physical world. With the existential entailments of arithmetic we aren't making a new claim about old things, we are making an old claim about new things.

This is my ontological story, but the considerations adduced in 8.III's discussion of the ontological argument for the existence of God put pressure on this approach to existence. We saw there that there were prima facie powerful reasons for thinking that existence claims cannot be analytic or otherwise trivially true. But as we have just seen, conventionalists are committed to the triviality of mathematical existence claims in precisely this supposedly problematic manner. Things could easily devolve into a standoff here, with opponents claiming that existence claims cannot be trivially true and conventionalists retorting that yes, they can, as the above argument shows. No versus Yes, modus tollens and modus ponens. Around and around we might go with little hope of escape. The standoff can't be broken at the merely ontological level. To break this deadlock, we

need go meta. Finding the right way to think about metaontology is the only way to advance matters.

Without necessarily being conventionalists, a number of previous ontologists have endorsed arguments for the existence of numbers that are something like the one I gave above, but they typically did so without offering detailed metaontological accounts of what they were doing.[13] By contrast, Neo-Fregeans have endorsed trivial existence and provided detailed metaontological justifications of their moves. But there remains a great deal of debate about exactly how the Neo-Fregean metaontology should be understood.[14] Rather than engaging in Fregean or Neo-Fregean exegesis, I will instead make a (mostly) fresh start, aiming to show how, properly developed, a variant of the metaontological views of Putnam, Hirsch, and (perhaps) Carnap provides exactly what is needed both in mathematics and in general.[15]

II The Metatheory of Existence and Reference

Inferentialism is a top-down approach to metasemantics: it explains the semantic facts about subsentential expressions in terms of the semantic facts about sentences. I presented a top-down inferentialist account of the logical expressions, including the quantifiers, in chapter 3 and developed it throughout chapters 4 and 5.

Top-down approaches contrast with bottom-up metasemantic theories, according to which the semantic facts concerning subsentential expressions explain or generate the semantic facts concerning sentences. Many bottom-up theorists think semantic theories for (a fragment of) a language are explanatory. The truth conditions of sentences are then seen as byproducts of the interaction of the semantic values of the components of the sentence, so that a basic arithmetical sentence like “3 is a prime number” is true *because* “3” refers to an object that has the property expressed by the predicate “is a prime number”. By contrast, while top-down approaches can easily accept these equivalences, they don’t accept the explanatory direction read into them by bottom-up theorists.

A truth-conditional semantic theory, with clauses like "'3 is a prime number' is true if and only if the referent of '3' has the property expressed by 'is a prime number'", does not itself entail a bottom-up metasemantics, despite what is sometimes thought.[16] These equivalences, even if necessary, have no explanatory direction built into them, *metasemantically* speaking. So top-down theorists, including inferentialists, can accept fully standard Tarski-style semantic theories for mathematics and the rest of our language, provided that they don't see the explanatory direction as running from subsentential semantics to sentential semantics, ceteris paribus.

This was already illustrated in 3.IV's demonstration that inference rules suffice to determine standard semantic values for the logical expressions. Here I am going to generalize and extend what was implicit in that discussion, in order to provide a general metaontology on inferentialist grounds. A metaontology is a theory of the nature of existence and reference, a metatheory for ontology, as it were.[17] As noted at the end of the previous section, the approach I will develop is related to ideas defended by Carnap, Putnam, and Hirsch. However, there is a lot of debate about how their views should be understood. So I will introduce some new terminology but will also build on the philosophical foundations offered by previous philosophers.[18]

In previous work I called my approach *quantifier deflationism*. The word "deflationism" is overused in contemporary philosophy, but I used it for its suggestiveness. I have now come to think the risk of confusion is too large and the impact of the term is too diluted for this practice to be warranted. So I now prefer to call my view a brand of metadeflationism, which is particularly fitting, since it is a *meta*semantic and *meta*ontological version of deflationism. Here I'll develop metadeflationism about quantification and reference. On my view, quantification and reference aren't playing fundamental explanatory roles in metasemantics, and thus any pretensions to such roles in metaontology stand deflated. In the next section I'll argue that plausible versions of quantifier metadeflationism, including my own, lead to pluralism about existence and reference. Here I will explain the basic metaontological view itself in a clear fashion.

On a metadeflationist picture subsentential expressions aren't explanatorily freestanding semantic natural kinds, so how are these expressions to be identified and singled out? To put the question less epistemically: in virtue of which features is a particular subsentential expression in a given a language a *quantifier*? For inferentialists, the correct answer is obvious: a quantifier is a subsentential expression that plays the inferential role of a quantifier. And the same is true for all of the other logical expressions as well (as discussed in 5.III, "inferential role" here should be understood *structurally* so that notation and grammar don't matter).

In my previous published discussion I called this idea "quantifier inferentialism", but given the danger of confusing this point with the more general inferentialist metasemantics of part II, here I will call it quantifier role priority:

> **Quantifier Role Priority:** Subsentential expression *Q* in a language *L* is an (unrestricted) type *i* quantifier expression just in case *Q* plays, in *L*, the inferential role of an (unrestricted) type *i* quantifier expression.

This isn't a brand of full metasemantic inferentialism since it doesn't say anything about meaning-determining inference rules. Instead, role priority is about what it is that makes something an expression of a given kind. The idea is that inferential role makes an expression a quantifier, but this thought can be endorsed without thinking that the inferential role is, itself, determined by automatically valid inference rules for the expression (3.I).

Here I'll assume for simplicity that the constitutive roles of the standard existential and universal quantifiers are given by the canonical natural deduction rules:

$$(\exists I)\ \frac{\phi}{\exists\zeta\phi^{[\alpha/\zeta]}} \qquad (\exists E)\ \frac{\exists\zeta\phi^{[\alpha/\zeta]} \qquad \begin{array}{c}(\phi)^n\\ \vdots \\ \psi\end{array}}{\psi}\,n$$

provided that in ($\exists E$) α isn't in $\phi^{[\alpha/\zeta]}$ (the result of replacing occurrences of α in ϕ with ζ) or ψ or in extra assumptions that were used in the derivation of ψ. And:

$$(\forall I)\ \frac{\phi}{\forall\zeta\phi^{[\alpha/\zeta]}} \qquad (\forall E)\ \frac{\forall\zeta\phi}{\phi^{[\zeta/\alpha]}}$$

Provided that in ($\forall I$) α isn't in $\phi^{[\alpha/\zeta]}$ or any assumptions that were used in the derivation of ϕ. Using the bilateralist quantifier rules from 3.IV wouldn't change anything of note, but I'll continue with the policy of eschewing bilateralism in part III in order to simplify my presentation. The *unrestricted* roles for these quantifiers are given by the fully open-ended versions of these rules, where they continue to apply in well-behaved expansions of the language in question (see 10.II).[19]

They key point is that if there is a subsentential expression in a given language that plays the (structural) inferential role of an unrestricted existential quantifier, then that expression *is*, really and truly, an unrestricted existential quantifier, *semantically* speaking. And what goes for the quantifiers goes for every other type of subsentential expression as well. We've already seen something like this with the logical notions in part II, but I mean for this to apply to predicates and names as well:

> **Predicate Role Priority**: Subsentential expression Δ in a language L is an n-place predicate expression just in case Δ plays, in L, the inferential role of an n-place predicate expression.
>
> **Name Role Priority**: Subsentential expression α in a language L is a name expression just in case α plays, in L, the inferential role of a name expression.

In first-order languages, predicate and name roles are very easily distinguished, since names but not predicates can be generalized upon by quantifiers and can flank the identity sign. These inferential differences are quite close to grammatical differences, but that is not

a problem. Quantifiers and identity too are, as above, singled out by their inferential role, but there is no deep circularity here, only the harmless type of circularity common to functional definitions of all kinds.[20]

Standard mathematical theories like PA and ZF are first-order, so for the purposes of part III it is enough to distinguish names from predicates by way of interaction with quantifiers. But it would be nice to have a more general account, applying even to second-order languages. In these languages predicates too can be replaced by variables and generalized upon by quantifiers, and while predicates can't flank the standard identity sign, they can flank higher-order relation signs expressing analogs of identity. Frank Ramsey once used points like these to challenge the very idea of a semantically relevant distinction between names and predicates.[21] The difficulty is that in a second-order context the name role and the predicate role are (seemingly) structurally identical, differing only in semantically irrelevant grammatical features. Responses to the challenge have been offered by Dummett, Strawson, Geach, Hale, and Brandom, among others.[22] The most popular idea for marking a difference between names and predicates goes back to Aristotle.

In modern terms, the key Aristotelian idea is that predicates but not names have *contraries*. Aristotle is wise. And there is something such that anyone is it just in case they are not wise. But we have no reason to think that there is someone who is something just in case it is not the case that Aristotle is that thing.[23] Logically speaking, this difference comes down to the presence of a second-order comprehension principle in the background (second-order) logic:

$$\text{(Comp)} \qquad \exists X^n \forall x_1, \ldots, \forall x_n (X^n x_1, \ldots, x_n \leftrightarrow \phi)$$

Where "X^n" does not occur free in ϕ.[24] To use the example just given, with "W" a one-place "is wise" predicate, (Comp) gives us a contrary for "W":

$$\exists X \forall x (Xx \leftrightarrow \neg Wx)$$

But we don't accept any similar principle for names or terms. This illustrates that we use names and predicates differently, considered as categories. Andrew Parisi has recently objected to drawing the distinction between names and predicates by appeal to comprehension:

> …the truth of comprehension schemas are not obviously more basic than the distinction between singular terms and predicates. Comprehension schemas … cannot be justified without recourse to a distinction between names and predicates.[25]

But for inferentialists, this is the wrong way to think about things. We have in our language a grammatical distinction between *A*s and *B*s, and to go along with this grammatical distinction we use *A*s but not *B*s in accordance with (Comp) understood in a fully open-ended fashion (or vice versa). This fact about usage is what *makes A*s predicates and *B*s names (or vice versa). So there is no prior task of "justifying" (Comp) for *A*s as opposed to *B*s, at least when working in an inferentialist background. Our usage is different for the items in the two categories, and so the name role and the predicate role are distinct, even in second-order languages. In the next chapter I will discuss how beings like us can understand schematic principles like (Comp) so that they have all of the power they seem to have, but for now the point is just that (Comp) and related principles can lead to inferential differences between names and predicates in second-order languages.

In natural languages, there will also be other features of use that might be relevant to differentiating the behavior of names and predicates. For example, natural language uses of names are tied to demonstratives, pronouns, and other features that don't play a direct role in most formal languages. But the only part of natural language that I will be directly concerned with here is the mathematical portion, so we can safely ignore these additional complications. The name role and the predicate role for the whole of natural language are probably much more complex, but while there is more to specify, there is also more to work with in drawing all of the relevant distinctions. I see no reason to think there is any in principle difficulty

in applying and distinguishing **Name Role Priority and Predicate Role Priority**, in general.

My approach is to see *all* semantic categories as deriving from syntactic/inferential categories, against the backdrop of a top-down approach to metasemantics. General brands of quantifier, predicate, and name metadeflationism arise from the combination of a top-down approach to metasemantics with the appropriate version of role priority:

> **Top Down + Quantifier Role Priority = Quantifier Metadeflationism**
> **Top Down + Predicate Role Priority = Predicate Metadeflationism**
> **Top Down + Name Role Priority = Name Metadeflationism**

I am using **Top Down** as a placeholder for any top-down metasemantics. In chapter 3, I endorsed an inferentialist metasemantic theory, which is one top-down approach, but there are others. This generality allows us to see that the view of quantifiers that is at the heart of Putnam's *conceptual relativity* is a version of quantifier metadeflationaism, as is the view at the heart of Hirsch's *quantifier variance*. Putnam tends to focus on use-based approaches to meaning, Hirsch focuses on charity in interpretation, and I focus on linguistic rules of inference. But as I pointed out in 3.VII, these all amount to roughly the same thing. As far as I can see, there isn't a great deal of metasemantic difference on these points between Putnam, Hirsch, and myself.

Neo-Fregeans like Wright and Hale have also been read as (accused of?) having a metasemantic view of this kind. The exegetical issue is complicated though. Neo-Fregeans have done more than anyone else to explicitly defend something like **Name Role Priority and Name Metadeflationism**. They have even taken steps to spell out the natural language role of names.[26] But they have been hesitant at best and hostile at worst about adapting their story about names to the quantifiers.[27] I don't think their hostility is justified. In fact, I don't think that the overall package of views that they seem to endorse is coherent. The inferential roles of quantifiers, predicates,

and names are all interdefined and interrelated, so if any one of them is treated in terms of role priority, then all of the others must be as well.

What role priority does, in effect, is deny the possibility of a perfect inferential simulation of a quantifier that is not, *semantically*, a quantifier. But if Neo-Fregeans allow for this possibility with quantifiers, then the non-inferential notion of quantification could be used to induce a non-inferential notion of a name, contrary to the explicit claims at the heart of Neo-Fregeanism. This kind of mixed approach harbors deep internal tensions. So I think we should be metadeflationists about either all or none of our subsentential categories. If the Neo-Fregeans really and truly accept metadeflationism about names while rejecting it for quantifiers and predicates, then their position borders on incoherence.[28]

This is related to a fact well-recognized in the metaontological literature—if one thinks that ontological debates, say over the existence of material objects like tables and chairs, are *merely verbal*, then one can't think that only the predicates "chair" and "table" vary in meaning.[29] That wouldn't be enough. If there is a single chair in a room, one party in the debate will say there are only *n* things in the room, all simples that happen to be "arranged chairwise".[30] The other party will say that there are these *n* simples, plus one chair that is not itself identical to any of the simples. In other words, the disputing parties also disagree about the truth conditions of the sentence "the number of objects in the room is *n*", so they also disagree about the contribution that quantifiers and identity make to truth conditions. Likewise if one tries to make only the identity predicate vary in meaning while leaving the quantifiers fixed.[31]

Given **Quantifier Metadeflationism**, we can make perfect sense of how it is that existence claims can be trivial or true by convention. The Peano rules prove the sentence "$\exists x \mathrm{N}x$" and so we have an inferentialist argument that the sentence is (necessarily) true. Additionally, the expression "$\exists$" in our language is an unrestricted existential quantifier expression, since it obeys the standard quantifier rules. As such, the real and literal existence of numbers is a byproduct of our linguistic conventions. A similar argument goes for reference as well: the Peano rules prove "$\exists x(x = 0)$", and since "0"

plays the role of a name, and "∃" the role of the existential quantifier (and "=" the role of identity, and "*x*" the role of a variable) then we are committed to saying that "0" refers. If we accept role priority then both existence and reference can be fully explained in this fashion against the backdrop of inferentialism. There is no remaining mystery about this, given a proper view of metaontology.

Yet there are lingering worries and concerns about metadeflationism itself. Mainly worries about how this approach deals with things like the ontological argument for the existence of God and the possibility of alternative practices that *deny* the existence of numbers. If no coherent sense can be made of the surrounding metaontological issues, then metadeflationism about existence and reference should be viewed with suspicion. I will address these points in stages, by first detailing metadeflationism's pluralist entailments, then defending the view against some objections, and finally returning to address the Kantian worries about the ontological argument from 8.III.

III Quantifier Pluralisms

One of the most interesting features of quantifier metadeflationism—one that has been embraced by Putnam and Hirsch but from which others have fled in terror—is that it leads to an interesting brand of quantifier pluralism. The pluralism is of a very particular type, and it leads directly to a form of ontological pluralism that some have found troubling, so it needs to be distinguished from other, adjacent *pluralisms* with which it might be confused.

One type of quantificational pluralism involves accepting many different *types* of quantifiers: "most", "infinitely many", "uncountably many", "more than 17", and so on. Presumably almost everyone is a *type pluralist* of this kind, but the controversial kind of pluralism concerns only the familiar existential and universal quantifiers, not exotic types of quantification.

Another brand of pluralism is *logical* or *role pluralism*, which involves there being languages with somewhat different rules for the same type of quantifier expressions. I argued for and characterized

this type of pluralism, at length and on inferentialist grounds, in 5.III. As was noted there, at some point, two inferential roles can differ so drastically that there is no real sense in which both are versions of the same quantifier role. Whether an expression is quantifier-like is a matter of relevant similarity of inferential role, with all of the indeterminacy that notion involves. There is no completely firm and context-insensitive line. While I fully endorse this type of pluralism, the kind of pluralism that falls out of quantifier metadeflationism arises even when the logical rules for the quantifiers are held fixed and assumed to be the standard, classical rules, as they were in the previous section.

Leaving these other kinds of quantifier pluralism aside, the ontologically relevant kind of pluralism involves the possibility of different languages each having (for example) unrestricted existential quantifier expressions that *do not mean the same thing* and so cannot be translated into each other. This isn't at all trivial: it is not the mere claim that the words "there is" or "exists" could have meant something different in some alternative language. That much is entirely trivial and is secured by the conventionality of linguistic signs (5.III). Nor is it the claim from above that the word "exists" or the like could have meant something *similar* while obeying slightly different logical rules of inference. Instead the claim is that the logical rules can be exactly the same for two expressions in different languages, and so both expressions can, by **Quantifier Metadeflationism**, be unrestricted existential quantifier expressions, *while still differing in meaning*.

We can express this kind of quantifier pluralism more explicitly as follows:

> **Quantifier Pluralism:** There are languages L and K with expressions Q_L and Q_K, respectively, such that (1) Q_L and Q_K are both unrestricted quantifier expressions of the same type and (2) Q_L and Q_K mean different things, that is, Q_L and Q_K cannot correctly be translated into each other.

It is important that both quantifiers be unrestricted, since otherwise one could be seen as a restriction of the other, at least in many

contexts. But everybody admits that some quantifiers are restrictions of others. The pluralism here is more radical than that.

The exact content of **Quantifier Pluralism** depends on what is required for there to be languages meeting this condition. Obviously the interesting claim in the vicinity is not about which languages there *actually are*, since that depends on various haphazard historical contingencies. It is instead about whether languages meeting this condition are *possible* (though the non-modal reading and the modal reading collapse if "languages" are understood as abstract objects). As such, I will understand **Quantifier Pluralism** *modally*, in terms of physical modality. It is true if there *could be* languages with unrestricted quantifiers that differ in meaning.

Plausible versions of **Quantifier Metadeflationism** entail **Quantifier Pluralism**. In my previous published discussion, I talked as if absolutely every type of top-down metasemantics entailed **Quantifier Pluralism**, via Quantifier Role Priority, but this was because I there dismissed top-down approaches that didn't see truth conditions as aligning very closely with language use.[32] But though they aren't very plausible, such approaches are certainly possible: imagine a top-down view entailing that the relationship between use and sentence meaning was so slack that *every* pattern of use determined the very same sentence meanings, even modally. Such a view would not lead to **Quantifier Pluralism**, even assuming **Quantifier Metadeflationism**, because it wouldn't even lead to pluralism about languages.[33]

However, most-top down approaches allow, trivially, that there can be different patterns of language use that will lead to different patterns of sentential semantics facts. On these views, this language pluralism will, together with **Quantifier Role Priority**, entail **Quantifier Pluralism**. As was illustrated in a different context in chapter 5, my own unrestricted inferentialist view allows that any rules of use that can possibly be followed can be meaning determining and so can found alternative linguistic practices. As such, my brand of quantifier metadeflationism also leads to very widespread pluralism concerning quantifier languages (languages with quantifiers).

Here's an illustration of this pluralism, based on an example used by both Putnam and Hirsch in many places. If a Martian community consisting of compositional universalists—those who think that for any two disconnected objects there is a third object, having both as parts—follows a rule like:

$$(\mathit{TroutTurkey}) \qquad \frac{\exists x\,\mathrm{Trout}(x) \quad \exists y\,\mathrm{Turkey}(y)}{\exists z\,\mathrm{TroutTurkey}(z)}$$

Then, by inferentialism, this and related rules will be necessarily truth- preserving in their language.[34] Of course, this rule is neither non-empirical nor topic neutral, but we can imagine this being an instance of general schematic rule that is plausibly non-empirical and topic neutral, and so that "∃" is still a *logical* symbol in Martianese (3.III). And the addition of such rules doesn't involve altering the (∃*I*) and (∃*E*) rules (section II), so "∃" will be still an unrestricted existential quantifier, in Martianese.

Yet in English, based on our linguistic rules, it seems clear that there is no such thing as a troutturkey. This point can be made even clearer by assuming that the compositional rules of Martianese entail that a troutturkey has a part that is a turkey and a part that is a trout. So the rule (*TroutTurkey*) is not valid in English.[35] Given this, the Martians will disagree with us (in a sense) about the existential facts. When there is exactly one trout and exactly one turkey in an otherwise empty room, we will truly claim that there are only two objects in the room, while the Martians will truly claim that there are three. This seems mysterious, but it is just an instance of **Quantifier Pluralism**—despite agreeing about the logic of "∃", the Martians have a slightly different concept of "existence" than we do. The same goes for their related concepts of "object" and "reference". Both our claim and the Martian claim can be true, despite seeming to contradict each other. In other words: against the backdrop of a Quinean approach to existence, **Quantifier Pluralism** entails a type of *ontological pluralism*.[36]

Despite this, we need to be very careful when arguing that ontological disputes between philosophers *in our own language*

provide instances of ontological pluralism. Putnam's discussions were often quite cavalier about this, but Hirsch has offered sophisticated arguments aiming to show that many actual ontological disputes are merely verbal. My own view is that philosophical ontology is built on fundamental and mistaken theoretical assumptions, so that rather than speaking a coherent language, ontologists often fall into massive and widespread foundational error. But here I'm not concerned to push this error-theoretic account of philosophical ontology.[37] Instead I want to distinguish the kind of **Quantifier Pluralism** discussed thus far from some variants of pluralism that both Putnam and Hirsch tend to focus on.

Putnam and Hirsch quickly go beyond the mere claim of **Quantifier Pluralism** and move to an egalitarian version of pluralism, according to which some of the distinct quantifier meanings are *equally good from an objective point of view*. In fact, they both build this condition into their basic terminology, and thus have no term solely for either the metasemantic view of quantifiers on an inegalitarian brand of pluralism. To illustrate, Putnam says of *conceptual relativity*:

> …conceptual relativity always involves descriptions which are cognitively equivalent ... but which are incompatible if taken at face value…[38]

Putnam's notion of cognitive equivalence is discussed in an earlier paper of his.[39] There he says that two theories are cognitively equivalent just in case they are mutually relatively interpretable in a way that preserves goodness of scientific explanation.[40] Though Putnam doesn't discuss it, this inter-theory notion of equivalence can be used to define an inter-language notion of equivalence by saying that two languages are equivalent just in case for every theory formulable in one language there is a cognitively equivalent theory formulable in the other and vice versa (Putnam's account of equivalence was briefly mentioned in 5.IV) Similarly, Hirsch says of *quantifier variance*:

> …truth-conditionally equivalent ontological languages are of equal meta- physical merit. That is the doctrine of quantifier variance. The doctrine says that there is no uniquely best ontological language with which to describe the world.[41]

Where two languages are truth-conditionally equivalent when they can express all and only the same coarse-grained truth-conditions, which Hirsch models as functions from contexts of utterance to sets of possible worlds.[42] In 5.IV I discussed equivalence between distinct logics in some detail, and many of the remarks made there are also relevant here.

Both conceptual relativity and quantifier variance go beyond mere quantifier pluralism to a type of egalitarian quantifier pluralism. For both Putnam and Hirsch, this egalitarianism is built upon the kind of expressive equivalence just discussed, where languages with non-synonymous unrestricted quantifiers are equally good in that both are able to—in their own way—express all and only the same facts. More generally, we can state this kind of egalitarian quantifier pluralism as follows:

> **Egalitarian Quantifier Pluralism**: There are languages L and K with expressions Q_L and Q_K, respectively, such that (1) Q_L and Q_K are both unrestricted quantifier expressions of the same type; (2) Q_L and Q_K mean different things, that is, Q_L and Q_K cannot correctly be translated into each other; and (3) the languages L and K are objectively equivalent.

Using this terminology, Putnam and Hirsch's respective terminology can be correlated with my own basically as follows:

> **Conceptual Relativity** = Egalitarian Quantifier Pluralism understood according to Putnam's account of equivalence (as applied to languages)
>
> **Quantifier Variance** = Egalitarian Quantifier Pluralism understood according to Hirsch's account of equivalence

This terminology has been further clarified in recent joint work of mine with Eli Hirsch.[43] There, basically what I have called “quantifier pluralism” is called “modest quantifier variance”, and the egalitarian version of quantifier pluralism, understood using truth-conditional equivalence, is called “strong quantifier variance”. So in my preferred terminology, which I think irons out most of the bugs and avoids the tired and misleading associations of both “deflationism” and “pluralism”, Quantifier Metadeflationism is a metasemantic view that leads to Modest Quantifier Variance, a type of pluralism about quantifier languages, and plausible views about equivalence lead from here to Strong Quantifier Variance, an egalitarian version of quantifier language pluralism.

Other philosophers have embraced versions of quantifier pluralism with a different, more metaphysically heavyweight notion of *equivalence*. Metaphysicians such as Cian Dorr and Ted Sider might be best interpreted as taking this option, at least in some of their work.[44] Here I won’t be much concerned with equivalence and egalitarian pluralism. Like most historical conventionalists, and as already mentioned at a couple of points in part II, I reject all heavyweight metaphysical notions on naturalistic grounds (1.IV).[45]

With some notion of equivalence in hand, we can formulate and entertain general claims about the metaphysical ordering of languages, with or without quantifiers. Perhaps the ordering is such that a single quantificational language is objectively the best? Or perhaps, as Hirsch seems to assume in the above quote (or thinks follows directly from his notion of equivalence), no single language is best, but instead a number of languages are of equal metaphysical worth. Analogously to my discussion of logical equivalence in 5.IV, I won’t be assessing or making such equivalence claims here, since it isn’t required for my purposes. Here I am only concerned with providing a theory of our mathematical language, which is itself a quantificational language, and not with the metaphysical or objective worth of our language as compared with the many other possible languages that might be imagined. But once again, for the record: like Putnam, Hirsch, Carnap, and most historical conventionalists, I don’t think that good sense can be made of metaphysically robust notions of equivalence like those suggested Dorr and Sider.[46]

Here I won't be using quantifier variance to argue for a deflationary position concerning ontological disputes or debates. My concern here is only that, when quantifier metadeflationism and thus modest quantifier variance are endorsed, the worries about trivial existence claims from the last chapter start to seem rather quaint. I will return to illustrating this directly, in section V, but first I need to address some of the philosophical worries about modest quantifier variance that readers are probably wondering about.

IV Antirealism, Collapse, and Other Antagonisms

The views of both Putnam and Hirsch have come under heavy fire in the philosophical literature. Some criticisms concern idiosyncratic aspects of their views, while others target the anti-metaphysical applications that they have (respectively) made of their views. A couple of criticisms attack modest quantifier variance directly and so need to be addressed here.[47]

The first is based on a simple misunderstanding. It claims that deflationism is an *anti-realist* or *ontologically constructivist* view, entailing that we make or create the objects that we talk about. Unfortunately a major reason for the close association of quantifier pluralism with anti-realism is that many of Hilary Putnam's early discussions of conceptual relativity bundled quantifier pluralism together with general anti-realist ideas, such as (something like) universal constructivism and an *epistemic* theory of truth.[48] Putnam later recanted these anti-realist views while still holding to conceptual relativity, but the damage had already been done.[49] By contrast, Hirsch's discussions have always made it clear that quantifier variance has nothing to do with anti-realism.[50]

The misunderstanding behind this objection is based on a failure to realize that quantifier pluralism concerns our *concept* of existence, not the things that exist (in this way it is somewhat related to the contingency objection to conventionalism discussed in 7.I). If we change our language so that the sentence "numbers exist" is true when it wasn't before, we haven't, thereby, *created* numbers. If we have created anything in this transition, it is new concepts of

existence and number. It is akin to a use/mention confusion to think otherwise. And this goes for reference as well as existence. Modest quantifier variantists can coherently maintain that humans did not create numbers or other abstract objects.

Of course, this isn't to say that absolutely all worries about the relationship between our practices and objects like numbers are cleared up by pointing out that simplistic attempts to convict the quantifier variantist of anti-realism are mistaken. Accordingly, I'll say a bit more about the metaphysical picture of numbers that my approach leaves us with in the final section of this chapter.

Before that, let's turn to the most important argument against deflationism and pluralism: the collapse argument. I have already addressed this argument several times in print.[51] The argument is based on one given, in a different context, by J.H. Harris.[52] It has been directed against quantifier variance by both Marcus Rossberg and the Neo-Fregeans.[53] I'll illustrate the standard collapse argument using the example of the Martian language from the last section before discussing a more sophisticated variant of the argument.

In the Martian language, *M*, in which composition is unrestricted, the sentence "$\exists z$ TroutTurkey(z)" is, as we have seen, true, but it is false in English. However, given the open-endedness of our English quantifier rules, we can give the following argument in English, with the Martian quantifier written "$\exists^M$" (for ease of reading, I'll again use the natural deduction format used in chapter 6):

1. $\exists^M x$ TroutTurkey(x) {1} *assumption*
2. TroutTurkey(α) {2} *assumption*
3. $\exists x$ TroutTurkey(x) {2} $\exists I$(2)
4. $\exists x$ TroutTurkey(x) {1} $\exists^M E$(1,3) *discharge*(2)

Obviously, this is a disaster. It seems to show that *shared inferential role*, as required by metadeflationism, undercuts the very possibility of pluralism. We seem to have shown either that it is true in English that there are troutturkeys, or that English is inconsistent. Neither is an appealing result.

Fortunately for variance, this argument rests upon two crucial mistakes. The first involves missing the point that accepting, in

English, the pluralist claim that the sentence "$\exists z$ TroutTurkey(z)" is true in Martianese is not to accept, in English, that there are troutturkeys. A simple illustration of the mistake here is to imagine a language where the sentence "Rudolf Carnap is taller than Wilt Chamberlain" expresses that 2 + 2 = 4 and so is true. We can cheerfully admit that this sentence is true in *that* language without thereby admitting that Carnap is taller than Wilt. This is, once again, a variant of The Translation Mistake that I so often warned against in chapter 5 and throughout part II. With this mistake cleared up, there is no justification, in English, for premise 1 of the collapse argument. At best, there is justification for the premise that "$\exists z$ TroutTurkey(z)" is true in Martianese, but we cannot simply disquote, we must instead translate this Martian sentence into our own, if we can. When we so translate—say by mapping it to "$\exists x$ Trout(x) $\wedge \exists y$ Turkey(y)"—nothing like the collapse argument will be forthcoming.

The second crucial mistake in the argument is related to the first. Even if we imagine expanding our language to include the sentence "$\exists^{M} x$ TroutTurkey(x)" and appealing to open-endedness to continue to apply *our* rules of inference, that isn't enough for the argument, since the argument uses, at line 4, the *Martian* rule of existential elimination, *in English*. But this, of course, is not a rule of English, it is a rule for the Martian quantifier and English doesn't include Martianese rules. It is no surprise at all that if you add another unrestricted quantifier to the language, it collapses with your own.[54] Adding new logical expressions and rules changes English into something else, as inferentialists will clearly recognize, so nothing at all *about English* (or about English's notion of existence) is established by this version of the collapse argument. The Big Picture situation is this: an *intra*-language collapse argument would be sound but is irrelevant to quantifier variance; the *inter*-language version we have been considering would target variance but is unsound.

What the collapse argument nicely illustrates is that, just as **Quantifier Metaeflationism** led to **Quantifier Pluralism** (modest quantifier variance), **Predicate Metadeflationism** and Name **Metadeflationism** lead to **Predicate Pluralism** and **Name Pluralism**, respectively. In fact, as I stressed above, all of these views go together.[55] Different languages may well have different

concepts of what it is (semantically) to be a quantifier, a predicate, and a name. This is why we can't simply import the expressions of another language into our own without translating, nor import their rules thinking it will leave all as it is. This point stands *even given open-endedness*. Open-endedness concerns names and predicates that "go with" our quantifier. Having an open-ended quantifier doesn't foreclose on the possibility that adding some name-like and predicate-expressions, together with rules for their deployment, could end up altering what we mean by "existence" and the like.

These points completely undercut the standard collapse argument. But Cian Dorr has recently offered a sophisticated update of the collapse argument against variance.[56] Dorr's paper is long, dense, and relentlessly technical; in effect though, the paper concerns versions of Harris-style "collapse theorems" that are overtly semantic. Dorr's collapse theorems have the form: if two expressions α, β each obey rules R, then, where $[\![\phi]\!]$ is the semantic value of ϕ, then $[\![\alpha]\!]$ is equivalent to (or identical with) $[\![\beta]\!]$. If α and β are, for example, existential quantifier expressions, then the result seems incompatible with quantifier pluralism, since it seems to show that quantifiers in different languages mean the same thing.

Dorr's approach involves formulating the quantifier rules in terms of a notion of *concept entailment* that isn't understood in terms of objectual quantification or any equivalent notion. Concepts here are the semantic values of predicates, so Dorr's notion of concept entailment allows us to apply entailment to open sentences. With the relevant notion of concept entailment explained and various complications handled, Dorr is able to formulate the following rules, which I'll call the *D*(orr)-rules for the existential quantifier:[57]

> $\exists$ is an existential quantifier in L just in case for any formulae ϕ and ψ, variable ξ not free in ψ, and non-repeating sequence of variables σ including all of the free variables in ϕ and ψ, $\exists$ obeys both (*i*) *D*-introduction: ${}^{\sigma}[\![\phi]\!]_{L}$ concept-entails ${}^{\sigma}[\![\exists\xi\phi]\!]_{L}$; and (*ii*) *D*-elimination: if ${}^{\sigma}[\![\phi]\!]_{L}$ concept-entails ${}^{\sigma}[\![\psi]\!]_{L}$, then ${}^{\sigma}[\![\phi]\!]_{L}$ concept-entails ${}^{\sigma}[\![\exists\xi\psi]\!]_{L}$[58]

He then proves that expressions that obey the *D*-rules have equivalent (or identical) semantic values.[59] This *seems* to undercut pluralism, but does it really do any better than standard collapse arguments?

Dorr's theorems require that the same space of concepts (semantic values of predicates) be available in each language. He defends this by claiming that the only plausible way to deny this involves not pluralism about concepts (essentially **Predicate Pluralism**, in my terminology) but rather the denial that other languages have anything like concepts at all. By parity of reasoning the apparent quantifier expressions of the alien language are, likewise, meaningless.[60]

But this claim has been consistently and roundly rejected by both Hirsch and myself, and it seems to be doing almost all of the *philosophical* work in Dorr's argument, so it is odd that it receives so little discussion.[61] Despite Dorr's protests, variantists can respond to Dorr's collapse argument in a well-motivated fashion by denying that the putative quantifier expressions of each quantificational language obey the full *D*-rules, in the sense required for collapse.[62] Variantists will instead see the very statement of the *D*-rules as varying in meaning in different quantificational languages. **Quantifier Role Priority** does not require anything as robust as the *D*-rules, but only the standard syntactic-inferential rules, so this response is not ad hoc. In fact, we have good evidence that this response involves no special pleading, since long before Dorr's paper appeared, Hirsch stressed that *this* was all that he meant by a "quantifier", and that if others with more robust requirements wanted to talk in terms of quantifier *elimination* instead of *variance*, that was of little matter to him.[63]

So neither form of the collapse argument undermines quantifier pluralism. The argument that actually targets quantifier pluralism is fallacious, and more sophisticated arguments are, even if not fallacious, aimed at a straw man (or, at least, their conclusions are irrelevant to modest quantifier variance). There are several other anti-variance arguments in the literature, but I think they can all be answered in a similar fashion.[64]

With quantifier metadeflationism and modest quantifier variance explained, clarified, and defended, we can now return to applying it to

mathematical ontology, explaining, at last, why worries about trivial ontology generalizing—to God, Santa Claus, and beyond—are easily met.

V Answering Kant's Challenge

In 8.III I discussed Kant's influential argument that ontological claims cannot be trivially true. Yet above I've shown that trivial ontological realism follows from conventionalism, given quantifier metadeflationism and its attendant pluralism. I think that this metaontological view is correct. I also think it dovetails neatly with inferentialism and conventionalism. But even granting this, various worries arise about exactly *why* Kant's argument fails. It is time to give an explanation.

Kant's point was that we could always reject the existence of Fs without contradiction. This is wrong because sometimes the meaning-determining rules that implicitly define "F" and related terms, themselves entail, as a byproduct, that Fs exist. When acceptance of an existence claim is built into the possession conditions of a concept, then ontological realism about F is a requirement of having the very concept of an F. This, I think, is basically what happens with numbers and other pure abstract objects. The sentence asserting the existence of numbers ("$\exists x \mathrm{N}x$") is true in virtue of the conventions of our arithmetical language together with the inferential properties of "$\exists$" and a standard, Quinean account of existence. So numbers exist, by convention. We do not need to check the world to see whether numbers exist. Our conceptual understanding, as rooted in the basic rules of our arithmetical language, suffices, both metaphysically and epistemologically, for establishing the real and literal existence of numbers. Our concept of existence is our concept of unrestricted existential quantification; there is no deeper notion of something *really* existing beyond the truth of a sentence like "numbers exist", in our language. All of this is vindicated and explained by quantifier metadeflationism, against an inferentialist backdrop.

The approach neatly sidesteps many metaphysical puzzles about ontology. But critics might still wonder why this approach doesn't

allow us to just as easily prove the existence of God?[65] Why, they ask, can't we inferentialists just add the following inference rule to our language:

$$(God) \quad \frac{}{\exists x(x = \text{God})}$$

There is no in principle barrier to our following this rule, so if we added this as a basic rule of our language, unrestricted inferentialists like me would have to admit that it is valid, and thus that the rule's conclusion is true. But, the objection continues, this conclusion entails that God exists and that the term "God" refers (to God)! So isn't this just to admit the validity of a simplified ontological argument for the existence of God? What has gone wrong here?

The simplest response to this is that *nothing* has gone wrong. Or at least nothing with inferentialism. The result the critic is pointing to is exactly what we should expect and want. The problem is only that the critic is badly misdescribing the result. The possibility of a linguistic practice like ours except with the additional rule (*God*) added, *does not show that God exists*. The term "God" means something in that language that it does not mean in our language. Even though the sentence "God exists" is true in that language, why should we think the same sentence is true in our language, on *that* basis? To simply import the claims of the extended language into our own is to, yet again, fall victim to **The Translation Mistake**.

Think of the matter dynamically: if we added the rule (*God*) to our language, we would thereby change the usage of the word "God" in a way that changes its meaning, so that the truth of "$\exists x(x$ = God)" in that extended language would no longer establish the existence of God, as we presently use the term "God". This is because as a matter of fact, we do not currently use the term "God" according to the rule (*God*). When we wonder about the existence of God in English, we are concerned with the sense of "God" in English, not about the sense of some other meaning for the same (syntactically individuated) term in an alternative language. So the objection fails for the same reason that naming your firstborn child "God" is

theologically irrelevant (it is also probably a bad idea for other reasons too).

This is why the ontological argument is not valid, in our current language. We simply do not use the term “God” or related terms so that it is a byproduct of our having the concept of God that God exists. Relatedly, theologians are not using “God” in the same way that literary scholars use a term like “Sherlock Holmes”; instead they are aiming to refer to a being with immense, perhaps even unlimited, causal powers.[66] Causal powers that could, in principle, make a difference in the physical world. To interpret them as speaking truly about a causally inert fictional character is a distortion of their practice, and inferentialists can fully admit this point without running into problems just by looking closely at the actual usage of the term “God” in our language.

The same point explains why it is incorrect to treat those who believe in Santa Claus as referring to an abstract object whose existence is trivially guaranteed, rather than as failing to refer to a man with a beard that lives at the North Pole. True believers have linguistic dispositions that are incompatible with their following a rule that *trivially* establishes the existence of Santa Claus. In this way, numbers and Santa Claus are very different. Even for true believers in Santa, Santa’s existence is a robust and potentially causally relevant fact about the physical world. This is not so for numbers, and the difference comes out in the differing rules followed for the respective terms, and is rooted in the linguistic dispositions of speakers with respect to these terms. This is why the trivial existence of mathematical objects doesn’t generalize.

Recognizing all of this, we can now see that despite its prima facie plausibility, Kant’s point is based on a subtle equivocation. It is simply not true that we can always reject the existence of Fs coherently. Sometimes having the concept of an F requires following rules that force the acceptance of a sentence expressing the existence of Fs. What we can always do is *reject the very concept of an* F. Elsewhere I have called this “conceptual rejection”.[67] There I gave an amusing historical illustration of the idea, which is worth repeating here. It concerns one of Oscar Wilde’s trials. On April 3, 1895, defense

attorney Edward Carson asked Wilde about John Francis Bloxam's story "The Priest and the Acolyte":

> CARSON: Do you think the story blasphemous?
> WILDE: I think it violated every artistic canon of beauty.
> CARSON: I wish to know whether you thought the story blasphemous?
> WILDE: The story filled me with disgust. The end was wrong.
> CARSON: Answer the question, sir. Did you or did you not consider the story blasphemous?
> WILDE: I thought it disgusting.
> CARSON: I am satisfied with that. You know that when the priest in the story administers poison to the boy, he uses the words of the sacrament of the Church of England?
> WILDE: That I entirely forgot.
> CARSON: Do you consider that blasphemous?
> WILDE: I think it is horrible. "Blasphemous" is not a word of mine.[68]

Obviously, when he rejected the *word* "blasphemous", Wilde was really rejecting the very concept of blasphemy (it isn't like he accepted the concept but preferred it to be called "blah-hemy"). Wilde's attitude toward blasphemy was that of conceptual rejection—he rejected the very concept.

Let's apply this to arithmetic. We have seen that you can't have the concept of number while clearheadedly rejecting the existence of numbers, since having the concept of number involves, at least, following the Peano rules. But what you can do is *reject the concept of number altogether*, and you can do so without any contradiction or incoherence. So from a certain point of view, Kant was right. The Universe will never take you by the throat and force you to accept an existence claim. You can always instead reject the conceptual scheme involved. As such, there is no mystery about how and why numbers can, in a sense, exist by convention, without being creations of the human mind. But what then, exactly, is the conventionalist picture of the metaphysics of mathematical objects?

VI The Metaphysical Nature of Mathematical Objects

Obviously, given everything I have said thus far—including part II's discussion of the metaphysics of logic—questions about the metaphysics of numbers shouldn't be taken in a *heavyweight* fashion. But there are broadly metaphysical questions about numbers and mathematical objects that arise even for conventionalists and other scientifically minded philosophers.

Frege infamously rejected the Hume's principle-based, Neo-Fregean-style implicit definition approach to numbers because of the so-called Julius Caesar problem.[69] Frege seemed to think that our *definition* of "number" should entail, among other things, that Julius Caesar is not a number. Apparently the Caesar problem was the central issue that pushed him to adopt his later and ultimately inconsistent approach to arithmetic using extensions and Basic Law V.[70] But an appeal to extensions, even if consistent, would have been no help with the Caesar problem.[71] What does conventionalism have to say about this issue?

One point worth noting is that the modal status of the Peano rules directly rules out any identification of numbers with contingently existing objects, such as Julius Caesar. In a language where necessity claims are explicitly made using the "□" symbol, we can easily prove, using the Peano rules and the logic of necessity, the sentence "$\Box\exists x(x = \mathrm{n})$", for any natural number n. So given the evident fact that we can presumably establish that Julius Caesar does not exist necessarily, Caesar isn't identical to any number n.[72] At the very least, we can show that given the modal status of the Peano rules and the existence of numbers, anything that doesn't have the same modal status is not a number, including Julius Caesar. And since the modal status of numbers is that they are necessary in the strongest possible sense (see the discussion in part II—especially 3.I's discussion of strong validity) virtually everything else in our ontology can easily be shown not to be a number. In effect, this much follows from the (implicit) *definition* of "number" provided by the PA-rules. So even if the Caesar problem is taken seriously, versions of the problem

that identify necessarily existing mathematical objects with contingently existing objects are quite easily answered.

What about *other* types of abstract objects, could they be numbers? It is well known that there are many different ways of interpreting number theory in set theory. These interpretations identify particular numbers with particular sets, but different interpretations disagree with each other about which sets are identified with particular numbers. This point was famously discussed in an influential paper by Paul Benacerraf.[73] Benacerraf used the point to popularize *structuralist* approaches to mathematical objects, according to which our mathematical theories talk only about abstract structures rather than particular objects. According to this way of thinking, "numbers" are really, in some fashion or other, just places in the overall number structure.

Can conventionalists be structuralists? Should they be? Pure structuralists say that there are no particular mathematical objects at all, just structures. An obvious issue with this extreme version of the view is that our mathematical theories will be empty if there aren't enough objects to fill out the various mathematical structures. To get around this type of worry, some pure structuralists endorse a strong metaphysical realism about structures that sits ill with conventionalism.[74] Alternatively, some have moved to *modal* structuralism, which doesn't claim that objects in the right structural relations actually exist, only that it is possible for the relevant structures to be filled out by objects.[75] Once again this goes far, far beyond anything that conventionalists need to endorse. For what could possibly justify us, independently of mathematics, in thinking there can be enough non-mathematical objects to answer the needs of *set theory*? We would need not just infinitely many objects, nor just uncountably many objects, but so many objects that they could no longer form a set.

Leaving aside pure structuralism, what about mixed structuralism, according to which only *some* mathematical theories are structural? Perhaps the most attractive mixed approach takes set theory to be about sets, but then sees all other branches of mathematics as purely structural, with structures identified with the "intended" set-theoretic models of theories. The relationship between conventionalism and

this kind of mixed structuralism is more equivocal. Of course, certain parts of mathematics are overtly structural—group theory, for example.[76] In these branches, the axioms don't categorically describe any class of structures within set theory. And these so-called *algebraic* branches of mathematics are typically studied in a way that makes it clear that they do not aim to be about particular collections of objects. They are more of a motley, and not much is lost by treating them in wholly structural terms. Conventionalists can and should admit this.

With arithmetic though, our actual practice *seems* to be *about* a *particular* collection of *objects*: the numbers. There is nothing in the practice that directly suggests a structural construal. But we shouldn't read too much into this fact. Even though our practice treats numbers as sui generis objects and allows us to differentiate them from sets via Leibniz's law—the number 3 has no members, so the number 3 isn't a (non-empty) set—if we altered our practice to treat number theory in a structural fashion, nothing of any importance would be lost. This is a claim about the uses of arithmetic, not a historical claim. Historically it might have been psychologically important for arithmetic to be understood non-structurally, perhaps so our ancestors could get on with the business of using numbers without worrying that they might run out of them.

So numbers aren't structures, but that isn't a particularly deep or important metaphysical fact about arithmetic, it is instead just another reflection of our linguistic conventions. It is a shadow cast by our practice of naming and talking about particular objects rather than places in a structure when doing arithmetic.

For conventionalists, there is not much else that is worth saying about the metaphysics of numbers and other mathematical objects, qua objects. Over and over again, the conventionalist's metaontology turns away metaphysical puzzles. Numbers and other mathematical objects are, like the logical truths, shadows of our syntactic rules. Mere reflections of our linguistic conventions, real in their own right, but with no deeper essence or nature.

≈≈≈≈≈≈≈≈

KANT'S CHALLENGE FOR trivial ontological realism is met by conventionalists. Numbers and other abstract objects exist. An inferentialist approach to arithmetic involving only the Peano rules already secures this much. In understanding arithmetic, we follow the Peano rules, and it is a straightforward consequence of these rules that there are numbers, so numbers exist. Conventionalism is not only compatible with ontological realism about numbers and other mathematical objects, *it entails it*, and quantifier metadeflationism answers the worries about how this could possibly be so without leaving any residual metaphysical mystery. But a deeper challenge for conventionalism still remains, even at the level of basic arithmetic, for the Peano rules fail to secure an intuitively satisfying theory of arithmetical truth, and it seems impossible, in principle, for any other rules we could follow to do any better.

---

1 See Densmore (2002).

2 In Quine (1948).

3 I say this instead of simply saying that the sentence "∃yXy" is true in the language since complications arise if, for example, the language doesn't contain a predicate for Xs—presumably a language could have included many true claims about squid long before a predicate like "squid" was introduced.

4 See Quine (1970*b*) for discussion of most of these topics. It isn't obvious that the total package of Quinean views is coherent; for instance, the philosophical line pushed in his (1968), or even the elimination of names argument in his (1948), when combined with the elimination of variables demonstrated in his (1967), seems to undercut the very idea that our theories have ontological commitments at all.

5 This is not say that Quine's approach is unanimously accepted, for a sampling of the resistance see Azzouni (2004), McGinn (2000), and Parsons (1980) and (1982). In terms of natural language, it seems clear that "natural numbers exist" can be read as either "*a* natural number exists" or as "*the* natural numbers exist"; Fine (2009) has argued, in effect, that Quineans can't handle the second reading, and so we need something like an existence predicate to account for natural language ontological commitment. I think this is the most powerful argument against a standard approach, but it can be handled on Quinean grounds by taking appearances seriously and using plural quantification (Boolos (1984), (1985)), see my (forthcoming *b*) for details.

6 From Quine (1958), among other places.

7 Quoted from Quine (1948).

8 Williamson (1990) makes a similar distinction between "one-level" and "two-level" criteria of identity.

9 See Rosen (1993) for some subtleties.

10 See Frege (1884) and the discussions of Dummett (1956) and Wright (1983), especially. Dummett's later discussions, such as in his (1991*b*), indicate a move away from this "thin" conception of reference.

11 This is not the heavyweight "ontological realism" of Sider (2009), but merely the denial of nominalism.

12 Boolos (2000) seems to think that we have intuitions above size that extend high into the infinite cardinal numbers. I deny this, but even if we did have such intuitions, I don't see why they should be given much weight (certainly not enough weight to overturn the deliverances of *set theory*!).

13 See Schiffer (1996) and Thomasson (2009), for example. Sider (2007) has, correctly in my view, claimed that these "easy ontology" views require supplementation with something like Hirsch's (2011) quantifier variance. I have discussed the interaction between views like these and the claims of philosophical ontologists in my (2016*c*).

14 For this debate see Eklund (2006), Hale (2007), Hawley (2007), Sider (2007), and Hale and Wright (2009).

15 See Putnam (2004), Hirsch (2011), and Carnap (1950). About whether Carnap's views should be classed with Putnam and Hirsch's, see Eklund (2009).

16 This is missed in the famous and influential discussion of Benacerraf (1973).

17 This terminology has recently become popular; see the introduction to Chalmers, Manley, and Wasserman (2009).

18 I first attempted this task in print in section 2 of my (2017*b*). Below I won't recapitulate that discussion, but will instead cover much of the same ground, and more besides, in a somewhat different way.

19 An alternative approach would be to account for a quantifier's being unrestricted, inferentially, by saying that an existential quantifier $\exists$ is $\exists_i$ unrestricted in L just in case for any formula ϕ and existential quantifier in L, $\exists_k$, $\exists_i x\phi(x) \vdash \exists\ x\phi(x)$; and a universal quantifier $\forall_i$ is unrestricted in L just in case for any formula ϕ and universal quantifier in L, $\forall_k$, $\forall_i x\phi(x) \vdash \forall_k x\phi(x)$. However, this approach is hostage to an independent notion of what it is for a quantifier to be "definable" in a language, so I will stick with the intuitive open-ended approach here. The notion of "expansion" and "open-endedness" will be discussed in much more detail in the next chapter.

20 As noted in a different context in 8.III, the standard method of definition here derives from Ramsey (1978), Carnap (1958), and Lewis (1970).

21 See Ramsey (1925).

22 See Dummett (1973*c*), Strawson (1974), Geach (1962), and (1975) Hale (1996) and (1994), and Brandom (2000). For recent criticism of some of these attempts and another stab at a solution, see Parisi (2017).

[23] These examples are similar to Dummett's examples on page 64 of his (1973*c*).

[24] For details see Shapiro (1991) or Enderton (2001).

[25] Parisi (2017), page 3.

[26] In Hale (1996) and (1994).

[27] See Hale and Wright (2009) and Hale (2007).

[28] In trying to walk this razor's edge and being driven further and further from the undiluted form of their metasemantic principles, I believe that Hale and Wright are going down something like the path Dummett already traveled between his (1956) and his (1991*b*).

[29] See Putnam (2004) and Hirsch (2002), see also Sider's discussion in his (2009) and (2011).

[30] The arranged Xwise locution comes from van Inwagen (1990).

[31] The argument for this is not in the published literature, to my knowledge, nor is the position itself. The position (called "identity variance") is found in Segal (2010) and the argument that it leads to quantifier variance was first given in Russell (2010). These items were presented at an APA meeting in December of 2010; I discussed these issues with both authors at around the same time. In unpublished work I give a full treatment.

[32] See pages 94–95 of my (2017*b*). Thanks to Callie Phillips for some discussion of this.

[33] He doesn't endorse this trivial view, but Horwich (1998*a*) is a "use" theorists who, in response to Kripke's (1982) anti-dispositionalist arguments, denies that meaning can be "read off of" language use. Extreme versions of this type of view will be top-down without entailing language pluralism. The view of Eklund (2006) and (2009) might also be read as endorsing quantifier metadeflationism but not quantifier pluralism (though I'm not convinced of this reading).

[34] By contrast, if an offshoot of our community seemed to start using this rule, it may be that they intended to defer to the broader community usage but were mistaken about this usage. External social facts have been recognized to be metasemantically important at least since Burge's (1979) influential discussion. Nothing in the pluralist approach here contradicts the importance of deference and social coordination in metasemantics.

[35] The term "TroutTurkey" comes from Lewis (1991), where it is used somewhat differently than I am using it here.

[36] This is ontological pluralism in basically the sense of Eklund (2009), which is different from the metaphysically heavyweight "ontological pluralism" endorsed by McDaniel (2010) and Turner (2010).

[37] For that, see my (2016*c*).

[38] Putnam (2004), page 48.

[39] Putnam (1983).

[40] Relative interpretability is introduced and explained in Tarski, Mostowski, and Robinson (1953); applying relative interpretability to the cases of interest to Putnam actually requires a generalization of the usual notion; see the appendix

to my (2015*c*). That appendix proves that plural mereological nihilism and atomistic classical extensional mereology are mutually relatively interpretable in this broader sense. In that appendix I was implicitly assuming, in applying the recursive clauses of the mapping defined there, that variables were to be relettered to avoid clashes. This is relevant in my mapping from N to U, since first-order and plural variables are both mapped to first-order variables. Kocurek (2018) has called this a "slight error" (page 181), but I think it more fair to call it a lack of explicitness in a fairly informal treatment.

41 Hirsch (2011), page xii.

42 As was noted in footnote 42 in chapter 5, Hirsch's truth conditions are essentially the "characters" of Kaplan (1989).

43 Hirsch and Warren (2019) and (forthcoming).

44 See Dorr (2005) and Sider (2009) and (2011).

45 See the discussions in my (2016*a*), (2016*b*), and (2016*c*).

46 In Dorr (2005) and Sider (2011).

47 See Hirsch and Warren (forthcoming) for brief responses to all of the major objections to quantifier variance.

48 See Putnam (1978) for his initial "internal realism" and his (1980) and (1981) for general constructivism.

49 Putnam's retraction is in his (1994*a*).

50 This is the central theme of Hirsch (2002). See also Searle (1995), pages 160–167, and Boghossian (2006), pages 32–41.

51 In my (2015*c*) and (2017) and in Hirsch and Warren (forthcoming).

52 In Harris (1982).

53 In Rossberg (2008) and Hale and Wright (2009).

54 Of course, we could alter the quantifier rules to mimic free logical rules so as to avoid this. Turner (2012) does this in defending a different version of ontological pluralism from the collapse argument. But the points I've made here show that there is no need for quantifier deflationists to revise our standard quantifier rules to block the collapse argument.

55 This point is explicitly made by Sider (2007) on behalf of quantifier variantists.

56 In Dorr (2014).

57 The D-rules are based on Dorr's "local open ∃-rules" (ibid, page 541), though as will be implicit below, I don't here mark any substantive distinction between these rules and Dorr's stronger and more complex "global open ∃-rules" (ibid, page 547). Dorr's approach seems related to the treatment of quantifiers in category theory—compare Dorr's discussion on pages 548–549 to section 9.5 of Awodey (2010), for example.

58 $\sigma[\![\ ]\!]$ " ϕ " represents the semantic value in L of ϕ relative to sequence σ; see page 540 of Dorr (2014) for a full explanation. L

59 See theorem 6 on pages 545–546, theorem 7 on page 547, and the arguments on pages 549–550.

60 See the discussions of pages 539 and 555 of Dorr (2014).

61 On page 531 Dorr also briefly claims that this type of response undercuts the initial motivation for positing quantifier variance, since it allows for a formal picture whereby each language's quantifier expressions mean the same thing but standardly operate on nonoverlapping spaces of concepts. But while this is formally possible (assuming that various complications are handled), I don't think variantists are likely to find it plausible, given their metasemantic commitments.

62 Dorr rejects this response, but does he think that the spirit of this response might be captured by positing variation in the meaning of "entails" (or, in my terminology: "concept-entails") across the various quantificational languages.

63 See page XIV of Hirsch (2011). Putnam (2004) has made similar claims; see also my (2017*b*).

64 In particular I'm thinking of the Tarskian argument from Eklund (2009) and Hawthorne (2006). This argument is answered, at length, and granting the assumptions of proponents of the argument, in Hirsch and Warren (2019). Criticisms continue to appear that fail to take quantifier variantists at their word about their own doctrines (for a recent example, see Finn and Bueno (2018)).

65 Field (1984*b*) raises this challenge for Wright's (1983) Neo-Fregean approach.

66 Schaffer (2009) claims that atheists use "God" to refer to a fictional character.

67 In section 3 of my (2016*b*). Conceptual rejection can be expressed by metalinguistic negation—see chapter 6 of Horn (1989). See also Brandom (2000) on slurs. Conceptual rejection is also used in an argument against robust concepts of consciousness in Cao and Warren (2019).

68 The transcript from which this excerpt is taken can be found online at http://law2.umkc.edu/faculty/projects/ftrials/wilde/Wildelibeltranowcross.html.

69 In Frege (1884).

70 In Frege (1893) and (1903).

71 Frege seemed to think otherwise, but his remarks about this are unsatisfying. In effect, he simply assumes that we have already solved the Caesar problem for extensions.

72 Surrounding issues quickly involve us with Quine's (1953) concerns about quantifying into modal contexts. See Smullyan (1948) for an early response to Quine.

73 Benacerraf (1965).

74 See Shapiro (1997) and Resnik (1997), for example.

75 See Hellman (1989), building upon Putnam (1967).

76 For details see any textbook in abstract algebra, for example, Dummit and Foote (1991).

10

Mathematical Determinacy

Unlike first-order logic, classical mathematics—even at the level of basic arithmetic—involves a computationally rich notion of truth. Accounting for the determinacy of basic mathematical truth was the second major obstacle for mathematical conventionalism mentioned in chapter 8. Denying the determinacy of arithmetical truth is extremely counterintuitive, but it is very difficult to see how arithmetical truth could, like logical truth, be a direct reflection of our linguistic conventions. Of course, a variant of this challenge faces *all* approaches to the philosophy of mathematics, but conventionalists don't have the option of offloading mathematical determinacy onto Platonic heaven and so must face the challenge in its starkest form, with no cheating (8.IV).

The metaontology developed in the previous chapter makes this last point particularly clear. Since the existence of numbers is *explained by* the truth of the sentence "$\exists x \mathrm{N}x$", which is itself explained by the rules of our arithmetical language, there are no external, explanatorily freestanding metaphysical facts about numbers for conventionalists to appeal to when explaining determinacy. The existence of mathematical objects follows as a trivial byproduct of our linguistic conventions, and so these objects have no substantial intrinsic nature or freestanding status. Platonic heaven is a myth. What then, can conventionalists do? This chapter tackles this question head on by giving a conventionalist-friendly account of arithmetical truth that will also (I hope) be appealing to many non-conventionalists.

I Determinacy, Pluralism, and Categoricity

In 8.IV I introduced the notion of *determinacy*. If mathematics is determinate, then there is a fact of the matter about every non-vague mathematical question. Obviously, this claim could be read trivially, so that a claim's being non-vague is cognitively equivalent to there being a fact of the matter about the claim. But I mean "vagueness" to apply only to terms like "small" and "large" which are indeterminate in the exact same manner that predicates like "bald" and "heap" are indeterminate.[1] Advocates of mathematical determinacy are not claiming that vague arithmetical claims like "97 is a small number" are determinate. Instead they are claiming that standard arithmetical claims, like Goldbach's conjecture, have determinate truth values, even if those truth values are forever unknown to us.

If we don't count sentences containing vague predicates as "mathematical", then the claim of determinacy can be put as follows:

> **Mathematical Determinacy**: All mathematical sentences have determinate truth values, that is, all mathematical sentences are either determinately true or determinately false.

This is a very strong claim. I doubt that it's true. But I do think that *arithmetic*, at least, is determinate:

> **Arithmetical Determinacy**: All arithmetical sentences have determinate truth values, that is, all arithmetical sentences are either determinately true or determinately false.

There is tremendous intuitive pressure in favor of the determinacy of arithmetic, but Church's thesis, together with standard mathematical results like the incompleteness theorems, makes it difficult to see how conventionalists could possibly account for determinacy even at the level of arithmetic (4.III and 8.IV).

Before trying to provide a conventionalist account of arithmetical determinacy, we need to be clear on the exact nature of determinacy. One relevant point is that determinacy is not obviously related to *pluralism*. In chapter 5 I gave a general account of pluralism that also applies to mathematics, particularly:

> **Mathematical Pluralism**: There are apparently incompatible languages of mathematics (that are all equally correct).

And even more particularly, we have:

> **Arithmetical Pluralism**: There are apparently incompatible languages of arithmetic (that are all equally correct).

In part II we saw that logical conventionalism entailed logical pluralism, and in chapter 11 we will see that mathematical conventionalism entails mathematical pluralism as well. The central point is the same in both cases: there are many incompatible rules for reasoning about sets (or set-like objects) that would be correct in any language that took them as basic rules (according to unrestricted inferentialism). Similar points go even for arithmetic (see 12.II).

But mathematical pluralism does not entail mathematical *in*determinacy, nor vice versa. Pluralism concerns alternative linguistic practices, while determinacy concerns truth in *our* practice. The two theses have no direct connection to each other. Alternative *X*-practices don't have to preserve every *X*-truth in our practice, considered syntactically. Conventionalists can coherently endorse both arithmetical determinacy and arithmetical pluralism. Pluralism and determinacy are sometimes run together in the literature, so it's important to recognize this point.[2]

If determinism isn't pluralism, then what is it? One thing that might be confusing is that **Mathematical Determinacy** uses the notion of "determinate" truth and falsity. But there is no deep circularity here. A general claim of determinacy concerns a *subject matter*; it says that every (non-vague) sentence in the part of the language dealing with that subject matter is either determinately true or determinately false. Say that a sentence is *determinate* if it has a determinate truth value. This is not an epistemic notion; it is instead a broadly metaphysical, or perhaps a semantic or conceptual notion. In any case, it has nothing to do with ignorance of the facts, even ignorance of the facts that is ineradicable in principle.[3] It instead concerns the very existence of a fact of the matter.

As noted briefly in 8.IV, it is standard to take the determinate truth (or falsity) of a sentence to be equivalent to the sentence being true

(or false) on all admissible interpretations of (the relevant fragment of) the language.[4] On this way of thinking, determinacy will seemingly follow from:

> **Categoricity**: A (fragment of a) language is categorical just in case every admissible interpretation of the language is isomorphic to any other, that is, all admissible interpretations of the language have the same abstract structure.

In logic textbooks, categoricity is standardly defined for theories, not languages, but the extension is straightforward. A full language (part II) determines an inaugural "theory" of sentences in the language.

Obviously, categoricity is a model-theoretic notion, so if it is used to do any explanatory work, conventionalists owe some account of what is happening. In section VI I'll call into question the worth of the standard argument from categoricity to determinacy (spoiler alert). Despite this, approaching determinacy with categoricity in mind provides a useful starting point. Though it is also worth keeping in mind that an important contrast between categoricity and determinacy will be important later in the chapter.

Given the result in the appendix to chapter 8, a straightforward application of bilateralism doesn't significantly change the philosophical situation here. So throughout this chapter I will stick with full languages formulated with standard natural deduction systems. But open-endedness, the other extension of standard proof systems introduced in 2.IV, will be crucial here. In fact, a number of mathematicians and philosophers have thought that the open-endedness of induction is key to explaining the categoricity of arithmetic. I agree with them, but the idea can only be properly evaluated once we have a clearer philosophical understanding of open-endedness itself.

II Open-Endedness Examined

A rule or principle is *open-ended* just in case it continues to apply as our language expands via the introduction of new vocabulary (2.IV).

Many logical and mathematical principles are open-ended; this much is not at all controversial and is very widely accepted.[5]

Of the Peano rules, the 5th rule—a version of the *principle of mathematical induction*—is schematic (8.II):

$$(P5) \qquad \frac{\varphi(0) \quad \forall\xi(\varphi(\xi) \rightarrow \varphi(\mathrm{S}\xi)) \quad \mathrm{N}\beta}{\varphi(\beta)}$$

Just as in the case of logical rules, the standard textbook understanding of the schematic letters in (*P*5) differs from the open-ended understanding. Shoenfield's influential graduate textbook *Mathematical Logic* explains the first-order induction schema as follows (where *L*(*N*) is our first-order arithmetical language):

> One statement of the induction axiom is that if a set contains 0 and contains the successor of every natural number in the set, then it contains every natural number. We cannot express this form in *L*(*N*), since we have no variables which vary through sets of natural numbers. Another form of the induction axiom states that if 0 has some property, and if the successor of every natural number having this property also has this property, then every natural number has this property. We cannot fully express this either, since we have no variables which vary through properties of natural numbers. However, we can express it for each property of natural numbers which can be expressed in *L*(*N*).[6]

If the textbook understanding was a faithful reconstruction of our actual arithmetical language we would be committed to induction only for arithmetical predicates in our current language. That is, according to the textbook understanding, if we add some new arithmetical predicate "F" to our language, it is an open question whether or not induction applies to "F".

This is clearly too weak to be a faithful reconstruction of our actual mathematical practice. If someone hesitates to apply induction to a newly introduced arithmetical predicate they are betraying conceptual incompetence. The proper way to understand the induction rule so as

to model our actual arithmetical practice is open-endedly. Here's Vann McGee on the open-endedness of induction:

> Whenever we adopt and name a stray dog or we introduce a new scientific theory with concomitant vocabulary, we expand the language. Fundamental principles of inference, like the principle of mathematical induction, are upheld even with the enlarged vocabulary. We don't have to reassess the validity of mathematical induction when we expand our inventory of theoretical concepts, for our current understanding of the natural numbers ensures that, no matter how we expand our language in the future, induction axioms formulated with the enlarged vocabulary will still be true. That's one reason for the remarkable stability of mathematics; scientific paradigms may come and go, but the principles of number theory, pure and applied, remain the same.[7]

I agree with McGee and many others that our arithmetic involves open-ended commitments that aren't perfectly modeled by fixed-language formal theories. This opens up the possibility, at least, that the formal limitative results that put pressure on determinacy and categoricity might be circumvented, in some fashion or other, by open-endedness.

If this is to have a hope of working, open-endedness cannot be a mere notational adjustment. In particular, open-endedness itself cannot be explained in terms of a non-open-ended background theory. To think of things in this way is to misunderstand what open-endedness is supposed to be. Unfortunately, many standard treatments of open-endedness encourage exactly this type of misunderstanding.

One popular approach formalizes open-ended commitments using a standard background model theory.[8] With this done, you can then prove a categoricity result about arithmetic within this background model theory. The problem with this approach is that if the background model theory is itself framed in a standard first-order language, then the original problem we were trying to solve recurs at the meta-level. The background model theory will itself have

nonstandard interpretations that can induce nonstandard interpretations of the open-ended theory embedded within it. On the other hand, if the model theory is itself presented in a language that includes open-ended schemas, then we are only "explaining" open-endedness with a further appeal to open-endedness.

Similar problems arise if we try to explain open-endedness using second-order logic. In second-order arithmetic, induction is not a schema, but a single universally quantified sentence:

$$\text{(Induction(SO))} \qquad \forall X((X0 \wedge \forall x(Xx \rightarrow XSx)) \rightarrow \forall x Xx)$$

The second-order quantifier here ranges over numerical properties in extension or (equivalently) sets of natural numbers. But while second-order arithmetic is categorical under its standard semantics, there are well-known alternative semantics for which categoricity fails.[9] Explaining the categoricity of arithmetic by appealing to second-order logic only replaces the mystery of arithmetical determinacy with an equivalent mystery about how our practices rule out nonstandard interpretations of our second-order quantifiers.[10] What is doing the work here is a comprehension principle in the background logic (9.II)—where "X^n" does not occur free in ϕ:

$$\text{(Comp)} \qquad \exists X^n \forall x_1, \ldots, \forall x_n (X^n x_1, \ldots, x_n \leftrightarrow \phi)$$

And this is a schematic principle that must itself be understood open-endedly if our grasp of (Induction(SO)) is to be categorical.[11] In essence, this is the *very same* problem that we started with, just pushed up a level (or maybe moved to a different spot).

A related formal approach to open-endedness understands open-ended schemas in a first-order logic augmented with second-order variables (but no second-order quantifiers).[12] This is formally equivalent to taking an open-ended schema as its universal closure in second-order logic. On this view, open-ended Peano arithmetic is a particular sub-theory of second-order arithmetic: second-order Peano arithmetic's Π_1^1-fragment. It seems unlikely that we can make any

philosophical gains by moving from full second-order PA to one of its proper sub-theories, understood standardly.

The lesson here is that open-endedness must be accepted on its own terms rather than reduced to a standard background theory. We follow the fifth Peano rule, (*P*5), in a fully open-ended fashion. This means that we not only have the dispositions that constitute following the rule for every instance that can be formulated in our current language, but we are disposed to continue to follow the rule as the language expands via the addition of new vocabulary. Our disposition to follow (*P*5) applies beyond our language. But this disposition is not explained in terms of some determinate totality like "all expansions of our language" or "all possible expansions of our language". Following Bertrand Russell, we might put this by saying we are disposed to accept induction in *any* expansion of our language, rather than in *every* expansion of our language.[13]

Without defining it in some background theory, we can attempt to clarify the notion of open-endedness by introducing some terminology. Let's say that we are committed to all of the sentences that follow from all instances of the Peano rules in our current language, and that we are precommitted to all sentences that follow from all instances of the Peano rules in well-behaved expansions of our current language. To understand these notions, we have to understand what an "expansion" of our language is and what makes an expansion "well-behaved".

As a simple model, we can think of an expansion of our full language as adding some vocabulary and thus new sentences to our language while keeping all of the old language in place, including our commitments. To make this all a bit more explicit, let us say that a synchronic language $\mathscr{L}$ is a pair whose first member is a formal language *L*, conceived of a set of sentences, and a commitment set *C*, which is the the set of sentences from *L* that are provable using the basic rules of the language. Another synchronic language is an expansion of $\mathscr{L}$ if its sentences and its commitment set are (possibly proper) supersets of $\mathscr{L}$'s. Synchronic languages provide a rough formal model of natural language arithmetic at a given point in time, with the current vocabulary held fixed.

The key insight behind open-endedness is that natural languages are *diachronic*, rather than *synchronic*—they constantly grow via the addition of new vocabulary and new commitments. This has been stressed since 2.IV and is an obvious feature of natural languages. To model this, we need to work not with synchronic languages but with sequences of expanding synchronic languages. But the expansions must meet certain conditions if this is to provide a plausible, though simplified model of actual arithmetic. This is why I limited things to "well-behaved" expansions above. We can say that an expansion is well-behaved just in case it is both consistent and lacks vagueness. Consistency here is simply syntactic consistency of the a synchronic language's commitment set. Why not the more demanding property of ω-consistency? Because we want ω-consistency to be forced by our arithmetical rules themselves, not by the definition of a "well-behaved expansion". This is simple enough, but the relevant kind of vagueness requires some unpacking.

Our mathematical languages are precise, not vague. Though this applies beyond arithmetic, to know how to apply this general constraint to our arithmetical languages we need to know what vagueness would look like in arithmetic. A natural first pass at this is to say that a predicate "F" is arithmetically vague just in case (*i*) there is some natural number *n* such that for any natural number *m* greater than *n*, ¬Fm $\in C_{\mathscr{L}}$; (*ii*) there is some natural number *k* such that any natural number *l* less than *k*, Fl $\in C_{\mathscr{L}}$; (*iii*) there is some natural number *a* such that neither "Fa" nor "¬Fa" is in $C_{\mathscr{L}}$ (or vice-versa, reversing direction). We could complicate this definition by adding something about penumbral connections, but the definition doesn't exactly work as stated, in any case.[14] Call a number "almost prime" if it has a small number of non-trivial divisors. The notion of being almost prime is no threat to induction, but it does seem intuitively vague without satisfying this definition. Its vagueness is owed to being defined in terms of the arithmetically vague notion of "a small number of".

There are other problems with this definition. The assumption of Cognitive Naturalism (1.IV) seems to demand that for any language that beings like us can possibly learn, at each point in time the language's commitment set is recursive. But then the commitment set

includes elementary arithmetic, is consistent (presumably), and recursive, and so it has its own Gödel sentence, G. Now we consider a predicate “F” defined as follows:

$$Fx =_{def} (x = 0 \vee (x = 1 \wedge G))$$

This particular predicate *seems* to count as vague by the above definition, since while 0 is F and for $\forall x > 1 : \neg Fx$, 1 is neither F nor not-F since the commitment set doesn’t include G by assumption. But it isn’t clear that this is truly vague, since if arithmetic is determinate, “G” is true in our current synchronic language, even if unprovable, and so 1 is F. The lesson here is that starting with the PA-rules, in each timeslice of our language, there will be general sentences (like “G” and “Con(PA)”) neither provable nor refutable in that timeslice, and that we can use these undecidable *generalities* to define predicates (like “F”) for which it is undecidable whether a given *particular number* satisfies the predicate. Given this, we clearly don’t want well-behaved expansions of our language to lack the resources to define predicates like “F”.

What we really want is for primitive arithmetical predicates to to be determinate over the standard naturals. This much is guaranteed if we restrict newly added *primitive* predicates to those that are determinate for each numeral—for predicate “F” and numeral “n”, either $Fn \in C_{\mathscr{L}}$ or $\neg Fn \in C_{\mathscr{L}}$. This will guarantee that our primitive notions aren’t “vague” in the above sense. Some complex predicates will be vague in the above sense, but the indeterminacy only enters in at the level of generalizations or notions defined in terms of generalizations. This approach rules out predicates like “small” and “almost prime” but without ruling out predicates or notions defined in terms of arithmetical undecidables.

Intuitively, the well-behaved expansions of our arithmetical language are those in which our old arithmetical vocabulary can be homophonically translated into the arithmetical vocabulary in the expanded language (this will be of more direct importance in 11.III). The characterization here does a passable job of capturing this intuitive idea, but it should not be thought of as aiming to do any more than that. Speaking in our natural language metalanguage, we are

open-endedly quantifying over well-behaved expansions of our language. This does not amount to defining open-endedness in non-open-ended terms. That would be to commit the same mistake that was warned against above. Open-endedness is not to be understood in standard terms. The discussion of expansions is only meant to clarify the notion, not to explicitly define it or explain it away.

The key idea that connects open-endedness to categoricity is this: interpretations of our arithmetical language are constrained not just by our *commitments*, but by our *precommitments* as well. As such, if we can find some way to rule out nonstandard interpretations using our precommitments, then we will have a plausible explanation of how our arithmetic manages to be categorical. Can this be done? A number of philosophers have thought so, but the two strategies that have been pursued in the literature are, I will argue in the following two sections, problematic. But discussion of them will lead to an approach that I find more promising.

III Strategy #1: Collapse

Perhaps the most popular approach to securing categoricity with open-endedness—pursued by Shaughan Lavine and Charles Parsons—is to mimic a version of Dedekind's categoricity theorem for second-order arithmetic in an open-ended setting.[15] In this setting, the categoricity theorem is effectively a collapse theorem, much like the argument discussed in 9.IV. Standard categoricity results are couched in model-theoretic terms, talking of both languages and their interpretations; by contrast, the collapse-style result uses only predicates and axioms in the object language.[16] I'm going to continue to call a categoricity proof of this kind, one that doesn't explicitly talk in terms of models or interpretations but instead shows that analogous terms are equivalent, a collapse argument.[17]

Imagine that we have two theorists, Alpha and Beta, each of whom speaks a different language (α and β respectively) and each of whom accepts the full open-ended first-order Peano axioms in their home language. Imagine now that Alpha and Beta want to know how to translate the arithmetical language of the other. Let "N_α" and "N_β" be

their respective number predicates (and similarly for their respective zero constants and successor functions). Without loss of generality, let's take Alpha's perspective to make things easier to follow.

Alpha wants to prove that his number theory is equivalent to Beta's number theory; his strategy for doing so is to simulate Dedekind's argument. To start Alpha defines a function f:

Definition. $f : N_\alpha \mapsto N_\beta$: (*i*) $f(0_\alpha) = 0_\beta$; (*ii*) $f(S_\alpha x) = S_\beta f(x)$

I'll be assuming that Alpha can define a function in this manner without making heavy duty mathematical assumptions that might be inappropriate or question-begging in this context. To complete the proof, Alpha must show that f is both one-one and onto Beta's numbers. By using the open-endedness of arithmetic in both *α* and *β*, Alpha can proceed as follows:

Lemma. f *is one-one*

Proof. Assume otherwise so that there is some smallest N_α, n such that for some smaller N_α, m, f(n) = f(m). By definition, $n = S_\alpha p$ for some $p \geq_\alpha m$, so $f(n) = S_\beta f(p)$. We have two cases to consider: (*i*) $m = 0_\alpha$ and (*ii*) $m \neq 0_\alpha$. If (*i*) then by the definition of f, $f(m) = 0_\beta$ and so since f(m) = f(n), $S_\beta f(p) = 0_\beta$ which violates one of Beta's Peano Axioms (the one saying that 0_β isn't a successor). If (*ii*) then $m = S_\alpha q$ so f(m) $= S_\beta f(q)$ by the definition of f, so f(m) = f(n) gives $S_\beta f(p) = S_\beta f(q)$ so by another of Beta's Peano Axioms (the one saying that S_β is one-one) we have f(p) = f(q), but this means there is a $p <_\alpha n$ such that f(p) is identical to the image of f for some smaller N_α, which violates our starting assumption. Since neither of these possibilities can be realized, we have reduced the assumption that f is not one-one to absurdity. __

Lemma. f *is onto*

Proof. Assume otherwise and let n be the smallest N_β that isn't in the range of f. By the definition of f, 0_β is in the range of f and so $n \neq 0_\beta$,

which means that n = S_βk but – by the induction hypothesis—k is in the range of f, that is: for some N_α, m, k = f(m) and so by the definition of f, n = f(S_αm) which violates the induction hypothesis and reduces the assumption that f is not onto to absurdity. –|

Theorem. f *is an isomorphism between* N_α *and* N_β

Proof. From the two lemmas above together with the definition of f. –|

As with every collapse argument, there are two different versions of this argument: an *intra*-language version and an *inter*-language version (9.IV). In the intra-language version, we can assume that Alpha and Beta are working in the same background language. In fact, for this type of case, it's easiest to imagine that Alpha and Beta are the very same person (Alpha). In this context, Alpha has two open-ended arithmetical theories and thus a notational variant of the above proof can be used by Alpha to prove, to himself, that his two theories of arithmetic are notational variants of each other. Nobody disputes the validity of the proof in the intra-language case. But all that this intra-language result establishes is that Alpha must regard each of his number predicates as equivalent to the other, not that his grasp of arithmetic is determinate. So the intra-language collapse argument isn't relevant to categoricity.

The inter-language situation is much more subtle. In this case, the proof is supposed to establish that the number predicates in languages *α* and *β* are equivalent and must be regarded as so by both Alpha and Beta. Since there was nothing special about *α* and *β* as languages, if this works it generalizes to all languages that include open-ended arithmetic. But Hartry Field has questioned the onto portion of the argument in the inter-language case.[18] In particular, Field objects to Alpha's implicit uses of Beta's induction in the onto direction of the proof, claiming that this use is question-begging in this context, since it amounts to assuming that Alpha and Beta agree about what counts as a well-defined predicate, which is close to the very point at issue.[19]

Perhaps this issue can be finessed.[20] It doesn't really matter though, because of an even deeper problem, also pointed out by

Field. In general establishing that “F” and “G” are equivalent and must be regarded as so by both F-speakers and G-speakers does *absolutely nothing* to establish determinacy. It always leaves open the possibility that “F” and “G” are both indeterminate in precisely the same way. Field provides a nice example to illustrate this crucial point: in a classical language, the predicate “bald(x)” will be provably equivalent to the predicate “¬¬bald(x)”, but “bald(x)” is a paradigm case of an indeterminate predicate (that is: a predicate without a determinate extension) and the predicate “¬¬bald(x)” clearly just *inherits* the indeterminacy of “bald(x)”.[21]

Is Field’s criticism too quick? If the arithmetical theories of Alpha and Beta, respectively, have two distinct admissible interpretations, M_1 and M_2, the inter-language collapse theorem seems to rule out Alpha interpreting his own arithmetic using M_1 while at the same time interpreting Beta’s arithmetic using M_2, and similarly for Beta’s interpretation of Alpha’s arithmetic. But this description tacitly assumes that Alpha can distinguish between M_1 and M_2 in his own language, presumably by proving that they aren’t isomorphic to each other. But this will allow him to prove that one of these interpretations, at least, is inadmissible for both Beta and himself. This undermines the description of M_1 and M_2 as admissible but distinct interpretations of arithmetic.

I think that Field’s criticisms mainly stand. At best inter-language collapse arguments only show that *if* a nonstandard interpretation is admissible for a version of open-ended arithmetic, *then* it also admissible for a version of open-ended arithmetic formulated in some other language. They do not show that all admissible interpretations of open-ended arithmetic are standard. This means that the Dedekind-Lavine-Parsons strategy is, by itself, unable to secure determinacy. What we need is some bridge connecting the results of open-endedness to interpretations of our arithmetical language. The second type of strategy in the literature attempts to build this bridge.

IV Strategy #2: Anti-Overspill

The second strategy is suggested in some of Vann McGee's discussions of open-endedness and categoricity (Hartry Field has noted that several distinguished logicians also endorsed this type of argument to him in conversation).[22]

Consider an expansion of our arithmetical language that includes a predicate, "ST", for the standard numbers—the numbers that can be reached by applying the successor function to 0 finitely many times. Now, by open-endedness and standard logic, the induction rule, (*P*5), precommits us to the following sentence:

$$(Magic) \qquad \neg(\mathrm{ST}(0) \wedge \forall x(\mathrm{ST}(x) \rightarrow \mathrm{ST}(\mathrm{S}x)) \wedge \neg\forall x \mathrm{ST}(x))$$

But this precommitment alone rules inadmissible any nonstandard model of our practice. Consider a putative nonstandard model of our practice, *M*, we know that 0 is standard, and that standardness is inherited under successors; but also, since *M* is nonstandard, that *M* $\models \neg\forall x \mathrm{ST}(x)$, and so *M* $\not\models$ (*Magic*). If open-ended induction constrains admissible interpretations by requiring that they vindicate our precommitments and (*Magic*) is one of our precommitments, *M* is not an admissible interpretation of arithmetic. But *M* was an entirely arbitrary nonstandard interpretation: all that we knew about *M* was that it is nonstandard. So by generalization, *all* nonstandard models of our practice are inadmissible.[23]

I call this type of argument an anti-overspill argument. Robinson's overspill theorem says, in effect, that non-empty successor-closed proper initial segments of the numbers in nonstandard models of arithmetic *M* are not definable.[24] The idea behind the McGee-style argument is that a way to avoid overspill is induced by open-endedness, so that nonstandard models won't vindicate all of our precommitments. This doesn't contradict the overspill theorem itself, since overspill concerns *a fixed language*. In McGee's argument, overspill is avoided by assuming there is a well-behaved expansion of our language in which we have a predicate that is true of all and only the standard natural numbers.

Taken baldly, this approach has all the advantages of theft over honest toil. Field's comments on McGee's argument are devastating:

> Of course, it's true that if I could add a predicate that by some magic has as its determinate extension the *genuine* natural numbers, then I will be in a position to have determinately singled out the genuine natural numbers. That's a tautology, and has nothing to do with whether I extend the induction schema to this magical predicate. But if you think that we might someday have such magic at our disposal, you might as well think we have the magic at our disposal now; and again, it won't depend on schematic induction. So the only possible relevance of schematic induction is to allow you to carry postulated future magic over to the present; and future magic is no less mysterious than present magic.[25]

It seems that McGee-style arguments must assume that in some well-behaved expansion of our language we can define or uniquely characterize the standard numbers, but as Field notes, this is to "explain" categoricity by, in effect, assuming it. Is there any way around this problem? I will consider two different strategies.

The first uses empirical applications of arithmetic to induce categoricity via an anti-overspill argument. Consider the schematic version of (*Magic*):

$$(\textit{Magic Schema}) \qquad \neg[\phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(\mathrm{S}x)) \wedge \neg\forall x\phi(x)]$$

By open-endedness, this schema will also apply to predicates concerning *applications* of our arithmetic. This opens up the possibility of using features of the physical world to rule out nonstandard models. I'll sketch a version of this idea that is based on Hartry Field's approach to securing the determinacy of arithmetic.[26]

Let $\Phi(Z)$ be a predicate requiring that Z is a sequence of events which (*i*) has an earliest and a latest member, and (*ii*) is such that any two members of Z occur at least one second apart. Let's also make the following cosmological assumptions:

1. Time is infinite in extent (there is no finite bound on the size of sets that satisfy Φ)
2. Time is Archimedean (any set satisfying Φ will have only finitely many members)

And let Ω(*α*) be a predicate requiring that *α* is a number that is the number of events in some set of events *Z* such that Φ(*Z*). Now substitute Ω in the (*Magic Schema*), to get the sentence (*Magic* Ω), and then consider a nonstandard model *M*. Trivially, 0 is Ω, and given (1), if some number *n* is Ω then the successor of *n* is also Ω, and given (2), not all items in *M* will be Ω, since the nonstandard (so to speak) elements of the model won't be Ω. So, if vindicating all of our precommitments is a constraint on interpretations of our arithmetic, then *M* is not an admissible interpretation, since given the cosmological assumptions and a fixed interpretation for the physical terms used to define the predicate Φ, such as "event", $M \nvDash$ (*Magic* Ω).

Unlike McGee's argument, the Field-style argument isn't obviously question-begging. Instead of gifting us with a determinate notion of standardness from the start, we are gifted with determinate physical notions like "event" and "one second apart" and so on. Our arithmetical notions then get their determinacy indirectly.[27] For this reason, even if Field's approach succeeded on its own terms, it would make arithmetical concepts partly empirical and so wouldn't cohere with conventionalism. For this reason, I won't discuss it in detail here. But I don't think it works on its own terms. As Charles Parsons says:

> I find it hard to see how someone could accept that assumption who does not already accept some hypothesis that rules out nonstandard models as unintended on mathematical grounds. If our powers of mathematical concept formation are not sufficient why should our powers of physical concept formation do any better? We are not talking about events in a common-sense context, but rather in the context of a physical theory developed in tandem with sophisticated mathematics.[28]

There are a number of worries about a Field-style approach, taken on its own terms, but this is perhaps the most pressing.[29]

Sharon Berry has recently developed a different variation on McGee's approach.[30] Her idea is based not on assuming the determinacy of our physical notions, but rather of our *modal* notions. She claims that it is possible for there to be infinite random sequences of coin flips, so that we can assume that *any* infinite sequence of coin flips—each flip being either heads or tails—is *physically possible.* We must be careful here though, since while there are results showing that supertask computations are possible in continuous Newtonian universes, and also in relativistic spacetimes, these types of universes are non-quantum.[31] So even granting that random universes are physically possible, and that supertask computers are physically possible, I don't know if we have any reason for thinking them *jointly* physically possible. Maybe this isn't too problematic though, since even if such universes are not *physically* possible, perhaps they are *metaphysically* possible, in some general sense. This is enough to allow for a development of Berry's argument.

So assuming these universes are possible *in some sense*, let $CF_v(\alpha)$ be a predicate requiring that α is a number such that the αth coin flip in some given world v comes up heads. Now substitute CF_w in the (*Magic Schema*), to get (*Magic* CF), and consider a nonstandard model M of our practice. By the assumptions about coin flips and supertasks, there is a possible world w where all and only the coin flips corresponding to some successor-closed initial segment of M come up heads. So in w, 0 is CF, and if any number n is CF then so too is its successor, but not all items in M will be CF_w, since the nonstandard elements won't be, by assumption. So if vindicating all of our precommitments is a constraint on interpretations of our arithmetic, then M is not an admissible interpretation, since $M \nvDash$ (*Magic* CF_w), given the determinacy of the modalities used here and a fixed interpretation for the terms used to define the predicate CF_w, such as "coin flip" and "heads".

Berry's approach has some nice features. For example, on Berry's approach, ruling out M doesn't require assuming that *only* the truly

standard initial segment of *M*'s flips come up heads, since the argument will still work if the initial heads segment isn't standard but fails to exhaust the "numbers" of *M*. But a major disadvantage of Berry's approach compared to the Field-style approach is that, since these coin flips aren't *actual*, there is no sense in which we could be picking them out in some externalist fashion; instead we must *describe them*, in our own language. If our initial description lets some nonstandard sequences of coin flips through the net, then nonstandard models with the structure of these sequences will not be ruled out. This problem threatens to reduce the argument to showing that *we can rule out the models that we can rule out*, which is a tautology.

If our starting notion of an infinite, successor-closed sequence of heads coin flips is, in fact, indeterminate, then there will be nonstandard sequences of coin flips that slip through the net and get counted as standard by our lights. The Field-style approach didn't have to deal with this problem, since the sets of events it appealed to were *actual*, but here the sequences of coin flips are *merely possible*. Because of this, there isn't an obvious way to offload the problem of determinacy in the modal case, as there was in the physical case.[32] If this worry is right, then Berry's approach falls victim to a more complex variant of the problem that plagued McGee's approach.

I think that Berry would respond to this charge by claiming that her realism about both the modal facts and the coin flip facts does, in fact, allow us to externalize in a way analogous to the Field-style approach. This doesn't seem very plausible to me though. Since these sequences are not actual, we cannot causally interact with them, but can only *describe* them. So we can't "point to" a coin flip in the "standard" portion of *M* that is actually tails but is counted as "heads" by the nonstandard interpretation. It is no use to protest that our descriptions will take some coin flips that *truly* come up tails to have come up heads—McGee could likewise protest that some *truly* nonstandard numbers are being counted as standard by *M*.[33] Is it any better that a non-actual coin flip is being misdescribed instead of an abstract object? Does the very idea even make sense? I doubt it.[34]

For these reasons, I don't think that Berry's approach succeeds on its own terms. But as with the Field-style approach, conventionalists

like myself will also reject the general framework used in Berry's argument. To use *possibilities* or *worlds* as explanatorily freestanding posits in our metasemantics is to violate **Metaphysical Naturalness** (1.IV). Another issue is that both the Field-style approach and Berry's approach are implausible if taken as *analyses* of our arithmetical concepts. So I think we must look elsewhere for a satisfying—and conventionalist-friendly—account of determinacy.

V Strategy #3: Communicative Generality

By building upon previous approaches, and learning from their problems, we can construct a straightforward argument from open-endedness to categoricity. This argument is part of the philosophical explanation—in naturalistic terms—of how we Turing machines manage to have a determinate conception of arithmetical truth. But it won't be the entire story.

What have we learned from extant strategies? From the collapse approach, we learned that we need to deal with interpretations directly; and from the anti-overspill approach, we learned that we need to avoid subtly assuming the very point at issue. A more general lesson, gleaned from both approaches, is that the right approach here will be *philosophical*, not *merely technical* (a famous methodological sermon again comes to mind here).

Let me present my argument baldly before discussing and defending it. So baldly, in fact, that at first glance it will look like a rabbit is being pulled out of a hat. Here's the argument:

1. Our open-ended arithmetical practice rules out any nonstandard interpretation of arithmetic that can, in principle, be communicated to us
2. Any nonstandard interpretation of arithmetic can, in principle, be communicated to us
3. *So*: any nonstandard interpretation of arithmetic is inadmissible (1,2)
4. *So*: arithmetic is categorical (3)

(4) follows from (3) by definition and (3) straightforwardly follows from (1), (2), and definitions, so the key to establishing the argument is justifying premises (1) and (2) and showing that, when properly understood, they don't beg any important questions.

Consider our synchronic language *L* and let A be all sentences in the language that follow from the Peano rules, and let *M* be a nonstandard interpretation of arithmetic. Against this background, the justification of premise (1) is relatively straightforward: for *M* to be communicated to us is for there to be some well-behaved expansion of our language *L*+, in which there is a predicate for *M*'s domain and an ability to see that *M* is nonstandard. We can imagine this as involving *L*+ including predicates for both the domain of *M* ("M") and *M*'s standard portion ("ST^{M}"), along with axioms for this new terminology entailing "$\exists x(\mathrm{M}x \wedge \neg \mathrm{ST}^{\mathrm{M}}(x))$".

Now, by open-endedness and standard logic, the induction rule, (*P*5), entails that we are precommitted to the following sentence of *L*+:

$$(\textit{Relative Magic}) \qquad \neg(\mathrm{ST}^{\mathrm{M}}(0) \wedge \forall x(\mathrm{ST}^{\mathrm{M}}(x) \rightarrow \mathrm{ST}^{\mathrm{M}}(\mathrm{S}x)) \wedge \neg \forall x \mathrm{ST}^{\mathrm{M}}(x))$$

Now take *M* as a model of our arithmetical practice. Obviously, 0 is in the standard part of *M* and *M*-standardness is inherited under successors, but also, since the standard portion of *M* doesn't exhaust *M*, by the axioms for our new terminology, $M \models \neg \forall x \mathrm{ST}^{\mathrm{M}}(x)$, and so $M \models \neg$(*Relative Magic*). Since the open-endedness of induction constrains admissible interpretations of our language by demanding that they vindicate our precommitments, and (*Relative Magic*) is a precommitment of ours, *M* is not an admissible interpretation *if* it can be communicated to us. Since no special assumptions were made about *M* beyond its being an interpretation of our arithmetic that is nonstandard and can be communicated to us, we have shown that *any* interpretation meeting these conditions is ruled out by our open-ended arithmetical practice, which is exactly what (1) says.

This argument was obviously inspired by McGee's argument, but it is importantly different. McGee's argument wasn't for (1), but rather for (4), via (3). Accordingly, McGee's argument required an extension

of our current language that included a predicate for the standard numbers, *in an absolute sense*. But my argument only assumed a *relative* notion of nonstandardness, not an absolute notion. The idea of (1) is basically that any communicable perspective that rules our practice non-standard is itself inadmissible, given open-endedness. This means that the argument is not assuming that the standard portion of *M*, is itself *absolutely* standard. For this reason, unlike McGee's argument, this argument doesn't cheat. It plays fair, but establishes only a more limited conclusion.

The argument here is also related to the collapse argument from section III. By the assumption that the nonstandard interpretation *M* can be communicated to us, we can add predicates to our language for *M*'s domain and for *M*'s standard portion. We can then rule out the interpretation of our number predicate that makes it coextensive with the totality of *M*'s domain. In either form then, the reasoning here isn't question begging.

Another challenge concerns (1)'s quantification over "all" interpretations of our arithmetical language. This sounds like it assumes a perfect grasp of a determinate totality of interpretations of our language, thus gifting us with something like semantic omniscience. But this charge is unfair. My argument for (1) used only standard universal generalization—from the assumption that *M* meets certain conditions, it is ruled inadmissible as an interpretation of our practice, and so, since no auxiliary assumptions about *M* were made in the argument, we conclude that *all* interpretations meeting these same conditions are likewise inadmissible. Surely using standard quantificational reasoning is allowed in this context, but if so, then this objection is misconceived.

Premise (2) claims that any nonstandard interpretation can, at least in principle, be communicated to us. At first glance, this premise also threatens to gift us semantic omniscience, this time in a quite direct sense. Perhaps for this reason, some defenders of open-endedness have stopped short of endorsing it. Witness Charles Parsons discussing what follows from his collapse-style approach to categoricity:

> This does not protect the language of arithmetic from an interpretation completely from outside, that takes quantifiers over numbers as ranging over a non-standard model. One might imagine a God who constructs such an interpretation, and with whom dialogue is impossible. But so far the interpretation is in the Kantian phrase "nothing to us." If we came to understand it (which would be an essential extension of our own linguistic resources) we would recognize it as unintended, since we would have formulated a predicate for which, on the interpretation, induction fails.[35]

Parsons seems content to rest with the conditional that *if* a nonstandard interpretation can be communicated to us, *then* we can rule it out—essentially my premise (1)—while dismissing the importance of nonstandard interpretations that can't be communicated to us. Personally I don't quite see how this could be "nothing to us", since if it isn't ruled out by our practice, then our arithmetic isn't categorical, and so is, apparently, indeterminate. That, surely, is *something* to us. But maybe Parson is thinking counterfactually, and maybe counterfactual ruling out is enough? Whether this is so or not depends on whether *some* interpretations are incommunicable to us, in principle.

But can any sense be made of an *in principle* incommunicable nonstandard interpretation of our arithmetic? I don't think so: for any nonstandard interpretation of our arithmetic *M*, we can conceive of a being or collection of beings that grasps *M* and uses it to interpret our arithmetic nonstandardly (Parsons does exactly this when imagining a God who constructs *M*).[36] At first glance, this step may seem controversial, but all that it really involves is the rejection of what we might call *ineffable interpretations* of our arithmetical practice—interpretations of our practice that can't be grasped linguistically, *even in principle*.

I don't think that ineffable interpretations are possible. How could it be incoherent to simply stipulate that there is an Oracle whose language includes a means of describing *M*? The thought behind this question is not that the Oracle would be physically possible, but that it is minimally coherent in that we can discuss and reason about it

without falling directly into nonsense. Proponents of ineffable interpretations must deny this, but I find this denial mysterious.

Rejecting these mysterious ineffable interpretations only takes us part of the way to establishing (2), since we must still argue that any "effable" interpretation can, at least in principle, be communicated to *us* (something else that Parsons apparently denies). Since the interpretation is effable, without loss of generality, let's assume that Martians have effed it. We can assume that the Martians have expressive resources going beyond our own in any way we like—uncountably many primitives, infinitary operations, and so on. We can even assume that the Martians have expressive resources undreamt of in our philosophy. By assumption, the Martians interpret our number theory as being "about" a nonstandard model *M*. If they could communicate this to us, then by premise (1), *M* would be inadmissible. So we must assume that the Martian interpretation is *in principle* incommunicable to us.[37]

What could possibly account for in principle incommunicability? A natural thought is to account for it using whatever complex resources the Martian language contains that are beyond our meager cognitive powers. There is nothing ad hoc about this response, but it confuses the question we were trying to answer with a different question, only tangentially related. We were trying to explain why we could not have a predicate in our language whose extension consisted of all and only the Martian-numbers; in answer we've been told that we cannot learn the Martian language because it includes resources that go beyond our cognitive powers. But this isn't responsive to the challenge. We aren't trying to *learn* the Martian language; we're trying to add a predicate—call it "N_{Mars}"—to *our language* that has the same extension as the Martian numbers. If adding such a predicate required us to learn infinitary or non-algorithmic notions, then we'd have a problem. But that isn't required.

In fact, one of the most widely agreed upon theses of modern metasemantics is the externalist point that sometimes a predicate or name only means what it does in our mouths because we systematically defer to others in their use of the predicate or name. This central idea was called *the division of linguistic labor* by Hilary Putnam, one of its discoverers.[38] The cases that Putnam and others

discussed concerned people who were all speaking the same natural language, but this isn't essential. I could count some stuff as an example of "water" just in case the world's leading Spanish-speaking water authority counted it as an example of "agua". What really matters is that my deference allows a name or predicate in my mouth to have the same extension as it does in someone else's mouth even when I lack the ability, on my own, to single out that extension. Applying this to our example, we could and would get the right extension for the predicate "N_{Mars}" in our language by simply deferring to Martian usage. For familiar externalist reasons, it simply doesn't matter what kind of complicated resources the Martians use to fix the extension of their number predicate. Whatever they do, we can piggyback upon it without doing it ourselves. The simplest way to do this is by description, using the "N_{Mars}" predicate. Because our arithmetic is open-ended, after we introduce this predicate by description, we can apply induction to it. And if we can apply induction to it, then we can show that our arithmetic isn't nonstandard relative to the Martian numbers.

The overall argument for premise (2) is also an application of universal generalization. Assume there is a given nonstandard interpretation of our arithmetic, *M*, then reflections on metasemantics and metaphysics (social externalism, deference, reference fixing by description, the metaphysical weirdness of ineffability) lead us to conclude that there is no in principle bar to *M*'s being communicated to us. Since we made no special assumptions about *M* itself, we generalize to get premise (2).

The rest of the main argument now proceeds smoothly. Open-endedness, together with various metasemantic and philosophical supplementation, has provided us with a successful argument for categoricity. I think this argument succeeds where others failed because it approaches things *generally*, and so has a different structure than the other arguments. Instead of trying to find some *particular* way to show our arithmetic standard, this argument tries to show that the *very idea* of a nonstandard admissible interpretation is as incoherent, in its way, as that of a round square.

Rather than assuming a non-relative notion of standardness or offloading the work to the (physical or modal) facts to be described,

this approach tries to work in complete generality, so that a nonstandard but admissible interpretation can be rationally rejected, *as such*. I think that this generality is essential in avoiding the master problem with the approaches of McGee and Berry. My approach uses only a relative notion of standardness, so it doesn't cheat. There is no requirement, at any point, that the relative notion of standardness is the absolute notion of standardness. The generality is gained by standard quantificational reasoning, rather than by appealing to descriptive abilities that are equivalent to the original problem's solution.

I don't think that this argument smuggles in any scientifically disreputable magic. But some may not yet be convinced of this. We can imagine a skeptic arguing as follows: your argument shows that by incorporating all possible well-behaved expansions of our arithmetical language into our current arithmetical competence via open-endedness, we thereby attain a categorical grasp of arithmetic. But in effect, this is to collect up all possible expansions of our arithmetical language into a supertheory, call it A^{Super}. Now A^{Super} cannot be algorithmic, because if it were, then it would admit of nonstandard interpretations and the argument would fail. So the friend of open-endedness must claim that A^{Super} is non-algorithmic. But then our current arithmetical practice amounts to a grasp of A^{Super} and so, since this theory is non-algorithmic, the open-ended argument gifts us with magical, non-algorithmic abilities, thus violating **Cognitive Naturalism** (1.IV).

If this is right, we've landed back in the very same trap we set out to avoid. But while this objection is quite intuitive, the move of collecting up all possible expansions is yet another attempt to fit open-endedness into a standard mold, something I strongly rejected in section II. Friends of open-endedness are not claiming that there is some Peirce-style arithmetical "end of inquiry" where the ultimate supertheory is to be found.[39] The supertheory you end up with in the "collecting up" objection is indeed unlearnable, but nobody, least of all friends of open-endedness, claimed that it was learnable, much less that it was *already learned*. You follow the very same induction rule as the language expands. Your general dispositions do not change. For any nonstandard interpretation of arithmetic *M*, you can add

predicates and principles to your arithmetic that explicitly rule *M* out. This is different than saying you can add predicates and principles that explicitly rule out every nonstandard interpretation of your arithmetic. Your commitments are recursive at any given time. But your precommitments aren't, and yet open-ended induction covers them all at once. Nothing I have been saying requires being able to do anything that a Turing machine cannot do.

This point is key. It is why using open-endedness to explain categoricity is an advance upon a mere move to a categorical but non-recursive theory of arithmetic (like second-order Peano arithmetic in standard second-order logic)—such moves gift us mysterious non-algorithmic mental powers, but the move to open-endedness does not. What the open-endedness of our arithmetic shows is not that there is some supertheory of arithmetic that we can enumerate, but rather that *the very idea of a nonstandard yet admissible interpretation of our arithmetical practice is as incoherent as that of a round square.*

VI From Categoricity to Determinacy?

Because of the argument of section V, I tentatively think that our arithmetic is categorical. Why "tentatively"? Because I am not maximally confident in the argument. I accept the conclusion on the basis of the argument, but I do so cautiously. Despite these doubts, I believe that the open-ended Peano rules give rise to a categorical conception of arithmetic. Their open-ended validity rules out all nonstandard interpretations by the **MVC**—if nonstandard interpretations were admissible, the open-ended induction rule would not be strongly valid.

The central reason that categoricity has been thought *philosophically* important is because it seems to lead to determinacy, via a simple model-theoretic argument. The argument runs as follows: take a sentence ϕ in the language of arithmetic, then in any model of arithmetic *N*, either $N \models \phi$ or $N \models \neg\phi$, by the nature of models. But since arithmetic is categorical, all models have the same structure as *N*, so if $N \models \phi$ then ϕ is true in *all* models of arithmetic,

and if $N \models \neg\neg\phi$, then ϕ is false in *all* models of arithmetic. So by cases and the characterization of determinate truth from section I, any sentence in the language of arithmetic is either true in all models or false in all models, and so is either determinately true or determinately false ($\triangle T[\phi] \vee \triangle F[\phi]$). This argument apparently leads from categoricity directly to determinacy.

This type of argument is entirely standard. As far as I know, it hasn't generated much controversy, if any. But I have worries about it. My worries are analogous to the worries about the simple, bivalence-based approach to determinacy that I discussed and rejected in 8.IV. My central worry is that it is only by assuming something like (Bivalence) for *models* of arithmetic that the argument manages to work. Given any model and sentence ϕ we accept a disjunction, either ϕ is true in the model or the negation of ϕ is true in the model. This seems very much like (Bivalence), only for the notion of "truth in a model" rather than "truth" simpliciter. Given this, have we really made much explanatory headway? Is the notion of "truth in a model" one for which we can simply stipulate a bivalence-like principle and be done?

I'm not a Platonist. So while I'm happy to work with a notion of "truth in a model", I don't think that models are explanatorily freestanding *things* out there in Platonic heaven. And I doubt that our notion of a "model" or "interpretation" of arithmetic can be assumed more determinate than the fundamental notion of arithmetical truth itself. The assumption of (Bivalence) or (Bivalence+) didn't solve the issue of determinacy because such assumptions neither guarantee a "fact of the matter" about every arithmetical question nor explain determinacy in a philosophically satisfying manner. I find myself with a similar worry at the meta-level, about our principles concerning "truth in a model".

This worry infects the argument from categoricity to determinacy with something like a version of Field's bald/¬¬bald point from section II. Perhaps "truth in a model" is indeterminate over models of arithmetic, but models of arithmetic are all indeterminate in exactly the same way? The categoricity proof will have bought us something, but not enough to think arithmetical truth is determinate, at least not with any confidence. According to this worry, the gain in moving from

collapse theorems to categoricity proper was illusory at worst and overblown at best.

There is a technical result that dovetails with my philosophical worry. Joel Hamkins and Ruizhi Yang have shown that two models of standard set theory M_1 and M_2 can have the very same standard model of arithmetic while at the same time *disagreeing* about the truth of some shared arithmetical sentence φ.[40] More philosophically, my worries are related to the role of "models" and "interpretations" for anti-Platonists. If models aren't explanatorily freestanding objects in Platonic heaven, then the facts about them must come from—where else?—our linguistic conventions.[41] Because of this, I don't think that we can rest content with the categoricity argument of section V *even if it is successful*.

There will still be an important role for categoricity to play (see section IX), but to answer these worries the conventionalist needs a clearer account of how arithmetical truth relates to our linguistic conventions. Only such an account can assure us that we aren't smuggling assumptions in through the back door that were denied entrance at the front. Fortunately, I think the open-ended account of categoricity can be supplemented with an entirely different approach, one that will give us a better purchase on arithmetical generalizations and allow us to securely bridge whatever gap there may be between categoricity and determinacy. The supplementary idea harkens back to Hilbert, Tarski, and Carnap.

VII The Omega Rule

Around the time of the incompleteness theorems, Hilbert and Tarski independently considered augmenting arithmetic with an infinitary meta-rule:

(OMEGA) If $\vdash \varphi(0), \vdash \varphi(S0), \vdash \varphi(SS0)$, ... then $\vdash \forall x \varphi(x)$

The antecedent here includes infinitely many provability claims, one for each of the standard numerals.[42] Shortly thereafter, in *The Logical Syntax of Language*, Carnap used a somewhat different version of

the omega rule to define "analytic" for his Language I.[43] I'll discuss the details of Carnap's approach in 13.II. Leaving fidelity to Carnap aside, a modern version of (OMEGA) formulated as an inference rule runs as follows:

$$(\omega) \quad \frac{\varphi(0), \varphi(1), \varphi(2), \ldots}{\forall x(\mathrm{N}x \rightarrow \varphi(x))}$$

This statement of (ω) uses standard numerals rather than successor notation and makes the restriction to numbers explicit. The mathematically interesting fact of note here is that closing even weak arithmetical theories like Robinson arithmetic (Q) under (ω) leads to true arithmetic (TA)—the complete first-order theory of the standard model of arithmetic (see the appendix for a proof of this fact).

Speaking *formally* this result seems to be exactly what is needed: a bivalent theory of (first-order) arithmetical truth characterized wholly in terms of inference rules. A particularly nice feature of (ω) is that it seems to tame quantification over *all* numbers by connecting arithmetical generalizations directly to the standard numbers. The rule-form of the induction principle, (*P*5), also concerned arithmetical generalizations, but the premises of the induction rule *already include a generalization over numbers*. Impressionistically, we might say that if someone was worried about the determinacy of quantification over natural numbers, induction could do little to allay those worries, given that induction only links one quantification over natural numbers with another. The omega rule, on the other hand, links the universal generalization to all of the relevant instances.

Speaking *philosophically*, things are less clear. It is almost universally agreed that beings like us cannot follow the omega rule without violating something like the **Cognitive Naturalism** principle (1.IV). In one of the earliest published discussions of the rule, Tarski recognized and discussed the difficulties involved, stating that

> …such a rule, on account of its 'infinitistic' character, departs significantly from all rules of inference hitherto used, that it cannot easily be brought into harmony with the current view of the deductive method, and finally that the possibility of its practical application in the construction of deductive systems seems to be problematical in the highest degree.[44]

Tarski's view is very, very widely shared by contemporary philosophers of mathematics. In discussing the omega rule in his excellent book on Gödel's theorems Peter Smith says:

> …finite minds can't really take in the infinite number of separate premisses in an application of the ω-rule: that's an impossible task.[45]

But despite being almost universally shared, I don't think that this concern has ever been developed in sufficient detail. Let me try to develop it here so that we can assess it. The central concern behind the objections voiced by Tarski and Smith seems to be that, since the (ω) has infinitely many premises, following the rule would involve performing an impossible supertask.[46]

Perhaps not all supertasks are impossible, but suppose that following (ω) required constructing, either in thought or in deed, an infinitely long proof like the following:

(1) $\varphi(0)$
(2) $\varphi(1)$
(3) $\varphi(2)$

… …

(n + 1) φ(n)

… …

(∞) $\forall x(\mathrm{N}x \rightarrow \varphi(x))$

In a proof of this kind, the conclusion, at line (∞), is only reached *after* infinitely many previous lines. Presumably, for beings like us, producing a proof of this kind is physically impossible—producing each premise takes some finite minimum amount of time, ϵ, and a supertask each of whose sub-tasks takes at least ϵ long to complete cannot be completed in a finite amount of time (this was noted in a different context in 7.III).

I think that this point has convinced virtually everyone that appealing to the omega rule is hopeless. Virtually everyone, but not everyone. A few philosophers have argued that while beings like us can't actually follow the rule, it still serves as a plausible idealization of our actual, finite arithmetical abilities.[47] If taken as a stopping point, this is problematic. It is one thing to idealize our finite capacities beyond any finite bound and quite another to idealize our finite capacities into infinite capacities. Unless it is explained exactly *how* the idealization relates to our practice, idealizing to (ω) has all of the advantages of theft over honest toil.

A few brave souls have simply helped themselves to the omega rule without saying anything about how beings like us are able to follow the rule.[48] This tradition started with Carnap himself, who, after noting the doubts of Tarski and others, simply said in response:

> In my opinion, however, there is nothing to prevent the practical application of such a rule.[49]

It's impossible to know exactly what Carnap was thinking here. It is most likely that he was imagining a background metalanguage in which we can reason with infinite sets of formulas and express all of the premises of (ω) at once while using only finitary rules.[50] If this is what he was thinking, it isn't adequate. Just like open-endedness, the omega rule must be taken seriously, on its own terms, rather than simulated in a standard, finite background theory.[51]

Skepticism about a human ability to follow the omega rule is a matter of philosophical consensus. I was a part of this consensus until the late spring of 2015; I even gave voice to the consensus in my earliest published comments on determinacy.[52] But as I continued to think about rule-following, I tried to think carefully about *why* beings

like us can't follow the omega rule, especially in light of my then nascent dispositionalist theories of rule-following and inference. I eventually realized that all of the arguments against our ability to follow (ω) were based on assumptions about rule-following or inference that I already independently rejected. More than that though, they were based on claims that almost everybody else rejected too.

In 2.VIII I developed a dispositionalist account of inference rule-following. Applying that account tells us that following (ω) requires dispositions not to violate (ω), to enforce (ω) in two directions, to comply with (ω) when there is any disposition to form attitudes toward all of (ω)'s component sentences, and to infer according to (ω) when given the chance and when disposed to continue to accept all of (ω)'s premises. Right away we can see that none of these conditions requires us to construct *proofs* using (ω), so the argument above is a nonstarter. The only conditions for which the infinitude of (ω) might cause problems are the dispositions to comply and to infer according to (ω) in certain special circumstances.

In both cases, the main potential stumbling block is the possibility of having attitudes toward all of the infinitely many premises of an instance of (ω). But despite the infinitude of premises, I think that not only is this *possible*, it is sometimes *actual*. Key here is the functional-dispositional account of acceptance from 2.V. According to such an account, accepting p does not require being in any particular occurrent or explicit state involving p. We can, in fact, accept infinitely many claims, at a given time. Accommodating this is one of the key selling points of dispositionalist approaches to attitudes. It is so important and plausible that even non-dispositionalists often feel compelled to add a notion of "tacit" or "implicit" belief to allow for it (see the discussion in 2.V). In whatever form, something like this is very widely believed.

Combining these two points—rule-following dispositionalism and acceptance dispositionalism—with the functionalist account of inference from 2.VI suggests that beings like us can infer according to the omega rule. If it is possible for Rudy to be in a mental state S that amounts to accepting all of the premises of an instance of (ω), there is no reason that S can't cause Rudy to go into the mental state S^* in

a way that plays the inference role, where S^* is the state of continuing to accept all of those premises while also accepting the conclusion of the rule. This is true at least when the premises themselves can be enumerated, since that seems required by the inference role's requirement of a disposition to cite the premises as reasons for the conclusion (2.VI). Aside from this, if we can, in principle, accept infinitely many premises, then that state can enter into causal transitions that play the inference role.

In short, and contrary to what nearly everyone seems to think, beings like us *can* follow the infinitary omega rule. But to show that something is possible is not to show that it is actual. In the next section I'll argue that we do, in fact, follow the omega rule.

VIII Supertasks and Mental Magic

A particularly clear illustration that we follow the omega rule is provided by supertask computation. These are computations that—from an outside perspective—run through each and every natural number, checking it for some property, finishing the entire check in a finite amount of time.[53] The Berry argument discussed in section IV appealed to the possibility of supertasks. For my purposes here it doesn't actually matter whether these supertask computations are truly possible, but let's suppose that they are for the time being.

Say that we set such a computation to check each number to see if it satisfies the conditions of Goldbach's conjecture (which will happen if the number is odd, less than 4, or the sum of two possibly repeated prime numbers). Let us imagine that the Goldbach computation finishes by falling into a singularity or otherwise being destroyed at the end of one minute of time.[54] This means that the minute expires without the computation halting and spitting out a counterexample.

In this situation, it would be justifiable to accept, of each particular number n, that n is not a counterexample to Goldbach's conjecture and then, after one minute, and on this basis, to accept that for all numbers x, x is not a counterexample to Goldbach's conjecture. This is clearly what we are disposed to do, in this and any other structurally identical case. We don't here reason from a general claim

to particular claims about numbers, but instead from all of the particular claims to a general claim, exactly as in (*ω*).

Some may challenge this, arguing that here we are only *complying* with the omega rule without *inferring according to it*. This would be the case if, for example, in this type of case what really happened is that we accepted the *conclusion* of the rule first and, on that basis, accepted each of the infinitely many premises. But I claim that this is not what happens, because there is a gap here that (*ω*) is needed to fill. To illustrate this gap, imagine that, alongside of us, there is a Venusian number theorist. Venusian number theory is like our own, founded on the Peano rules, but in addition let's assume that Venusian arithmetic is *ω*-inconsistent—for some Π_1 claims in arithmetic, every instance is true in the Venusian language but the negation of the generalization is true. Let's also assume that the Venusians take the negation of Goldbach's conjecture as an axiom in their arithmetic. Goldbach's conjecture is a Π_1-claim, and since the Peano rules are Σ_1-complete, the Venusians will agree with us about the truth of all instances of Goldbach's conjecture. Unlike us then, the Venusians do not follow the omega rule. Because of the weird features of Venusian arithmetic, we must be careful not to supplement it with what are, to us, quite natural principles (like uniform reflection). That can lead to inconsistency.

Let's assume again that our supertask computation finishes without halting. Up to a point, we would both react the very same way to this bit of news, by adding to our stock of beliefs each instance of Goldbach's conjecture. Venusian arithmetic might be shown inconsistent by a finite proof of Goldbach's conjecture, but it is not shown inconsistent by the supertask computation, given the assumptions we've made. So at the point of the non-halt, both can be described as accepting all of the infinitely many premises:

$$\mathrm{GB}(0), \mathrm{GB}(1), \mathrm{GB}(2), \mathrm{GB}(3), \ldots$$

But *on this basis*, we—but not the Venusian—go on to accept "$\forall x(\mathrm{N}x \rightarrow \mathrm{GB}(x))$". There is a gap between these two states, but the gap is bridged by (*ω*). The fact that there is a gap here illustrates why the

best account of what is happening is that we reason to the generalization by way of (*ω*), not from the generalization to the instances. At least in cases like this. The reasoning here isn't based on mathematical induction. In fact, the omega rule itself is needed to secure the universal premise of any relevant instance of induction.

As I said at the outset, it does not matter if the imagined supertask computation is truly physically possible. The point here still stands. For even if supertask computation isn't possible, it is clearly possible for us to believe that it is. If we were being tricked by a mad scientist about the reality of the supertask computations checking the Goldbach conjecture, our reasoning would be the same. As would the Venusian's. This case and others like it illustrate that we do sometimes infer according to the omega rule.

Nor does it matter that *explicitly describing* the set-up here requires quantifying over all of the numerals, determinately. I am making a claim about the rules we in fact follow. The description I give here is not itself playing some independent explanatory role (for more on this and related points, see 12.III).

This thought experiment illustrates that we can have the disposition to infer according to the omega rule when given the chance, when accepting all of the premises. Together with the discussion at the end of the last section, this shows that it is possible for human beings to have the dispositions that constitute following the omega rule. For most other approaches to the philosophy of mathematics, the question is then whether or not these dispositions to follow the omega rule have anything at all to do with the validity of the omega rule. But for inferentialists, there is no gap between these two things. Given that we *can* follow (*ω*), as unrestricted inferentialists, we can take the Peano rules plus the omega rule to be meaning determining for our arithmetical vocabulary, by **MAC**. All of these rules will then be automatically valid, by the MVC, and so, truth in our arithmetical language coincides with membership in true arithmetic (TA). So arithmetical truth is determinate and bivalent.

I don't think that following the omega rule requires any spooky mental magic, but some challenges to this need to be considered. The challenge of determinacy as posed in 8.IV is to account for the intuitive amount of mathematical determinacy without violating

Cognitive Naturalism from 1.IV. I want to answer some worries about whether I've succeeded in this. Given the varied amounts of idealization involved here, it isn't always easy to make the charge perfectly precise. But clearly if following the omega rule gave us an ability to algorithmically enumerate the non-recursive theory TA, this would violate **Cognitive Naturalism**.

Here is an attempt to argue from an ability to follow (ω) to an ability to enumerate TA. Suppose that Rudy follows (ω) along with the PA-rules. Consider an enumeration of all of the sentences in Rudy's (synchronic) arithmetical language, say in alphabetical order:

$$p_1, p_2, \ldots, p_n, \ldots$$

Imagine that we query Rudy about whether he accepts each of these sentences, in sequence. Since only the sentences of TA hold in all admissible interpretations of the rules Rudy follows, he will say that he accepts a sentence just in case it is in TA, producing the sub-enumeration:

$$p_\alpha, p_\beta, \ldots, p_\eta, \ldots$$

Since every sentence in the language appeared in the original enumeration, this sub-enumeration is a full enumeration of TA, produced by Rudy's brain. But since we know that TA is not recursively enumerable, Rudy must have used non-algorithmic mental powers to produce the enumeration. And the key here was assuming that Rudy follows the omega rule, so that is the assumption that must be rejected.

Obviously though, this argument fails. One issue here is that the dispositionalist account of rule-following and acceptance in chapter 2 does not require that if a sentence "follows" from claims Rudy accepts by rules he follows, then Rudy can divine this in some finite amount of time. In fact, in chapter 2 I took pains to make sure that following a rule, even for finite rules, doesn't require anything like logical omniscience. And what goes for finite rules goes for the omega rule as well. Nothing in my account of rule-following justifies assuming that Rudy can enumerate TA when he follows (ω). So, in gifting Rudy

with a kind of logical omniscience, this argument didn't play fair. Like everyone else, dispositionalists reject the idea that we *automatically*, if not making some mistake, accept everything that follows from things we accept by way of rules that we follow. But even if dispositionalists were forced into this implausible view, the argument would fail. Since Rudy can't complete a supertask, there will be no point at which he as gone through infinitely many premises of any instance of the omega rule when running through the original enumeration. So he can't be asked about the generalization "after" being asked about each of its instances. So for multiple reasons, this argument fails.

In the previous section I argued that, when closely examined, the popular (though largely inchoate) reasons philosophers have for thinking us unable to follow (ω) all fail. Here I have argued that we do follow (ω), in actual fact, and that this doesn't require any spooky mental magic. This means that inferentialists can adopt a variant of Carnap's omega rule–based approach to determinism. But does this obviate the need for open-endedness?

IX Is Open-Endedness Still Needed?

Closing the Peano rules under (ω) results in true arithmetic (see the appendix). This is a key point about determinacy. The central philosophical details of the inferentialist-conventionalist approach will be discussed in the next chapter. The important question now is whether the induction rule ($P5$) is even needed given that we have (ω). Was the discussion of open-endedness a complete dead end?

No it was not. One immediately relevant technical point is that, while (ω) gives us an inferentialist argument that first-order arithmetical truth is negation complete and determinate, and thus that all models of our arithmetic are elementarily equivalent, it does not give us a categorical theory of arithmetic, which is what we were after with the open-endedness arguments. This is because TA itself has nonstandard models.[55] This seems to show that ignoring open-endedness, even with the omega rule included, allows for nonstandard interpretations of arithmetic. These nonstandard models

won't interfere with the first-order notion of arithmetical truth, but as we add richer expressive resources to our language, differences may become evident. The open-endedness–based categoricity argument from section V heads off this possibility.

But (ω) is also a schematic rule, so maybe we can do without induction by using the open-endedness of (ω) to rule out nonstandard models. To work in a language in which a model *M* of TA can be described as "nonstandard" is to work in a context in which we have predicates distinguishing the standard part of *M* from the nonstandard part. Let "SM(x)" be a one-place predicate expressing the "is a standard number of *M*" property in this metatheory. Since (ω) is an open-ended rule schema, we can now consider an instance of (ω) for the predicate "SM(x)" and note that zero is standard, one is standard, and so forth and so on. From this, it follows that every "number" is standard ("$\forall x$(N$x \rightarrow$SM(x))"), but, by assumption, this is false when *M* is taken as our arithmetical domain. So if vindicating the open-ended, infinitary omega rule is a constraint on admissible interpretations of arithmetic, then nonstandard model *M* is inadmissible. And since the only assumption we made about *M* was that it was nonstandard, this argument generalizes to show that any nonstandard interpretation of our arithmetical practice is inadmissible. So if categoricity is all that we are after, then we don't *obviously* need an induction rule, though we do still need open-endedness.

Metaphysically, open-endedness is required to constrain the interpretations of our arithmetical language in a way that can become relevant as we move beyond first-order arithmetic. This is a point about metaphysics; in 11.VI I will also show that a proper account of the epistemology of arithmetic still requires the open-ended induction rule, not just the open-ended version of (ω). So we are stuck with open-ended induction.

It is a good thing, too. For when we move beyond arithmetic to mathematical truth quite generally, the need for open-ended schematic but non-infinitary rules becomes apparent. Many branches of mathematics, such as set theory, deal with more objects than can possibly be put into a canonical infinite list. If our language is countable, there will be real numbers and sets that are unnamed in our language. We could attempt to add more complex infinitary rules,

with more premises, but that soon becomes unwieldy (to say the least). I think that we can follow the omega rule, but I don't think that we can follow rules with uncountably many premises—to say nothing of rules with as many premises as there are sets. One reason is that, as noted in section VII, inferring requires dispositions to cite the premises, and this plausibly requires that the rule's premises are enumerable. This is true for the omega rule, but not for many other infinite rules that we might dream up.

To illustrate how the inferentialist approach generalizes beyond arithmetic, consider standard Zermelo-Fraenkel set theory (ZF).[56] Like PA, ZF is a schematic, first-order theory; and like PA, the ZF axioms can be reformulated into natural deduction rules for a full set-theoretic language. Let "Set" be our one-place set predicate and let "$\in$" be our two-place membership predicate taking set terms on the right. For convenience let's imagine that we have defined terms for the empty set ("$\varnothing$"), the subset relation ("$\subseteq$"), the term forming operation for turning lists into sets $(\{\ldots\})$, and a union function ("$\cup$") using standard definitions. The ZF axioms give way to the following *Zermelo-Fraenkel rules* or ZF-rules:

$$(\textit{Existence})\ \frac{}{\exists x \mathrm{Set}(x)} \qquad (\textit{Extensionality})\ \frac{\mathrm{Set}(x) \quad \mathrm{Set}(y) \quad \forall z(z \in x \leftrightarrow z \in y)}{x = y}$$

$$(\textit{Foundation})\ \frac{\mathrm{Set}(x) \quad \exists z(z \in x)}{\exists y(\mathrm{Set}(y) \wedge y \in x \wedge \neg \exists z(z \in x \wedge z \in y))}$$

$$(\textit{Separation})\ \frac{\mathrm{Set}(z)}{\exists y \forall x(\mathrm{Set}(y) \wedge (x \in y \leftrightarrow (x \in z \wedge \phi)))}$$

$$(\textit{Pairing})\ \frac{}{\exists z(\mathrm{Set}(z) \wedge x \in z \wedge y \in z)}$$

$$(\textit{Union})\ \frac{\mathrm{Set}(x)}{\exists y \forall z \forall w((w \in z \wedge z \in x) \rightarrow w \in y)}$$

$$(\textit{Infinity})\ \frac{}{\exists x(\varnothing \in x \wedge \forall y(y \in x \rightarrow x \cup \{x\} \in x))}$$

$$(\textit{Power})\ \frac{\mathrm{Set}(x)}{\exists y \forall z((\mathrm{Set}(y) \wedge z \subseteq x) \rightarrow z \in y)}$$

$$(\textit{Replacement})\ \frac{\mathrm{Set}(w) \quad x \in w}{\exists y \phi(x,y) \rightarrow \exists u \forall y(\mathrm{Set}(u) \wedge y \in u \leftrightarrow \exists x(x \in w \rightarrow \phi(x,y)))}$$

In (*Replacement*) it is required that u can't be free in $\phi(x,y)$. Just as with the induction rule, (*P*5), the rules (*Separation*) and (*Replacement*) are actually rule schemas, and should be understood in an open-ended fashion, continuing to hold through well-behaved expansions of our language.

We saw in section III that versions of Dedekind's categoricity theorem for PA can be adapted to the open-ended Peano rules. Similarly, a version of Zermelo's *quasi-categoricity* theorem for second-order ZF can be adapted to the open-ended ZF-rules. Any two interpretations of ZF set theory are either isomorphic or one is isomorphic to an initial segment of the other that meets certain conditions. This isn't a full categoricity theorem, but it is close.[57]

Some philosophers have argued that this type of result, while not quite showing that set-theoretic truth is *determinate*, at least shows that the continuum hypothesis and many other famous set-theoretic undecidables have determinate truth values. They think this, presumably, because the continuum hypothesis concerns only a very low portion of the universe of sets. I am not quite sure what to think about this matter. I worry that a variant of the concerns from section VI apply here as well. In fact, the danger is even more acute here, since models are sets. More generally it is unclear how much weight should be put on results about set-models of set theory, or even class model results. At the very least, I think we should be cautious. If interpretations of set theory are themselves indeterminate with respect to the continuum hypothesis, then categoricity-style arguments don't seem to give us a reason to think there is truly a determinate fact of the matter about the size of the continuum.

Ultimately I'm unsure how much indeterminacy there is in set theory. I'm inclined to think that there is quite a lot, but I don't want to stake much on this. My position on arithmetical truth is different. I suspect that arithmetic is determinate, at least for first-order sentences. I think that we should be conventionalists regardless, but the determinacy of basic arithmetic is quite intuitive. This chapter has tried to show how rules that we could actually follow might account for this. This completes my response to the second major problem for extending conventionalism from logic to mathematics proper. In the next chapter I will bring everything together to argue from

mathematical inferentialism to mathematical conventionalism, while also fleshing out the resulting conventionalist theory of mathematics.

≈≈≈≈≈≈≈

WITH THE TWO major obstacles overcome, we can now turn to more general matters, showing how unrestricted inferentialism leads to mathematical conventionalism, even despite the puzzles of ontology and determinacy. According to unrestricted inferentialists, the basic rules of our mathematical language are meaning constituting, in whatever form they come, by **MAC**, and automatically valid, by the MVC. Mathematical truth, like logical truth, is a reflection of our linguistic rules. Bigger, scarier, and more ornate, but still just shadows of syntax. Or so I will argue.

Appendix: The Omega Rule and True Arithmetic

Instead of working with the Peano rules augmented with the omega rule, I'm going to consider the much weaker theory starting with Robinson arithmetic (Q). We want to show that, in some sense, the "closure" of Q under the omega rule results in true arithmetic. The central idea of almost any proof of this result is that Q is arithmetically sound, and so is its closure under the omega rule, and this closure is also negation complete.

If there is a standard way of proving this result, it involves indexing proofs with infinite ordinal numbers. Basically this allows proofs to be infinitely "long", but the result can be proved without this (at the price of some additional complications in the proof of lemma 1 below). With this in mind, I'll define the closure of a set Γ as *cl*(Γ) = Γ where "$\vdash$" is standard first-order provability and define the unary ω-closure of a set Γ as $\omega(\Gamma) = \{\phi \mid \Gamma \vdash_{1\omega} \phi\}$, where $\Gamma \vdash_{1\omega} \phi$ just in case $\phi = \forall x \psi(x)$ and $\forall n \in \mathbb{N}$, $\psi(n) \in \Gamma$. In effect, the unary ω-closure of a set of sentences Γ is the result of adding to Γ all sentences provable from Γ by one use of the ω-rule. Define the theory Q^{ω} as follows:

$$\mathrm{Q}^0 = \mathrm{Q}$$

$$\mathrm{Q}^{n+1} = cl(\omega(\mathrm{Q}^n))$$

$$\mathrm{Q}^\omega = \bigcup_{n\in\mathbb{N}} \mathrm{Q}^n$$

A series of lemmas establishes several important results about Q^ω, building to the result that Q^ω is true arithmetic—TA, the theory of all first-order sentences true in the standard model of arithmetic, $\mathbb{N}$. I think that these definitions more closely correspond to use of the omega rule in reasoning than more general, ordinal-based accounts.

Lemma 1. Q^ω *is closed under the ω-rule—if* $\forall n \in \mathbb{N}$, $\varphi(n) \in \mathrm{Q}^\omega$ *then* $\forall x\varphi(x) \in \mathrm{Q}^\omega$

Proof. If all numerical instances of a one-place formula are members of the theory at level *n* of the sequence of Q-theories, then the generalization of the formula is in the theory at level *n*+1 by definition. So the only way this lemma could be false is if there is a formula each instance of which is in the theory at *some* finite level of the Q-sequence, but there is no finite level containing *all* such instances. In effect, instances of the formula must be added indefinitely, as we ascend levels. But this is impossible, because Q itself is negation complete for atomic sentences, so an inductive argument ascending levels of the arithmetical hierarchy of sentences shows that if each arithmetical instance of a formula is in some theory in the Q-sequence, then every instance is in the theory at some particular finite level of the Q-sequence. This holds not just for one-place formulas, but also for *k*-place formulas, for any finite *k*. __

Lemma 2. Q^ω *is sound with respect to the standard model of arithmetic—if* $\phi \in \mathrm{Q}^\omega$ *then* $\mathbb{N}\mathbb{N}\phi$

Proof. Assuming consistency, this follows by definition since every axiom of Q is true in $\mathbb{N}$, and all logical rules, as well as the omega rule, preserve truth in $\mathbb{N}$. __

Lemma 3. Q^{ω} *is negation complete—for all* ϕ*, either* $\phi \in Q^{\omega}$ *or* $\neg\phi \in Q^{\omega}$

Proof. As mentioned in the proof of lemma 1, Q itself is negation complete for atomic sentences. The remainder of the proof of this lemma proceeds via a standard induction on the complexity of formulas. The induction will use lemma 1 in establishing the result for quantified formulas. __

Theorem. Q^{ω} *is True Arithmetic*—$Q^{\omega} = TA = Th(\mathbb{N})$

Proof. By lemmas 2 and 3. __

---

1 See Dummett (1975) on Wang's paradox. In the next section I will provide a more precise account of the type of arithmetical vagueness that I mean to rule out.

2 See also Warren and Waxman (forthcoming *a*).

3 Williamson (1994) and Horwich (2005) see vagueness-based indeterminacy as epistemic.

4 This is roughly equivalent to notions of "supertruth" as used in supervaluationist theories; again see Mehlberg (1958), Van Fraassen (1966), Fine (1975), and Williamson (1994), chapter 5.

5 See Feferman (1991), Field (2001), Harris (1982), Lavine (2001), Lavine (2006), McGee (1997), McGee (2000), McGee (2006), Parsons (2001), Shapiro (1991), and Williamson (2006*a*). Lavine (1994) discusses the historical antecedents of the idea of open-endedness.

6 Shoenfield (1967), page 204.

7 McGee (2006), page 187.

8 This approach is taken in McGee (2000).

9 Introduced in Henkin (1950); for details see Shapiro (1991).

10 See Putnam (1980) and Weston (1976).

11 See Restall (2008) for discussion.

12 This approach is called "second-order-real-variable logic" or "sorv logic" in Button and Walsh (2015).

[13] This distinction is from Russell (1908); see also Lavine (1994).

[14] See Fine (1975) for penumbral connections.

[15] Dedekind's categoricity theorem was first proved in his (1888). See Parsons (2001) and Lavine (2001) for their approaches.

[16] Thinking of "models" or "interpretations" as predicates is analogous to the way that "class models" are usually understood in set theory, see Kunen (1980), page 112, for example.

[17] It is called an "internal" categoricity result in Walmsley (2002).

[18] In the appendix to chapter 12 of Field (2001).

[19] Parsons (2001), pages 15–16, responds to Field by appealing to a presumption of homophonic translation.

[20] Both Field (2001) and Parsons (2001) discuss some strategies. Even more options are discussed in some of my unpublished papers.

[21] See again the appendix to chapter 12 of Field (2001).

[22] McGee (1997) and Field (2001).

[23] Similar reconstructions of McGee's argument are also given by Field (2001) and Berry (2017).

[24] See Kaye (1991), pages 70–71.

[25] Field (2001), pages 355–356.

[26] Field (1994*b*), (1998*a*), and (1998*b*).

[27] If a "finitely many" quantifier, , was determinate, we could determinately axiomatize arithmetic by adding to PA the axiom "$\forall xy(y < x)$"; in fact, this is the approach that Field actually takes, rather than dealing with open-endedness directly.

[28] Parsons (2001), page 22.

[29] Also pressing is the potential falsity of Field's first cosmological assumption. Some of Field's critics, including Putnam (2015*b*), have seen Field as relying on a determinate grasp of finiteness in stating his cosmological assumptions. But this is a mistake. The statement itself isn't doing any real work, it is the fact that the world is some particular way. Sharon Berry noted to me that a Field-McGee hybrid approach could also be reformulated so that the cosmological assumptions were modal; I discuss Berry's own distinct but directly modal variation on McGee's approach below.

[30] See Berry (2017).

[31] See Davies (2001) and Earman and Norton (1993) and (1996).

[32] One possibility would be to use reference magnetism based on intrinsic naturalness in the manner of Lewis (1984), Merrill (1980), and Sider (2011), this time applied modally. I think that all of these approaches appeal to anti-naturalist notions (1.IV).

[33] I am not denying that assuming standard reasoning linking generalizations to instances, given this initial indeterminacy, some "heads" sequences will contain coinflips that are, intuitively, tails within their nonstandard portions (assuming the other vocabulary's interpretation is fixed).

[34] Berry's approach is also threatened by the type of worries that Parsons raised for Field's approach, I think, but I won't press this here.

[35] Parsons (2001), pages 19–20.

[36] This can be used as the premise in a conceivability to possibility argument—see Chalmers (2002) and Yablo (1993)—but I won't dwell on this here, since it is likely to be distracting and potentially confusing (neither I, nor Parsons presumably, are arguing that the "God" or Oracle who interprets our arithmetic nonstandardly is possible in any robust sense).

[37] In unpublished work I presented the same argument using God (following Parsons). Unfortunately this led many early readers of those papers to assume that I was assuming the "metaphysical possibility" of God. This misunderstanding persisted no matter how explicit and tedious my protests became, so I have switched to a less distracting illustration here. I have some work in progress in which I develop some of the ideas of this chapter without assuming conventionalism. One of these papers focuses only on the argument of this section. In addition, another unpublished paper develops the discussion of the omega rule later in this chapter. I feel the need to mention this here partly because a footnote in my (2015*a*) referenced a paper then under review which is yet to appear. The front portion of the paper formed part of what was developed into Warren and Waxman (forthcoming *a*). I am comforted somewhat by the fact that my (2015*a*) and (2015*b*) also promised that the book you are now reading was "forthcoming" (in the former under an earlier title), so I am (very slowly) keeping my promises.

[38] In Putnam (1975).

[39] For end-of-inquiry approaches to truth and other semantic notions, see Peirce (1878) as well as, apparently, Putnam (1981).

[40] In Hamkins and Yang (2014). This is perhaps related to concerns about how the quasi-categoricity of Zermelo (1930) impact the notion of set-theoretic truth—see the discussion in section IX.

[41] This is related to a point that is accepted by non-conventionalists as well, see Dummett (1963) and Parsons (2001).

[42] The rule was first introduced in Hilbert (1931*a*) and (1931*b*). There Hilbert puzzlingly claims that the rule is finitely acceptable, but in writing this Hilbert was perhaps operating under the mistaken belief that Ackermann had shown number theory to be consistent using finitely acceptable methods; see Buldt (2004) for some discussion of this. Section 10.2 of Potter (2000) offers a different diagnosis of Hilbert's optimism. In his (1933*b*) Tarski claims that he introduced the omega rule in lectures given in 1927.

[43] See Carnap (1934) and (1937). Carnap had a different approach to analyticity for his Language II; see Coffa (1987) for discussion.

[44] Tarski (1933*b*), quoted from page 295 of Tarski (1956).

[45] Smith (2013), footnote on page 22.

[46] See Thomson (1954) for a flawed argument that all supertasks are impossible and Benacerraf (1962) for a compelling response.

[47] Chalmers (2012) appeals to the omega rule as an idealization by imagining a being with infinite parallel reasoning capacities, so that each premise can be evaluated at once; Weir (2010) seems to argue that there is no real difference between the idealization involved in astoundingly large finite proofs and infinitely long proofs.

[48] See Raatikainen (2005) for some discussion.

[49] Quoted from Carnap (1937), page 173.

[50] This interpretation is offered in Coffa (1991), for example, in note 3 to chapter 16.

[51] 12.III will return to this and related themes.

[52] At the end of my (2015*a*).

[53] Many of the classic essays on supertasks are collected in Salmon (1970). See Davies (2001) for a proof that supertask computations are possible in continuous Newtonian universes and Earman and Norton (1993) and (1996) and Hogarth (2004) for discussion of socalled Malament-Hogarth machines as allowed in some relativistic spacetimes.

[54] The idea of using supertask computations in the philosophy of mathematics is hinted at (positively) in the introduction to Benacerraf and Putnam (1983) and (negatively) by Weyl (1949). More recently Berry (2014) has attempted to put supertask computations to work in making an epistemological criticism of Tait's (2001) axiomatic conception of mathematics. Warren and Waxman (forthcoming *b*) argue that these computations are no help in explaining determinacy without something like the omega rule being assumed in the background. Here I use them for a different purpose—to illustrate that we do follow the omega rule.

[55] That true arithmetic has nonstandard models is proved in most graduate logic texts, see chapter 7 of Bell and Machover (1977), for example. It was first proved in Skolem (1934).

[56] For ZF see any standard set theory textbook, for example, Kunen (1980).

[57] See Weston (1976) for discussion. McGee (1997) gives a full categoricity proof for set theory in an open-ended setting, but his result requires both a set urelemente axiom and the assumption that the quantifiers range over absolutely everything (this guarantees homophonic translation between the quantifiers of any two set languages).

11

Truth, Paradoxes, Freedom, Applications, and Knowledge

In 8.II I noted that a simple version of inferentialism about mathematics along the lines of part II's theory of logic faced at least two major obstacles. In chapters 9 and 10, respectively, I discussed how to overcome those obstacles. This implicitly involved a move from *simple* mathematical inferentialism to *sophisticated* mathematical inferentialism. In the case of arithmetic, this required augmenting the Peano rules in order to deal with both ontology and truth.

These augmentations have been developed and defended at length in the previous two chapters, but it is important to bring everything together and to show that sophisticated mathematical inferentialism leads to mathematical conventionalism, just as logical inferentialism led to logical conventionalism. In logic this move was fairly direct, but in mathematics the various subtleties introduced make the move to conventionalism more complicated. Once these points are accommodated, a number of general issues in the philosophy of mathematics also need to be addressed on conventionalist grounds, including the set-theoretic paradoxes, the freedom of mathematics, the abundant applications of mathematics to the physical world, and the epistemology of mathematics and consistency.

I From Inferentialism to Conventionalism (Again)

Let's bring together the discussions of chapters 8, 9 and 10. For any given branch of mathematics M, the axioms of M, in natural deduction form, function as meaning-constituting rules for the

primitive terminology of M. Assuming Inferentialism as applied to *mathematics*, by an extension of the MAC principle, unrestricted inferentialists can say this for *any* M-rules that we follow, so can apply this approach to all branches of mathematics. By the **MVC**, the M-rules will be strongly valid. And by Totality (applied to mathematical inferences), all M-validity comes only from the automatic validity of the M-rules. Since mathematical truth is necessary, for any ϕ in the language of M, ϕ is true just in case ϕ is necessarily true. This means that, since the argument with no premises and ϕ as a conclusion will be valid whenever ϕ is true, the M-rules alone explain the truth (and necessary truth) of ϕ. As with logic proper, unrestricted mathematical inferentialism is a version of mathematical conventionalism (1.II). The automatic validity of the basic mathematical rules is the source of all mathematical truth.

Various complications have been suppressed in this big-picture sketch. Let's dig into the details for the fundamental case of arithmetic. The strategy of 8.II was to read the arithmetic rules off of first-order PA, the most widely used axiom system for arithmetic. But this didn't lead to a satisfying account of arithmetical truth. So as chapter 10 discussed, the syntactic Peano rules must be understood in a fully open-ended fashion. I argued that this leads to a categorical conception of arithmetic (10.V). But because of some general uneasiness about the connection between categoricity and determinacy (10.VI), I further augmented the PA-rules with the syntactic but infinitary omega rule (ω) and argued that, despite what is usually thought, beings like us can and do follow the omega rule (10.VII and 10.VIII). I'll denote the combination of the open-ended Peano rules with (ω) by "PR$^{\omega}$". Sophisticated arithmetical inferentialism claims that the PR$^{\omega}$ rules are the basic, meaning-determining rules for our arithmetical vocabulary (N,S,0).

What I said for the Peano rules in 8.II goes for PR$^{\omega}$ as well—I am not making a precise historical claim about the actual origins of arithmetic. Our arithmetical concepts grew up slowly, bit by bit, arising out of our activities of counting and making cardinality comparisons. For our current abstract understanding of pure arithmetic though, the PR$^{\omega}$ rules are descriptively accurate. If they differ from the basic rules governing our arithmetical language, they

don't do so in any substantial way, or so I claim. A full language based on the PR^{ω} rules is, by and large, a descriptively accurate account of modern-day natural language arithmetic.

With the metaontology developed in chapter 9, we don't need to worry too much about the fact that the PR^{ω} rules ontologically commit us to infinitely many numbers. This is no barrier at all to unrestricted inferentialism. By the MVC any sentence that follows from the PR^{ω} rules is automatically true, even sentences expressing ontological commitments. For inferentialists, the rules are explanatory, with no hostage left to the world. So the infinite realms of mathematical ontology are utterly non-mysterious, once the lessons of chapter 9 are heeded.

Turning to arithmetical truth, we can say that a sentence in the (first-order) language of arithmetic is true just in case it "follows" from the PR^{ω} rules. This parallels the case of logic, with an important added wrinkle. Logical truth coincides with finite provability from logical rules (4.I and 4.III). But the PR^{ω} rules include the infinite omega rule, so some true sentences in any timeslice of our language will not be finitely provable in that timeslice (see 10.II's discussion of "synchronic" arithmetical languages). My preferred way of talking here, presaged back in 3.I's discussion of Totality, is to say that a sentence ϕ in the language of arithmetic is *made true* or *forced to be true* by the PR^{ω} rules just in case, the automatic validity of the PR^{ω} rules fully determines that ϕ is true.[1] In other words: a sentence ϕ in the language of arithmetic is true just in case the automatic validity of the PR^{ω} rules requires ϕ to be true. This is extensionally equivalent to saying (1) that ϕ holds in all admissible models of PR^{ω}, (2) ϕ is "provable" from PR^{ω} in a proof system that allows proofs to be infinite either in length or in depth, or (3) that ϕ is a logical consequence of of the PR^{ω} rules.

The PR^{ω} rules lead to a complete and categorical first-order theory of arithmetic. Categoricity is taken care of by open-endedness, completeness by the omega rule. The usual proofs that closing weak arithmetical theories under (ω) result in true arithmetic (TA) involve indexing proofs by infinite ordinals, in effect allowing them to be infinitely long. The proof of this result in the appendix to

chapter 10 doesn't require this. That proof can be adapted to a proof system that allows infinitely "wide" natural deduction rules like (ω) to figure in proofs.[2] The only source of arithmetical truth is the automatic validity of the PR^{ω} rules. A sentence in the language of arithmetic is true just in case its truth is explained by the automatic validity of the PR^{ω}rules. Likewise for arithmetical falsity. For the logical truths, a sentence's truth is explained by the automatic validity of the basic rules just in case it is finitely provable using those rules (4.I and 4.III). For the arithmetical truths, *finite* provability *in a given timeslice* of our language is not enough. Some true sentences won't be finitely provable in the current timeslice of our arithmetical language, only in well-behaved expansions. But their truth, even in the current timeslice, is forced by the automatic validity of the open-ended PR^{ω} rules.

This account of arithmetical truth easily applies to the Gödel sentence for our current (synchronic) arithmetical language. Let A be the arithmetical theory induced by our finite commitments in the current timeslice of our language, so that A is recursive and so has its own Gödel sentence, G_A. Assuming the consistency of A, there is no proof of G_A available in A, but by construction, G_A is a Π_1-sentence that "says" of all numbers; that they don't code for a proof of a certain sentence in A. The instances of this generalization are Σ_1-sentences saying, for example, "3" doesn't code for a proof of this claim in A. But since A is determined by finite provability from the PA-rules in a fixed language, and PA is Σ_1-complete, each instance of G_A is provable in A. So by (ω), G_A must be true in A even though it is unprovable in A. If it was not, the (ω) would be invalid. And since we made no special assumptions about A, this argument goes through for the Gödel sentence of any timeslice of our arithmetical language. It also works for the consistency sentence of any timeslice of our language too.

This establishes truth, but what about provability? While we can't always produce a finite proof of a given true claim in a given timeslice, we can always produce a finite proof of a true claim in *some* well-behaved expansion of our language. Of course, this is very different than saying there is some well-behaved expansion

where *every* true claim is provable (10.V). This difference is key to understanding why my claims here don't run afoul of the **Cognitive Naturalism** principle (1.IV). Our general assurance that this notion of truth is determinate comes by way of more general arguments, involving constraints on interpretation and infinitary rules. Worries abound about these assurances. Some of them are more technical and more pointed variants of the general worries discussed in 7.IV—worries about the role of syntax, consistency, the appeal to standard mathematical results in my discussion, and the ever-looming threat of circularity. These worries will be addressed at length throughout chapter 12, but especially in 12.III. I leave them aside for now; modulo these worries, we can see that the sophisticated inferentialist approach to arithmetic is a version of arithmetical conventionalism.

An arithmetical sentence is made true by the automatic validity of the meaning-determining rules. There is no appeal to an independent Platonic realm of ethereal mathematical facts. The only source of arithmetical truth is the automatic validity of the meaning-determining rules of our arithmetical language. The only source of mathematical truth is the automatic validity of the meaning-constituting rules of our overall mathematical language. The resulting theory of arithmetic accounts for the existence of numbers without violating **Metaphysical Naturalism**, and accounts for the determinacy of arithmetical truth without violating **Cognitive Naturalism**. Mathematical facts are byproducts of the conventions of our language, and can be fully integrated into a scientific picture of the world.

As with logic, I think that conventionalism requires *unrestricted* inferentialism (though see 4.II for some subtleties). This point is even less controversial in the case of mathematics than it was in logic. Most restricted inferentialists will have long parted ways, since the power of mathematical theories makes them fail nearly every restriction on the **MAC** principle that has ever been proposed. Even the PA-rules by themselves entail every first-order counting sentence, so they are wildly non-conservative over most base theories, in the standard sense. Some restricted inferentialists will be able to follow me along a portion of the path to mathematical conventionalism, but as we consider more and more powerful

mathematical theories, even the most liberal of restricted inferentialists will eventually balk and turn away.

The inferentialist account of logical *understanding* from 3.V can also be extended to mathematics. Appropriately modifying the basic understanding clause gives us:

> **Basic Understanding (BU)**: If someone uses a mathematical expression in accordance with its meaning-constituting rules of inference, then they have a basic understanding of the expression.

As in 3.V, we can build a more sophisticated account of mathematical understanding on top of this principle. Perhaps we would also count someone as understanding arithmetic when they can see what following the PR^{ω} rules would involve. This is fine since **BU** only makes following the rules a sufficient condition of basic understanding, not a necessary one. In any case, the psychological application of mathematical inferentialism to understanding is exactly the same as in the logical case.

Historically, arithmetical conventionalism has been closely associated with *logicism*—the view that mathematics is a species of logic.[3] Logicists typically tried to explicitly define mathematical terminology in logical terms. Arch-conventionalist Carnap often flirted with logicism and saw it as closely related to conventionalism.[4] Though Neo-Fregeanism—sometimes even called neo-*logicism*—is closely related, logicism proper has been almost universally rejected in contemporary philosophy.[5]

If our mathematical rules aren't "logical", then the inferentialist principles from 3.I and 3.II require a slight reformulation in order to apply to mathematics. Is my brand of mathematical conventionalism a version of logicism? In 3.III I offered the following account of logical notions:

> **Logicality**: An expression is a logical constant just in case its meaning-constituting rules are both (*i*) deeply non-empirical and (*ii*) highly topic-neutral.

Using this principle, the question of whether arithmetic is a branch of logic is the question of whether the PR^{ω} rules are deeply non-empirical and highly topic neutral. Obviously, the PR^{ω} rules are non-empirical, so the question is whether or not they are topic neutral. On one fairly natural reading, they are not. They have a very particular topic: numbers and arithmetical properties. But I don't put much stock in this, because the applications out of which arithmetic originally developed are completely topic neutral—we can *count* any objects whatsoever.

So while I think there is a very intuitive sense in which the arithmetical constants are "logical", any dispute on this score seems largely insubstantial. It will turn on whether the subject matter of arithmetic is *the numbers* or whether the numbers are merely a device for facilitating general applications of arithmetic. The most accurate thing to say is that applied arithmetical notions count as "logical", but pure arithmetical notions don't. But since these are the very same notions, it is hard to see one of these descriptions but not the other as objectively correct. Whether or not they technically count as part of logic, just as with logical truths, arithmetical truths in particular and mathematical truths in general are reflections of our rules, yet further shadows of our syntactic conventions. It is just that, in the case of mathematics but not logic, our syntactic conventions are *wider* and *taller* than the finite proofs we construct in particular timeslices of our language. Despite this, they are still rooted in the syntactic rules of inference that we follow.

I won't offer a general definition of "mathematics" or "mathematical" notions. Pure mathematics is a well-segmented human activity. We know it when we see it. There are many possible simple symbol manipulation games that seem of a kind, but that barely deserve to be called "math", if at all. In some ways the mathematical notions are a motley, connected more by matters of historical contingency and cognitive impenetrability than by a shared essence discoverable by the philosopher. I believe that my approach can be applied to give a philosophical account of *any* branch of pure mathematics, though I admit at this level of detail many questions about non-arithmetical cases are being left open. I have left open the exact extent of set-theoretic truth, for example (10.IX).

To make the power of my conventionalist approach clear, I will discuss how to deal with mathematical paradoxes, the freedom of mathematics, applications of mathematics, and mathematical epistemology. Before that though, I want to deal with another purely philosophical objection to my approach. The objection is quite general and, as might have been expected, is inspired by the work of Quine.

II As Always, Carnap versus Quine

Anyone who clearheadedly rejects inferences sanctioned by the PR^{ω} rules must mean something different than what we mean by “N”, “S”, or “0” (assuming they agree with us about non-arithmetical vocabulary). And since the PR^{ω} rules include open-ended induction, anyone who clearheadedly refused to extend induction to a (non-vague) arithmetical predicate would have to mean something at least slightly different than we do by their arithmetical vocabulary.

In arguing that induction is open-ended, I came perilously close to begging many relevant questions. For there are many apparently possible situations where we would refuse to apply induction to a newly introduced predicate. The issue brings us back once again to the heart of the decades-long dispute between Carnap and Quine. Something needs to be said about this issue, lest it be doubted that open-ended induction, and thus the PR^{ω} rules, are truly meaning-constituting for our arithmetical vocabulary.

What are some cases where we might stop applying induction? One type of case is where we simply stop doing any arithmetic at all. Surely that is possible; but it can be left to one side, since there won’t be any issue about whether we mean the same thing by arithmetical terms in our expanded language since the “expanded” language has no arithmetical terms. But even after leaving this aside, there appear to be easy-to-describe possibilities where we continue to do arithmetic while refusing induction on a newly introduced predicate. Here’s a possibility proof involving a fanciful case:

> *The Oracle*: While digging in the sands of Egypt, archeologists discover a statue that speaks. The world is astonished, and modern science is completely unable to explain how the statue does it. What is more amazing is that this statue, whose name in hieroglyphs translates into English as “The Oracle”, answers any question one puts to it, including mathematical questions. One day a team of Princeton mathematicians, after waiting their turn for over a year, ask the Oracle about mathematical induction. Astonishingly, the Oracle tells them that there is an arithmetical predicate “F” such that 0 is F, and F is inherited by successors, but yet there is some number that is not F! This information gradually gets disseminated by philosophers and mathematicians, and everything goes on much as before, except that English speakers, while seemingly fully competent in arithmetic, reject induction on “F”.

Nothing like this is actually possible (presumably). The example is meant merely as a proof of concept showing that there is nothing baldly incoherent about our refusing to extend induction to a predicate. This puts pressure on the idea that open-ended induction is strongly valid.

The problem is obviously not that *The Oracle* is a real example of a failure of induction. Nor is the problem that inferentialist-conventionalists took it to be inconceivable for humans to reject a *sentence* that is syntactically identical to an instance of induction (compare 7.I). Nor is it even that a revised but alternative arithmetical practice is supposed to be inconceivable for conventionalists. Although philosophical history is often written in a way that would suggest otherwise, this last point was clearly recognized by Carnap nearly two decades before Quine’s “Two Dogmas of Empiricism”:

> …all rules are laid down with the reservation that they may be altered as soon as it seems expedient to do so. This applies … also to the L-rules [logico-mathematical rules—JW], including those of mathematics.[6]

So even Carnap, arch-conventionalist of philosophical folklore, agreed that logical and mathematical rules could be given up to accommodate recalcitrant experience. So what then is the difference between Carnap and Quine? It is that Carnap but not Quine distinguishes between revisions that change meaning and those that don't. Here is Carnap illuminating this crucial distinction in a late reply to Quine:

> …I should make a distinction between two kinds of readjustment in the case of a conflict with experience, namely, between a change in the language, and a mere change in or addition of, a truth-value ascribed to an indeterminate statement (i.e., a statement whose truth value is not fixed by the rules of language …). A change of the first kind constitutes a radical alteration, sometimes a revolution, and it occurs only at certain historically decisive points in the development of science. On the other hand, changes of the second kind occur every minute. A change of the first kind constitutes, strictly speaking, a transition from a language L_n to a new language L_{n+1}.[7]

As Carnap notes, mere revisability is a red herring. What truly matters is whether meaning is changed in the revision. Quine often glossed over this because of his infamous skepticism about meaning.[8] But even Quine reinstated basically the *same distinction* in his discussion of norms of translation applied to deviant logics.[9]

The case of *The Oracle* is supposed to be a problem for arithmetical inferentialism because it seems to be a case where our open-ended induction rule is clearheadedly rejected, but without any attendant change in the meaning of the arithmetical vocabulary. But as in my reply to Williamson (7.V), I deny that homophonic translation is appropriate in this case. Whatever intuitive pressure there is toward homophonic translation is being generated by irrelevant features of the case—that we are imagining *ourselves* speaking to the Oracle, for example. Attempts to use *The Oracle* to create a problem for inferentialism about arithmetic commit **The Translation Mistake**.

We imagine that we can translate homophonically between the counterfactual community in *The Oracle* and ourselves because we are treating them like other speakers in our actual community. Being part of a shared social practice comes with a defeasible presumption of homophonic translation. But the mathematical community in *The Oracle* case defers to the Oracle, and so, via the division of linguistic labor, they end up speaking the Oracle's arithmetical language, *not our own*.[10] If we ask ourselves whether we would translate homophonically between the arithmetical language of someone who denied induction for (consistent, non-vague) arithmetical predicate "F" and ourselves, the answer is clear. There is no evidence that could convince us to give up induction, save deference to some authority without any understanding on our part. What is easy to miss in considering these cases is that deference amounts to accepting the authority's language, and so cancels the presumption of homophonic translation that comes from being part of the same linguistic practice.

There are also starker cases we might imagine, where translating homophonically between a given, non-vague expansion of our mathematical language and our current language results in an inconsistency. But this is not something to be feared or shunned, and is, in fact, the key to taming the terrifying paradoxes of set theory that so troubled Hilbert and Russell.

III A Conventionalist Look at the Paradoxes

Early in the history of set theory it was realized that quantifying over *all* sets, ordinal numbers, or cardinal numbers gave rise to paradox. It is worth briefly discussing if and how conventionalism makes sense of this situation, both for the historical paradoxes and for the threat of paradox in contemporary set theory.

The most famous set-theoretic paradox is Russell's paradox (which was independently discovered by Zermelo at around the same time). Russell's paradox arose because many early set theorists accepted, in some fashion, an axiom of naïve comprehension:[11]

$$\text{(NC)} \quad \exists y \forall x (x \in y \leftrightarrow \phi(x))$$

This amounts to assuming that for any particular formula in our language there is a set of all and only the things satisfying the formula, but this leads to a contradiction:

1. $\exists y \forall x (x \in y \leftrightarrow x \notin x)$ [instance of (NC)]
2. $\forall x (x \in \text{r} \leftrightarrow x \notin x)$ [introducing a name witnessing (1)]
3. $(\text{r} \in \text{r} \leftrightarrow \text{r} \notin \text{r})$ [instantiating for "$\forall x$" in (2) with "r" according to standard quantifier rules]
4. $\bot$ [from (3) using classical logic]

The other key paradoxes—Cantor's paradox and the Burali-Forti paradox—are structurally similar to this one.[12] The mathematical community's response to the paradoxes was to reformulate set theory so as to avoid the principles, like (NC), that led to paradoxes. The ZF-rules from 10.IX, which are based on the axioms of modern ZF set theory, don't include or entail (NC) and are (probably) consistent.

But unrestricted inferentialists think that (NC) was, since a meaning-constituting rule of pre-Zermelo set theory (let's assume), automatically valid in the *original* set-theoretic language. Arguably, this means that the language of set theory was originally a trivial language, like the *naïve* tonk language from chapter 5. Or perhaps it was more like a *sophisticated* tonk language, where the inconsistent package of rules is not actually followed in a naïve fashion.[13] In that case, given that Frege and other early set theorists didn't actually have a disposition to apply the rules naïvely, it is not even clear that they actually followed all of those rules together as basic rules of their set-theoretic language. Instead there was probably some indeterminacy about the rules that they actually followed. I don't want to put too much weight on this, since it would quickly turn into a descriptive/historical question about the linguistic dispositions of Frege and other early set theorists. And most of the relevant evidence has been lost to the ravages of time.

A more interesting point is that the metaontology of chapter 9 gives the inferentialist/conventionalist an option that dovetails with

the discussion of the previous section.[14] The conventionalist can make sense of the paradoxes in an ultimately consistent fashion by explaining how a shift in our set-theoretic quantifiers is taking place in any standard argument for paradox. This basic idea, that in the so-called *logical* or *set-theoretic paradoxes* the concepts involved are *indefinitely extensible*, goes back to Bertrand Russell:

> …the contradictions result from the fact that ... there are what we may call *self-reproductive* processes and classes. That is, there are some properties such that, given any class of terms all having such a property, we can always define a new term also having the property in question.[15]

Michael Dummett later gave a similar but more complex characterization:

> What the paradoxes revealed was not the existence of concepts with inconsistent extensions, but of what may be called indefinitely extensible concepts. The concept of ordinal number is a prototypical example. The Burali-Forti paradox ensures that no definite totality comprises everything intuitively recognisable as an ordinal number, where a definite totality is one quantification over which always yields a statement determinately either true or false. For a totality to be definite in this sense, we must have a clear grasp of what it comprises: but, if we have a clear grasp of any totality of ordinals, we thereby have a conception of what is intuitively an ordinal number greater than any member of that totality. Any definite totality of ordinals must therefore be so circumscribed as to forswear comprehensiveness, renouncing any claim to cover all that we might intuitively recognize as being an ordinal. … The intuitive concept of ordinal number, like those of cardinal number and of set, is an indefinitely extensible one.[16]

If the concept of set is indefinitely extensible, then we cannot quantify over absolutely all of the sets. Attempts to do so lead to quantification over a larger totality of sets.

The indefinite extensibility response to the paradoxes says that the issue here is not that (NC) is problematic, but that the Russell set r, introduced at premise (2) in the argument above, lies outside the range of the quantifier "∀*x*" in (2), hence the move from (2) to (3) is invalid. As Russell and Dummett indicate in the quotes above, according to this line of thought, the very attempt to talk about all of the sets (or in this case, all of the non-self membered sets) necessarily involves the "introduction" of a new set not in our original totality. And our inability to quantify over all of the ordinals or cardinals is explained in the same way.

Indefinite extensibility is puzzling on standard approaches to metaontology: how can we be barred from quantifying over all of the sets, as a matter of principle? That would be deeply mysterious, metasemantically.[17] But modest quantifier variance allows us to make good sense of indefinite extensibility, leading to what I have previously called the variance theory of indefinite extensibility, highlighting again the connection to quantifier variance.[18]

The basic idea is that when we attempt to form a conception of the set of all non-self membered sets, we introduce a term "r" for the Russell set and so expand our set-theoretic language from *L* to *L*+. This expansion is not consistent *if* the quantifiers of *L* are translated into those of *L*+, but the change of meaning in this case is quite different than the one described in *The Oracle* case considered in the previous section. In that case there was no obvious way to translate between the two languages. But here the expanded language can easily see the quantifiers of the unexpanded language as a restriction of their own. What the paradox shows is just that the quantifier's meaning has changed in the move from *L* to *L*+. In other words, when we expand our language in the move to generate a paradox, *L* and *L*+ form an instance of quantifier language pluralism or modest quantifier variance. The set-theoretic quantifiers of *L* and *L*+ are closely related, but they mean subtly different things. *L*'s unrestricted quantifiers are properly translated as restrictions of *L*+'s quantifiers.

We can say both that "∀" in *L* is unrestricted and, from the perspective of *L*+, that "∀" in *L* is a restriction of "∀" in *L*+. There is no incoherence here, since these two claims don't contradict each other

when properly understood. The first only claims that there is no quantifier *in L*, for which "∀" in *L* is a restriction, but that is different than the claim that there is no expansion of *L* from the perspective of which "∀" is seen as restricted. The quantifiers are both unrestricted, but not *absolutely unrestricted*, since there are languages from the perspective of which each can be seen as restricted.[19] The claim is not that it is possible for there to be sets that don't currently exist. The sense in which our unrestricted quantifiers are "restricted" concerns expansions of our language. Failure to recognize this leads to muddles (see 7.I). No quantifier is absolutely unrestricted, on this approach; given (NC), the extensibility here is ineliminable. We can continue going on to form ever more expansive set-theoretic quantifiers with no stopping point in sight.

This doesn't require endorsing any egalitarian thesis about this instance of pluralism, and on the metasemantics of **Quantifier Metadeflationism** the indefinite extensibility itself isn't mysterious. In the paradoxical case there is merely a subtle shift in meaning, not a drastic shift. Many of the set-theoretic claims from the unexpanded language can be understood, in a slightly altered fashion, within the expanded language. In this type of case I say that the rules are expansive—they force a rejection of homophonic translation on pain of contradiction, but allow a quantifier-restriction translation. Rules for set theory based on **NC** are expansive, but the ZF-rules are not. There is no deep mystery about expansive rules, for the inferentialist-conventionalist mathematical truth and ontology are both explained in terms of the rules of our mathematical language rather than vice versa.

The crucial point is that paradoxes do not undermine conventionalism. In any given language, even given indefinite extensibility, set-theoretic truth is simply a byproduct of our basic set-theoretic rules of inference. Even paradoxical mathematical principles don't undermine conventionalism, and the metaontology of chapter 9 neatly provides us with the resources for making sense of standard set-theoretic paradoxes without any serious difficulties. Presumably current mathematics has purged the paradoxes forever by moving to the ZF-rules, but that wasn't really necessary, metaphysically. Inferentialists can make good sense of our current

set theory while also making good sense of those, perhaps including Russell and Dummett, who favored a different philosophical reaction to the paradoxes. This is an extremely satisfying theoretical situation to be in with respect to the set-theoretic paradoxes, and it is open to conventionalists and no others.

Mathematical paradoxes, whether historical or waiting to be found, pose no insuperable difficulties for mathematical inferentialism or conventionalism. Conventionalists can make good sense of many different mathematical theories, including inconsistent theories. This point is related to explaining another key feature of mathematics that Platonists find inexplicable: the *freedom* enjoyed by mathematicians.

IV A Conventionalist Look at Mathematical Freedom

Mathematicians enjoy a type of freedom that scientists do not. A mathematician is free to set up any formalism he may like. If a hot-off-the-press mathematics paper starts by saying “define a Carnapo as follows ...”, nobody would so much as bat an eye. Nor would they object that no evidence had been provided of the *existence* of Carnapos. This illustrates the widely recognized fact that mathematicians are entirely free to set up any mathematical formalism they may like, while scientists and others do not have analogous freedoms—giving a name to a new type of particle is fine, but scientists would need to produce direct or indirect evidence for the particle’s existence in order for it to be accepted. There is nothing like this evidential burden in mathematics.

Freedom poses no problem for conventionalists. Like the logical conventionalism of part II, the mathematical conventionalism of part III is a version of unrestricted inferentialism. Any rules of inference that may be followed for “Carnapos”, or other newly introduced mathematical terms, are automatically valid. This means that the sociological fact of freedom in mathematical practice leads to the normative freedom of mathematical practice. Unrestricted inferentialism is what explains the freedom of mathematics. Those who do not endorse unrestricted inferentialism—which is to say

nearly everyone who has ever written about the topic—won't be able to explain the full nature and extent of the freedom enjoyed by mathematicians in their theorizing.[20]

To say that mathematicians enjoy freedom is not, of course, to say that there are no constraints on our mathematical practices. Standard mathematical practice is normed by classical logic. Here, unlike with natural languages, very few would dispute this descriptive claim. Given classical logic, if a new mathematical formalism turns out to be inconsistent, it is also trivial, which makes it mathematically uninteresting if interpreted standardly. Unfortunately, unintended inconsistency has plagued many proposed theories—Frege, Church, Quine, Reinhardt, and other great logicians have all created systems that turned out to be inconsistent.[21] So it goes. Conventionalists can make sense of this given the triviality induced by inconsistency in classical settings, but they—unlike any competing views—can also make sense of *inconsistent* mathematics being permissible, in other contexts, where a paraconsistent logic is being used, for example.

For completely new types of mathematical objects consistency is the main practical constraint on actual mathematical practice, given classical logic. But in many standard cases "new" mathematical objects are simply defined in terms of old ones, like when a mathematician says "define a Carnapo to be a set meeting the following six conditions ...". To determine whether or not a Carnapo exists, in this type of case, appeal is made to our previously understood rules for reasoning with "set" terminology. If the set-rules guarantee that a set meeting the six conditions in question exists, then Carnapos exist, not otherwise. Conventionalists can fully account for this, since it is the *basic* rules of the language that are meaning constituting. If some new bit of mathematical terminology is defined in terms of old mathematical terminology, then no new basic rules have been introduced, and so you must use the rules for the old vocabulary to see if there is anything answering to the given definition. This is all as it should be and harmonizes completely with mathematics as actually practiced.

Some cases are on the borderline, like when new axioms for "set" talk are proposed that differ from standard set theory (for just one example, consider Quine's system NF).[22] We can call differently

defined objects “sets” without using our pre-existing rules for “set”-talk. But we would balk at calling some newly defined type of object a “set” if the new “sets” were too different from our already understood concept of a set. We can use the notions introduced in 5.III to make this precise: we think a newly introduced type of mathematical object can share a name with an already understood type of object if the new notion and the old are semantic counterparts, not otherwise. Newly introduced but set-like things can be called “sets”, but not newly introduced non-set-like things. This constraint is, of course, highly informal and not always rigidly applied, but it nonetheless explains these apparent borderline cases perfectly.[23]

Traditional versions of mathematical Platonism are helpless in the face of mathematical freedom. If mathematical objects are explanatorily prior to our mathematical practices, then we can only truly claim that Carnapos exist if, in actual fact, Carnapos exist in Platonic heaven. Our practices must answer to the world, rather than the world trivially conforming to our practices. This means that the hardcore ontological realism espoused by some philosophers of mathematics makes standard mathematical practice epistemologically problematic and completely unjustifiable.[24] Some non-traditional realist theories do much better with freedom, but even these theories don’t explain the full range of freedom (paraconsistent mathematics), nor do they connect up to a general and appealing metaontology applicable to mathematics.[25] In both of these respects, alternative approaches fail to explain things that conventionalism does quite easily.

This is why conventionalism, unlike many rivals, will have no problem at all accounting for future developments in mathematics, whatever they may be. If category theory or some new exotic approach supplants set theory as the standard foundation for mathematics, that is no problem at all, as far as conventionalism goes. Of course, if mathematics radically changes in character, like if mathematics becomes akin to physics, that would significantly change matters, and conventionalism would no longer be the right philosophical account. But by “mathematics” we currently mean pure mathematics as practiced now, as it has been practiced since the

ancient Greeks. And conventionalism explains the freedom found in our actual mathematical practices perfectly. It can even—by appealing to the conventionalist account of logic developed in part II—explain why inconsistency is shunned in standard mathematics but not in paraconsistent systems. Other approaches have difficulties with this and related points. On all fronts, conventionalism does better than its philosophical rivals in accounting for and rationally justifying the actual, observed practice of mathematicians, both positively and negatively.

Let us turn now to an aspect of mathematics that isn't obviously so easy to explain on conventionalist grounds: the all-important applicability of mathematics to the physical world. Do conventionalists have anything useful to say here?

V A Conventionalist Look at Applications

The mathematician Eugene Wigner famously claimed that mathematics was "unreasonably" effective at describing the physical world.[26] Whether or not the effectiveness of mathematics is "unreasonable", it is certainly ubiquitous—all of our physical theories are formulated mathematically, and mathematical relationships and principles can also be applied in the "soft" sciences in various ways. To put things metaphorically: mathematics is the language in which the book of nature is written. This fact is widely thought to be puzzling, in itself, and it might also be thought particularly puzzling for conventionalists, so it is worth saying a bit about how conventionalists can handle the applications of mathematics.

Something that is worth saying right up front, since it is often lost sight of in discussions of the applicability of mathematics, is that, fundamentally, the applicability of mathematics is not deeply puzzling. Branches of mathematics like arithmetic grew out of our human practices of counting and making cardinality comparisons. So it is not puzzling or surprising that numbers can be applied to count objects in the world, for this is almost certainly what arithmetical notions *were created to do*. This is only as puzzling as it is that hammers can nail items into place. All of the canonical branches of

mathematics are similar in this respect: set theory grew out of more general and abstract cardinality comparisons and counting processes, geometry grew out of spatial applications, rational and real numbers grew out of measuring processes, and so on. The mathematician Saunders MacLane pointed out that even beyond these canonical examples, virtually all of our mathematical concepts are closely related to particular human activities.[27] As such, the applicability of mathematical concepts to the activities and aspects of reality that they were *designed* to apply to is not mysterious at all.

As it is in mathematics, so it was in logic. There was no puzzle at all about how formal logic managed to apply to natural language arguments. Even more generally, if only the canonical applications of mathematical theories were at issue, the "unreasonable" effectiveness of mathematics would be exactly as mysterious as the "unreasonable" effectiveness of *language* and no more.[28] Our tools were created to be tools, so their effectiveness as tools is neither unreasonable nor unexpected. This point is rarely made, but it should be made, often and loudly, whenever the applicability of mathematics is discussed.

Perhaps the point isn't often made because on hardcore realist views of the nature of mathematics the point itself is puzzling. Since our mathematical concepts have to answer to some independent reality, saying they were cooked up for some particular purpose is not the end of the matter. By contrast, conventionalism rejects external constraints on our concepts. This means that conventionalism is particularly well placed to accommodate the close relationship between our mathematical concepts and the physical world. Conventionalism recognizes, as most other philosophical theories of mathematics do not, that mathematics is a byproduct of human conceptual practices, as embodied in our linguistic rules. These practices have a history, and they were initially designed for particular purposes. The idea that a concept like *number* could have been *created for* application to (particular aspects of) the world is not at all mysterious for conventionalists. And once created, it is automatically in good standing, so no lingering mystery remains. So not only can conventionalism explain the natural applications of

mathematics, it is perhaps the only theory of mathematics which can do so without adding various epicycles.

All of this is fine for the canonical applications of mathematics, but as Wigner famously noted, not all applications of our mathematical theories were intended from the start. In fact, mathematical principles and notions tend to pop up in unexpected places. A famous example involves arithmetic; it turns out that various principles involving prime numbers are extremely useful in cryptography. The popular RSA public key cryptosystem is based on fundamental arithmetical principles.[29] These applications of the theory of numbers were not, in any sense, built in or intended from the start. This is not an isolated example, for in both scientific and everyday contexts, our mathematical theories have been fruitfully applied to aspects of the world for which they were never intended. These unintended applications, in contrast to the intended applications, are apparently puzzling and maybe they even seem "unreasonable", as Wigner claimed.

But their unreasonableness is overstated, as can be seen if we look at the details of what happens when pure mathematical theories are applied to physical systems. Applications proceed via *bridge laws* mapping theorems of the mathematical theory to states of the physical system (sometimes these bridge laws derive from a direct mapping between objects of the theory and objects in the world, but the same general structure is in place). Conventionalism explains why we are free to stipulate these bridge laws for use as rules in the service of any attempted application. Given this particular freedom, and the more general freedom that mathematicians have (discussed in the previous section), scientists have a nearly bottomless mathematical toolbox. And the tools in the box are often related to each other in subtle ways that can be exploited. If mathematical theory T_1 is directly applied to a physical system *S*, a connection between T_1 and another mathematical theory, T_2, often means that T_2 can be used, indirectly, to obtain information about *S*.

Given the "boundless ocean of limitless possibilities", the fact that mathematical theories find applications in the world for which they were never intended isn't as puzzling as it seemed at first glance.

Conventionalism explains how it is that our mathematical theories grow and develop from old theories, since mathematics is a byproduct, for conventionalists, of our conventional rules. Successful applications, via bridge laws, require some type of structural similarity (often only partial) between the theory (or its standard models) and the physical system it is being applied to.[30] Given that so much of our mathematics was initially built to model worldly structures, together with the plausible assumption that nature is fairly uniform, it is not all that surprising that applications that were unintended are waiting to be found. Especially given the diversity and freedom of mathematics, which is fully backed up and explained by conventionalists (even for *inconsistent* mathematics). For conventionalists, this is only a particular example of the familiar phenomenon of exaptation or general repurposing of existing tools. No more, no less.

But there is a final issue that makes the applicability of mathematics seem nearly miraculous. Mathematical research is often driven by *aesthetic* factors. Mathematicians seem to rely on a sense of beauty in their search for mathematical truth. But the human aesthetic sense is, like all of our capacities, the output of our evolutionary history and social background. It seems mysterious that our aesthetically selected mathematics applies to the objective physical world. Why should our idiosyncratic sense of beauty turn out to be a reliable guide to the objective world?[31] Unless the Universe was build for us, by God or some other designers, this seems hard to explain.

This sounds convincing, but it is ultimately mistaken. The aesthetic sense in mathematics tracks features that aren't idiosyncratic—generality, fruitfulness, simplicity, and other objective features of theories that are not defined in terms of human psychology.[32] The human aesthetic sense when applied to mathematics isn't independent of *truth* in mathematical theories. It likely that our sense of theoretical aesthetics developed, in part, to track truth in the world, rather than truth in the world being keyed to our theoretical aesthetic sense. If there is such an alignment, this order of explanation is the most plausible. So even given an alignment between mathematical

aesthetics and mathematical epistemology, it is not obviously mysterious, let alone “unreasonable”.

All told, mathematics isn’t *unreasonably effective*, and it isn’t effective *as if by magic*. You don’t actually get more out of a bit of applied mathematics than you put in (so to speak). The triviality of mathematical truth isn’t a stumbling block here; instead it once again dovetails with the effectiveness of mathematics. You can infer that Jared is a male from the fact that Jared is a bachelor, even though it is a conventional truth that all bachelors are male, and a similar story goes, often at a much, much greater level of complexity, for all mathematical truths. Given the vast range of mathematical structures dreamt of in our philosophy, many of which were initially designed to apply to physical systems, and the various ways of mapping these structures to any given physical system, it is not very surprising that worldly structures can be modeled in unexpected ways by human mathematics. Conventionalism neatly explains the applicability of mathematics, far better than alternative views in the philosophy of mathematics, I think.

VI A Conventionalist Epistemology for Mathematics

Chapter 6 developed an externalist, proof-based approach to the epistemology of logical truth and validity. At bottom, the approach holds that a speaker can come to know ϕ via proof if they have an epistemic right to the rules used in the proof, in a given context. Logical inferentialism was central to this strategy in vindicating the meaning entitlement connection (MEC), so it should come as no surprise that I think we can apply the very same strategy to mathematics, with mathematical inferentialism playing the role that logical inferentialism played before.

In mathematics, the proof-based strategy is not only philosophically plausible, it is tremendously intuitive. It perfectly dovetails with the actual practice of working mathematicians. Mathematics proceeds via proofs, forever building upon our store of mathematical knowledge by proving more and more theorems. The

central philosophical challenge that the proof-based epistemology had to overcome in the case of logic was that the proofs in question were often rule-circular. But for the bulk of mathematics, this isn't a concern at all, since the proofs will instead be analogous to the non-rule-circular proofs that were used to found our knowledge of logical truths (as opposed to validity claims).

As a very simple illustration of this approach to mathematical knowledge, take the following proof that two is a number using the Peano rules. In the full language modeling our arithmetical practice, the sentence "NSS0" can be proved using (*P*1) followed by two applications of (*P*2) and a tacit (structural) appeal to transitivity:

$$\dfrac{\dfrac{\mathrm{N0}}{\mathrm{NS0}}}{\mathrm{NSS0}}$$

Since all of the the rules used in this proof are meaning constituting for the arithmetical vocabulary, by the MEC we have an automatic entitlement to them. If someone accepts that two is a number on the basis of this proof, they thereby come to *know* that two is a number. A similar approach works for justification as well.

This will work for the epistemology of provable sentences in every branch of mathematics. And as mathematics advances, previously proved claims and principles can then be used in knowledge-conferring proofs as a type of shorthand. Things will always ultimately trace back to proofs that use the basic, meaning-determining rules of our mathematical language. This is exactly what we actually find in mathematical practice itself. Conventionalism's proof-based epistemology for mathematics is fully visible in the practice and self-conception of working mathematicians. As a description of the human activity of mathematics, this approach is ideal.

This is all well and good for *provable* mathematical claims, but what about claims that are true but unprovable using the resources of our current language? There will always be sentences like this, given that arithmetical truth (say) is not computable, but proof in any given timeslice of our language is. Take section I's example of the

Gödel sentence, G_A, of the current timeslice of our arithmetical language. Though G_A is not finitely provable from A, it is finitely provable in some well-behaved extensions of A (though of course these extensions will have their own Gödel sentences). There are various ways to extend our current arithmetic so as to prove the Gödel sentence of our starting arithmetical theory, say by adding a truth predicate and compositional rules, adding the consistency sentence for the current commitment set, or adding the Gödel sentence itself. Each of these strategies can be iterated without limit.[33]

What this shows is that the open-endedness of arithmetic would be needed *epistemologically* even if it wasn't needed *metaphysically* (of course, I argued in 10.IX that it is still needed metaphysically as well). We can get a determinate notion of arithmetical truth by closing Robinson arithmetic, Q, under the omega rule, but Q itself has no induction axiom. If the corresponding Q^{ω} rules were our meaning-determining arithmetical rules, then various arithmetical sentences wouldn't be provable without serious additions to or enhancements of our language, even though they would be true in our starting language, since they would follow from Q using the infinitary omega rule.

But our practice does include induction. The induction rule (*P*5) is a meaning-determining rule in my full language reconstruction of natural language arithmetic. Via the open-endedness of induction, the extensibility of our arithmetical means of proof is already built in to our current arithmetical competence. Even the fully open-ended omega rule doesn't do this for us, since beings like us cannot directly construct infinite proofs, though we can follow the omega rule in a less-demanding sense (10.VII and 10.VIII). Induction or some other finite schematic system of extension is needed for an adequate account of the epistemology of arithmetic. By adding further and further arithmetical predicates to our language, via open-endedness, we extend our finite means of proof further and further, and put ourselves in the position to know more and more arithmetical truths. We can even come to know truths that were already true but

unprovable in our starting language by—how else?—constructing finite proofs of them.[34]

With open-endedness on board we have a satisfying proof-based epistemology for mathematics that harmonizes with the practice of actual mathematicians. Since the proof-based epistemology applies so readily to mathematics there is no need to belabor things here. The basic approach of chapter 6 easily extends to mathematics. But there is an issue in the epistemology of mathematics (and logic) that cries out for special discussion: our supposed knowledge that our mathematical theories *are consistent.*

VII A Conventionalist Look at Knowledge of Consistency

A theory T is consistent just in case there is no sentence ϕ in the language of T such that both $T \vdash \phi$ and $T \vdash \neg\phi$. In classical logic, a theory is consistent just in case some sentence in the language of the theory is *not* provable from the theory. Consistency is a logical notion, but I didn't treat the epistemology of consistency directly in chapter 6 because mathematical results, most famously Gödel's second incompleteness theorem, bear directly on the epistemology of consistency.

The consistency of standard theories can be expressed, via the technique of arithmetization (12.I), by a sentence in the language of arithmetic—"Con(T)" for a theory T. Gödel's second incompleteness theorem says that no consistent, recursive theory of arithmetic can prove its own consistency sentence. A corollary of this is that if theory T can relatively interpret theory S, S cannot prove T's consistency sentence, since if it could, it could also prove its own consistency sentence, thus violating the second theorem.[35] The second incompleteness theorem seems to show that Hilbert's program of proving infinitary mathematics consistent using finitary mathematics is impossible to carry out.[36] But there are a number of fine points about this that need to be carefully noted, lest the epistemological importance of Gödel's theorem be overstated.

The first is that the second incompleteness theorem only holds for "standard" consistency sentences, that is: consistency sentences of a particular kind.[37] There are other sentences in the language of arithmetic that, through the arithmetization of syntax, both intuitively express the consistency of theory T and are also provable in T.[38] In addition, there are weak arithmetical (or arithmetic-like) theories T in which even T's standard consistency sentence is provable; these are called "self-verifying" theories.[39] Finally, there are consistent theories of arithmetic T that prove their own *inconsistency* sentences, "¬Con(T)", for example: Q+¬Con(Q).[40] For all of these reasons, we should be wary of blithe attempts to analyze informal consistency claims about a theory with consistency sentences for the theory that are formulated in some recursive theory of arithmetic or syntax.

Consider Peano arithmetic (PA)—can we come to know that PA is consistent by proving "Con(PA)"? In stronger mathematical theories, we can prove the sentence "Con(PA)", so the epistemology of the previous section suggests that we can come to *know* it, by way of such a proof. I have been told by some that this is beyond belief, but I think the apparent implausibility here is based on a simple misunderstanding. All that such a proof would tell us is that the sentence "Con(PA)" is mathematically true. But this is not at all the same thing as showing that the formal system PA is consistent. The crucial point here is that arithmetization is a bit of *applied* mathematics. Thus a proof (in ZF, for example) that PA is consistent cannot be the entire story here, since it concerns only a sentence of *pure* mathematics.

As discussed in section V, applying mathematics requires bridge laws between the world and the mathematical theory in question. For the consistency of a theory like PA, the coding has been set up so that there is an exact correspondence between the truth of a consistency sentence in the standard model of arithmetic and the consistency of the theory PA, so that: $\mathbb{N} \models$Con(PA) if and only if PA is consistent. So a full argument for the consistency of PA that made this correlation explicit, rather than blithely equating the consistency of PA with the provability of "Con(PA)," would go something like this:

1. $\mathbb{N} \models$Con(PA) if and only if PA is consistent
2. Con(PA)
3. *So*: PA is consistent

But as stated, this argument is a non-sequitur. Proving "Con(PA)" in ZF or some other background theory is different than proving $\mathbb{N} \models$ Con(PA). But of course, in ZF we can do enough model theory to prove "$\mathbb{N} \models$Con(PA)". In other words, ZF $\vdash$ "$\mathbb{N} \models$Con(PA)", so we can (linguistically descending) prove the left-hand side of premise 1, which might seem to be enough for a variant of the argument.

The reasoning involved in a ZF proof of consistency is complex and works in a very rich background theory that interprets PA. As such, it would be unlikely to assuage the worries of someone who truly wondered if PA was consistent.[41] It is worth keeping in mind that if our background theory is inconsistent, our proof of premise 2 will still go through, but we will no longer be able to move from this to a claim about consistency. This is because in an inconsistent theory, there will no longer be any link between the provable claims and facts about the standard model of arithmetic. We will know this not by way of pure mathematics, but by considering the bridge laws that link consistency sentences with consistency facts. It is for this reason that the extreme permissivism of unrestricted inferentialism doesn't lead to nonsense here. Not all consistency sentences express consistency (see 12.II for much more on this).

It is hard to see how someone who was truly worried about PA's consistency could trust ZF or any other theory that interprets PA. The most famous proof of the consistency of PA is Gentzen's proof, which can be carried out in primitive recursive arithmetic plus transfinite induction up to ϵ_0 on logically simple sentences—PRA + TI_{ϵ_0}. This theory isn't an extension of PA, so a consistency proof carried out inside of it might be dialectically convincing, at least if PRA + TI_{ϵ_0} can itself be justified on independent grounds that don't already assume the consistency of PA.[42]

Of course, if we have a determinate grasp of arithmetic, which I have argued we do (in chapter 10), then we will also have a

determinate grasp of *syntax*, at least for theories like PA. That is what seems needed for us to really and truly say and prove that PA is consistent. Obviously, if this means working in true arithmetic (TA), then since TA extends PA, it would seem uninteresting to prove Con(PA) in TA. But this probably isn't how we should think of things here. The dialectical and philosophical subtleties here will be explored in detail in 12.II and 12.III.

If we are really and truly worried about the consistency of PA, we also shouldn't ignore inductive and abductive evidence for PA's consistency. The standard tricks for generating an inconsistency don't work in PA, one of the best studied formal theories ever devised. The best explanation of the failure to generate an inconsistency in PA may well be that PA is, in fact, consistent. We also seem to have an intuitive grasp of a model of PA, having a sense of what the model "looks like", and a first-order theory is consistent just in case it has a model.[43] This type of evidence is open to everyone, conventionalists included. It is sometimes the strongest and least tendentious evidence we have of the consistency of our mathematical theories. But it isn't sacrosanct or uncontroversial, on its own.

Most of what I've said about PA's consistency also goes for other mathematical theories as well, though the facts on the ground will vary a bit from case to case. For a much more powerful theory like ZF, the risk of inconsistency is greater, our grasp of a model is dimmer, and thus our confidence and perhaps our epistemic standing with respect to consistency is weaker than with PA. All of this is as it should be. None of the points made here undermine the general, proof-based approach to mathematical knowledge from the previous section. They only illustrate that when it comes to applied mathematics in general, and this weird case of mathematics being applied to itself in particular, things are much trickier and more involved than they are for canonical cases of pure mathematical knowledge, where everything proceeds quite smoothly.

≈≈≈≈≈≈≈≈

I THINK IT'S clear that the conventionalist theory of mathematics has no special problem accounting for mathematical truth, mathematical knowledge, mathematical paradoxes, mathematical freedom, or mathematical applications. But several technical objections to mathematical conventionalism have been raised in the philosophical literature. These even include among them the claim that arithmetization itself is incompatible with conventionalism. This and other technical objections to mathematical conventionalism need to be considered and answered before victory can be declared.

---

1 The distinction between being provable using some rules and being "made or forced true by" or "following from" them calls to mind Carnap's (1934) distinction between derivability and consequence. See 13.II for a discussion of Carnap's views.

2 This is done explicitly in Tennant (1978).

3 Logicism was famously endorsed in Frege (1884), (1893), and (1903) and Russell (1903).

4 See Carnap (1931); the quasi-logicist views of Wittgenstein (1922) were a major influence on Carnap's pre–*Logical Syntax* views on mathematics.

5 In fact, Benacerraf (1981) basically argues that there have been no true logicists since Frege.

6 Quoted from Carnap (1937), page 318. See Quine (1951) for the most influential discussion of these matters.

7 Quoted from page 921 of Carnap (1963*b*), in response to Quine (1960*a*).

8 Chapter 2 of Quine (1960*b*). See also the discussion in the final section of my (2018).

9 In Quine (1970*b*), see my (2018) and Burgess (2004) for some relevant discussion.

10 See Putnam (1975) and Burge (1979) for the introduction of this type of semantic externalism.

11 See Frege (1893), (1903), and Russell (1903), (1908).

12 See Priest (1995).

13 See Warren (2015*b*).

14 This approach is developed in most detail in my (2017*b*).

15 Russell (1906).

16 Dummett (1991*b*), page 316–317.

17 This is one half of the puzzle of indefinite extensibility developed in my (2017*b*).

18 In my (2017*b*).

19 Fine (2006) makes a similar distinction.

20 Carnap (1934) was the first attempt to use something like quantifier variance to explain mathematical freedom; Berry (2015) uses a restricted form of quantifier variance for the same purpose.

21 The formalism of Frege (1893) and (1903) was shown inconsistent by Russell (1903), that of Church (1932) by Kleene and Rosser (1935), that of Quine (1940) by Rosser (1942), and the large cardinal axiom of Reinhardt (1967) by Kunen (1971).

22 Named for the title of Quine (1937), in which the system was introduced.

23 Boolos (1971) seems to suggest that extensionality, but perhaps not the other ZF axioms, is required of any possible concept of set worthy of the name.

24 This alone puts some pressure on Platonism—see the discussion of the relationship between mathematics and the philosophy of mathematics in Lewis (1991).

25 The main alternative views I have in mind are those of Balaguer (1998) and that suggested in Putnam (1980); see Field (2005) for some discussion of how these views answer the epistemological challenge of Benacerraf (1973), a point which is closely related to how they deal with freedom.

26 In Wigner (1960).

27 In MacLane (1986).

28 A comparison like this is made in Sarukkai (2005).

29 Rivest, Shamir, and Adleman (1978).

30 If nonstandard models were used in this context, it might trivialize the problem, so that only cardinality would matter—see Newman (1928), Skolem (1922), and Putnam (1980). The approach of chapter 10 can help to ward off these concerns.

31 Steiner (1998) develops related concerns at length.

32 This type of point is made, and illustrated with several detailed case studies, in Waxman (2019*a*). Related points are also made in Grattan-Guinness (2008).

33 See Franzén (2004) for some relevant discussion.

34 Of course, the claim here is not that there is any given extension of our arithmetic where *every* truth is provable, since that would be impossible, assuming Church's thesis.

35 The second incompleteness theorem was announced in Gödel (1931).

36 See Hilbert (1926) for the original Hilbert Program, and Detlefsen (1986) for a more positive assessment of the program's fate.

37 Hilbert and Bernays (1939) give three conditions on the provability predicate used in formulating the consistency sentence that, if satisfied, suffice for Gödel's second theorem to hold for the associated consistency sentence.

38 Gödel's lectures show that he was aware of the possibility of provable consistency sentences; the issue is explored in Feferman (1960).

39 See Willard (1993) and (2001) .

40 For a discussion of this, see my (2015*a*) and chapter 12.

41 Edward Nelson famously thought that PA was inconsistent, and at one point even announced a proof that he later withdrew after errors were found. Nelson's own, much weaker, theory of arithmetic is developed in his (1986).

42 The proof is in Gentzen (1936). The epistemic worth of Gentzen's proof is discussed in Chow (2019); the philosophical issues are addressed in detail in Waxman (2019*b*).

43 This type of approach to knowledge of consistency is pursued in Waxman (2019*c*), drawing on the notion of composite dispositions introduced in my (forthcoming *a*). I reply to Waxman an unpublished paper.

12

Metamathematics versus Conventionalism

In chapter 7 I responded to the most influential general arguments against truth by convention. The discussion in chapter 7 also shields *mathematical* conventionalism from those very same objections. But one of those general objections takes a particularly pointed form in the light of modern metamathematical results, and so must be dealt with anew (7.IV). In addition, a number of technical arguments against mathematical conventionalism have been raised in the literature. Some of these seem to be very widely endorsed.

Important arguments of this kind offered by Kurt Gödel, Peter Koellner, John Pollock, and Hilary Putnam will be answered in detail here. All of these objections are at least somewhat technical, often relying upon sophisticated logical and metamathematical results and techniques. Accordingly, this chapter is a bit more technical than the other chapters in the book. To make things more accessible I begin with a very brief overview of the arithmetization of syntax.

I The Arithmetization of Syntax

The early twentieth century produced several justly celebrated results in mathematical logic, the most famous of which are Gödel's two incompleteness theorems, which have already been discussed at several points in the book. Gödel proved the incompleteness theorems using *arithmetization*—the coding of syntactic properties using arithmetical predicates.[1]

The basic idea behind arithmetization is to code the syntactic items in formal systems using natural numbers. This can be done in many ways, but generally we want to map each symbol in the language of the theory we're considering to a distinct natural

number. We call the number associated with an expression ϕ the "Gödel number of ϕ". Here I'll use brackets to pick out the Gödel number of an expression, so that "[ϕ]" picks out the Gödel number of ϕ. The essential feature is that there are simple algorithms that allow us to move either from an expression in the language of the theory to its Gödel number, or from a Gödel number to an expression of the theory. Our coding scheme can then be extended to allow us to code sequences of sentences, thus allowing for proofs to be coded as well.

Once we have a suitable scheme of Gödel coding, we can define primitive recursive *arithmetical* predicates and relations that *express* syntactic properties. So, for example, there will be a numerical predicate "$\mathrm{Wff}_T(x)$" that holds of a natural number n when n codes a well-formed formula of the language of theory T; and there will be a numerical relation "$\mathrm{Prf}_T(x,y)$" that holds of the natural numbers n, m when the first number, n, codes a T-proof of a formula ϕ with Gödel number m. Say that an arithmetical theory T *captures* a k-place relation R by the formula $\varphi(x_1, \ldots, x_k)$ if and only if for all $n_1, \ldots, n_k$, (*i*) if $n_1, \ldots, n_k$ stand in relation R then $T \vdash \varphi(n_1, \ldots, n_k)$ and (*ii*) if $n_1, \ldots, n_k$ do not stand in relation R, then $T \vdash \neg\varphi(n_1, \ldots, n_k)$. An arithmetical theory that can capture all primitive recursive relations is able to "say" a tremendous amount about the syntax of various theories, including itself.

Interestingly, this can be done even in very weak arithmetical theories, such as Robinson arithmetic (Q), which has already been encountered at several points above. Of course, this is not to say that every syntactic property is captured by Q. Some intuitively some syntactic properties turn out not to be primitive recursive. The first incompleteness theorem entails that any consistent, recursive theory that extends Q is negation incomplete. Gödel even shows how to construct a sentence G_T, the so-called "Gödel sentence," for theory T meeting these conditions such that $T \nvdash G_T$ and, if T is in addition ω-consistent, $T \nvdash \neg G_T$.[2] In addition, with our arithmetized provability relation we can express the consistency of an arithmetical theory T using a sentence of T as follows:

$$\mathrm{Con}(\mathrm{T}) =_{def} \neg\exists x \mathrm{Prf}_{\mathrm{T}}(x, [0{=}1])$$

This sentence "says" that there is no T-proof of the arithmetical sentence "0 = 1", and since T is assumed to extend Q, T ⊢ 0≠1, so this suffices to show that T is consistent.[3] In the classical logic in which standard mathematical theories are formulated, inconsistent theories are trivial, they prove everything. So a statement that something is not T-provable is also a consistency statement, in this context. The second incompleteness theorem tells us that if T meets the above conditions, T ⊢Con(T) just in case T is inconsistent (provided that "$\mathrm{Prf}_{\mathrm{T}}$" meets certain conditions).[4]

Both "G_{T}" and "$\mathrm{Con}_{\mathrm{T}}$" are Π_1-sentences in the language of arithmetical theory T that cannot be proved in T. This means that incompleteness arises very quickly, in our simplest attempts to universally generalize over numbers. With this background in place, let's turn to attempts to use arithmetization and results proved using arithmetization to argue against mathematical conventionalism.

II Pluralism, Consistency, and Consistency Sentences

As seen throughout chapter 10 and in 11.I, my theory of mathematics doesn't simply identify mathematical truth with provability from the meaning-constituting rules. But I do take provability from the basic rules as a sufficient condition for mathematical truth. I am also a pluralist about mathematics (chapter 5 and 11.IV); I think that many alternative mathematical—even arithmetical—rules can be used to found alternative practices.

The role of "proof" in conventionalism seems to require that syntax is a matter of fact, not convention. This harkens back to the variations on Quinean themes discussed in 7.IV, but in the light of arithmetization, the challenge takes a particularly pointed form. The previous section discussed the familiar fact that syntactic claims can be encoded as arithmetical claims via arithmetization. There is a perfectly good sense in which the *arithmetical* sentence, "Con(Q)"

expresses the *syntactic* claim that Q is consistent. But if Q ⊢Con(Q), then Q is inconsistent, by Gödel's second incompleteness theorem.[5] Conventionalists, given their pluralism, must allow alternative practices in which "¬Con(Q)" is finitely provable using the basic rules of the practice, and is therefore *true* (by the **MVC**). How can this be, given that the facts about syntax are not under our control? Something, it would seem, has to give.

I agree that insofar as they are determinate, linguistic rules are matters of fact. Given that a community follows the Q-rules, it will be a matter of fact, not convention, whether or not the sentence "0 = 1" is provable using the rules. A determinate package of rules is either consistent or not, with little sense to be made of a borderline case. As a conventionalist, I accept all of these points, as I think everyone should. But several prominent philosophers have argued that these seemingly benign observations sit ill with arithmetical conventionalism. Here is Hilary Putnam:

> …once we grant that there is at least *one* mathematical fact which is not simply our stipulation … that, for example, the *consistency* of our stipulations/practices is not itself just another stipulation or practice—then logical positivist/Wittgensteinian accounts of logical and mathematical truth are seen to be bankrupt.[6]

Peter Koellner made a similar point more recently, in a discussion of Reichenbach's pluralist, conventionalist views:

> This radical form of pluralism came to be challenged by Gödel's discovery of the incompleteness theorems. To begin with, through the arithmetization of syntax, the metamathematical notions that Reichenbach takes to fall within the provenance of "critical investigation" were themselves seen to be a part of arithmetic. Thus, one cannot, on pain of inconsistency, say that there is a question of truth and falsehood with regard to the former but not the latter.[7]

These arguments are related to Copleston's charge that consistency itself cannot be a matter of convention, discussed in 7.IV.[8] But they are even more worrying for the conventionalist, since they point out an apparent tension at the very heart of the theory, as applied to arithmetic.

Of course, in the exact form in which they are stated, they don't apply to my conventionalist view (nor to the views of the more sophisticated logical positivists, like Carnap). Putnam seems to think conventionalism simply identifies mathematical facts with our stipulations. But not only does this implausibly take conventions to be stipulations (2.II), it also understands conventionalism in one of its implausible forms (already rejected in 1.II). Koellner likewise seems to focus on the implausible, non-cognitivist form of conventionalism, rather than the more attractive explanatory version that I have been developing and defending here.

But it would be intellectually dishonest to simply note this and move on. Both Putnam and Koellner are pointing to a serious tension that needs to be reconciled by every version of conventionalism, including my own. The tension is created by the combination of the factuality of syntax, the possibility of arithmetization, and pluralism about arithmetic. Perhaps it won't be surprising that I think the tension is eased by our old friend, the translation mistake, this time applied to arithmetical languages. This is best illustrated with an example.

Imagine that we encounter a Plutonian community whose arithmetic includes all of the axioms of Q in addition to the sentence "¬Con(Q)", so that, in effect, Plutonian arithmetic is the theory Q + ¬Con(Q). We know, in our arithmetic, that the theory Q + ¬Con(Q) is consistent if and only if Q is consistent. Also, from the way that consistency sentences are constructed, Q + ¬Con(Q) $\vdash$¬Con(Q+¬Con(Q)), so assuming that Q is consistent (a safe bet!), Plutonian arithmetic is based on a consistent theory that proves its own *inconsistency* sentence! Hardcore realists will say that the Plutonians have a false theory, since their "¬Con(Q)" axiom is false. Conventionalists cannot say this, since by **MAC** and the **MVC**, the Plutonians' basic rules are automatically valid in the Plutonian language.

What conventionalists will say instead is that translating the Plutonian sentence "¬Con(Q)" into our language, homophonically, commits the translation mistake. Since (I have argued) our arithmetic is determinate, the sentence "¬Con(Q)" is false in our language if Q is consistent. But the same sentence, syntactically individuated, is true in Plutonian. So they don't mean the same thing that we mean by the sentence. This is possible because the Plutonians have an *ω*-inconsistent theory of arithmetic. Every model of their theory is—from our point of view—nonstandard. Arithmetization is a bit of applied arithmetic, and in *ω*-inconsistent theories this application goes haywire. In models of Plutonian arithmetic, predicates like "Wff_Q" and "Prf_Q" will be satisfied by nonstandard elements that don't really code for actual well-formed formulas or proofs in the Plutonian language.

Arithmetization is a piece of applied mathematics and should be treated as such. Given the independent facts about syntax, an arithmetical language like ours can model them, but the Plutonian language cannot. Nobody, least of all conventionalists, thinks that *every* mathematical theory can be applied to exactly the same worldly phenomonena. As discussed in 11.V, applications of mathematics are based on features of the world being (at least partially) mirrored by aspects of our theories. Since Plutonian arithmetic has different structural features than our arithmetic, it is not surprising that some applications of our arithmetic cannot be aped by the Plutonians. There is nothing more mysterious than this going on here. But as usual, the translation mistake obscured this from view. Once the mistake is avoided, the matter can be seen clearly.[9]

So despite how it initially seemed, there is no sense in which the conventionalist is forced to say that some arithmetical claims are non-conventional. Nor is the conventionalist forced to place restrictions on which sentences can be made true by convention. Arithmetization is a piece of applied mathematics, and like all pieces of applied mathematics, it only works as intended in certain theories. The Plutonians can mimic some applications of our arithmetic, but not all. Their acceptance of an existential claim ("¬Con(Q)") as an

axiom, without any numerical witness, leads them to mean something slightly different than we do with arithmetical terminology.

The particular argument here was fairly easy to answer, but the worries behind the argument are related to very general worries about the coherence of mathematical conventionalism. These general worries must also be addressed before we can turn to other specific arguments against conventionalism.[10]

III Circularity and Syntax

Imagine that the Plutonians of the previous section try to distinguish between the theories Q and Q + ¬Con(Q), for themselves. Given that their syntactic metatheory includes Q + ¬Con(Q), the Plutonians will think that Q is inconsistent. Because of this, they won't be able to distinguish between the theories Q and Q + ¬Con(Q) at all.[11] Yet surely these are *different* theories, and there could be linguistic communities that followed rules corresponding to one theory but not the other. Has something gone wrong here? Does this show that syntax is up for grabs, a matter of whatever our arithmetical or syntactic metatheory tells us?

No, not at all. Syntax in the relevant sense is still a matter of fact. The facts about the rules followed by speakers of a language are a matter of the speakers' dispositions to use language in certain ways. The Plutonians are willing to accept the sentence "¬Con(Q)" on the basis of nothing at all, but we are not. This is a real difference. Once again, arithmetization is a piece of applied mathematics. In ω-inconsistent theories, facts about the recursive syntax of real languages will be missed. Languages that are distinct will blur together. The theory of the Plutonians is worse than ours, as a general syntactic metalanguage. This doesn't mean that the Plutonian arithmetical theory is false in Plutonese. It just means that their theory can't be applied in all of the ways that our (quite different) arithmetical theory can be applied.

Syntax is a matter of fact, rooted in actual and possible behavior with sentences. Conventionalists don't need to be pluralists about these facts. They can be pluralists about *theories* of syntax. But

since these are, by their very nature, applied theories, they are more like theories of physics or applied mathematics, rather than pure arithmetic. Just as a theory of applied geometry might be unable to correctly describe the behavior of objects in physical space, the Plutonian theory of syntax is unable to correctly describe the syntax of possible natural languages. This means that *egalitarian* pluralism about theories of syntax is implausible, even for conventionalists.

The skeptic might further worry about the actual facts of syntax. The syntactic conventions are linguistic rules. The unbounded application of these rules was accounted for using the idea of composite dispositions (2.VII). But how, the skeptic wonders, can we know whether these finitely iteratable "composite" dispositions aren't actually only "finitely" iteratable in the sense of the Plutonians. Is some question being begged by talk of "finite" here? What rules out nonstandard interpretations of my discussion of "composite" dispositions or "finite" iterations? Exactly analogous questions can be raised at several other points. To consider but one: my conventionalist account of arithmetic appealed to the infinitary omega rule (10.VII, 10.VIII, and 11.I), but we cannot finitely state the infinitely many premises of an omega inference, we can only indicate or gesture at them. So the finitary indications in our language might seem themselves to be posed in a standard background metatheory. What is it that rules out nonstandard interpretations of this metatheory? Just as with "composite", our claims about "omega" will have nonstandard but admissible interpretations. Or so worries the skeptic.

It makes no difference to the skeptic's challenge if this background theory is a version of arithmetic where syntax is arithmeticized or a direct theory of syntax itself. A theory of syntax, in this sense, is equi-interpretable with arithmetic.[12] We obviously can't be assuming that our syntactic metatheory is determinate in order to establish the determinacy of arithmetic if we actually want to make progress. The skeptic's challenge is that many of the claims I have made in this book—about syntax, about rules, about dispositions, about open-endedness, about the omega rule, and so on—will go awry unless the language of the book, in which all of my philosophical claims are being made, is somehow magically free of nonstandard

interpretations. The skeptic's charge is that nonstandard interpretations of these notions have only been ruled out by a tacit appeal to a metalanguage without nonstandard interpretations!

The skeptical dialectic here is fairly standard in discussions of general indeterminacy arguments. An analogous move is made by Kripke to defend his rule-following paradox and by Putnam to defend his model-theoretic argument against realism.[13] The skeptic gives an argument against determinacy, in response some theoretical account T of determinacy is offered, the skeptic replies by applying the original argument to T itself. In this way T is treated as "just more theory" of the very same kind that was originally under attack. This is a standard reply on behalf of the skeptic, but it is often extremely unfair. The move is only fair when T really is merely more of what was originally attacked. When T is not "just more theory", this skeptic's reply is aimed at a straw man. In a discussion of Putnam's argument David Lewis puts the crucial point quite neatly:

> Suppose that we say it is constraint *C* that saves the day. …*C* is *not* to be imposed just by accepting *C*-theory. That is a misunderstanding of what *C* is. The constraint is *not* that an intended interpretation must somehow make our account of *C* come true. The constraint is that an intended interpretation must conform to *C* itself.[14]

This is closely related to a point I stressed repeatedly in 10.II when discussing open-endedness—we cannot simply *define* open-endedness in some standard background theory. Analogous points answer all of the skeptic's other charges as well.

When I say that we really and truly follow the omega rule, and that this is a key part of what explains the determinacy of arithmetic, I mean that we *actually follow the omega rule*. Not that we accept some standard metatheory in which the "omega rule" is stated finitely. I instead mean that we can actually perform infinitary inferences of the kind required by the rule. I do not simply mean that what is explanatory are the claims of some standard metatheory. I mean that we infer in such-and-such a way. If this is right, then the original challenge cannot be reposed. Given that our dispositions are

actually a certain way, we have a determinate grasp of the natural numbers. Our usage allows us a determinate conception of the naturals, that is how. Given that this is the case, we can generalize and talk about the naturals in a perfectly standard fashion in our philosophical theorizing. There is no problematic circularity involved.

It is decidedly *not* the case that my claims about open-endedness, the omega rule, and composite dispositions are themselves to be understood in terms of some standard, finite, language-closed metatheory. If that were intended, the "just more theory" objection would be spot on. Nonstandard interpretations of our terms "composite dispositions", "open-endedness", and "omega" are not ruled out by appeal to a standard metatheory, with its own problem of nonstandard interpretations to deal with. Instead the natural language in which my discussion is presented is a language whose meaning-determining rules are themselves open-ended and non-finitary. The snake eats its own tail.

If the facts about syntax and arithmetic are *as we all agree that they are*, then we, but not the Plutonians, have an adequate account, at least for the syntax of theories like Q and PA. If we instead imagined trying to apply mathematics to the "syntax" of the language of God, which might have more primitive symbols than there are sets, things are much less clear. We can't even make a real start on arithmetizing the syntax of God's language. Our arithmetical commitments at any given point are recursive, and we have an adequate and determinate conception of syntax just in case we have a determinate conception of arithmetic. And we do, so we do (chapter 10).

So despite the arguments of Putnam, Koellner, and others, I think that conventionalism evades the worries here, both sharp and particular, and soft and general. But there are other anti-conventionalist arguments in the literature that use arithmetization-based metamathematical results in attempting to refute conventionalism.

IV Pollock, Provability, and Paradox

During the waning days of conventionalism, John Pollock published a sophisticated technical argument aimed at refuting a version of conventionalism.[15] His target wasn't exactly conventionalism about mathematical *truth*, but something more like conventionalism about mathematical *proof*. In effect, Pollock argues that since conventionalists think that the notion of classical mathematical proof is a byproduct of a finite number of syntactic principles and finitary inference rules, they are committed to the existence of a single axiomatic theory M that suffices for classical mathematics and is able to "talk about" its own provability predicate in a way that leads to inconsistency.

Pollock's argument is related to Montague's paradox (the paradox was previously discovered by John Myhill).[16] In general: let T be a theory extending Q, and let "A" be a one-place predicate in the language of T that obeys the following principles:

(T) $\quad A([\phi]) \rightarrow \phi$

(NEC) $\quad$ If $T \vdash \phi$ then $T \vdash A([\phi])$

Under these conditions, T is inconsistent:

1. $T \vdash p \leftrightarrow \neg A([p])$ (for some sentence "p" by the diagonal lemma)
2. $T \vdash A([p]) \rightarrow p$ (from (T))
3. $T \vdash \neg A([p])$ (from 1 and 2)
4. $T \vdash p$ (from 1 and 3)
5. $T \vdash A(\ulcorner p \urcorner)$ (4 and (NEC))
6. $T \vdash \bot$ (from 3 and 5)

To apply this to conventionalism, Pollock argues that conventionalists are committed to a single axiomatic theory M that extends Q and satisfies (T) and (NEC) for a one-place arithmetical predicate "P_M" in the language of M corresponding to M-provability.

Are conventionalists really committed to (T)? Pollock argues that they are, since for any ϕ in the language of M, $P_M([\phi])$ is true just in case $M \vdash \phi$ and if $M \vdash \phi$ then ϕ is true. Combining these we have that if $P_M([\phi])$ is true then ϕ is true. Pollock claims that this shows

that $P_M([\phi]) \rightarrow \phi$ is true. He claims that since this is a sentence in the language of M, the conventionalist thesis also requires that M $\vdash P_M([\phi]) \rightarrow \phi$ because of the equation of mathematical truth with M-provability.[17] Pollock doesn't mention it, but if M is primitively recursively axiomatized and the predicate "P_M" obeys the Hilbert-Bernays derivability conditions, then by Löb's theorem, this is already enough to establish that M $\vdash \phi$, for all ϕ.[18] Also, if M meets the condition of the incompleteness theorems, then plugging a contradiction in to (T), we get M $\vdash P_M([\bot]) \rightarrow \bot$. So by classical reasoning ($\phi \rightarrow \bot \vdash \neg\phi$) we get M $\vdash \neg P_M(\ulcorner \bot \urcorner)$. But this is to say that M $\vdash$Con(M) so M is again inconsistent, this time by way of Gödel's second incompleteness theorem.[19]

One conventionalist of the time, Carlo Giannoni, responded to Pollock.[20] Giannoni pointed out that if (T) is replaced with:

> For every sentence ϕ in the language of M, if M $\vdash P_M([\phi])$ then M $\vdash \phi$

then conventionalists can avoid the contradiction. Though he did think that by using this premise we could establish that M was negation incomplete, and so—via an equation of truth with M-provability—conventionalists had to reject the law of excluded middle. He wasn't bothered by this himself, since he thought conventionalists would already reject excluded middle on "anti-Platonist" grounds. But this is misleading. In a classical system, every instance of the law of excluded middle is an M-theorem. I think that what Giannoni actually meant is that conventionalists should reject the determinacy of mathematical truth. Similar reasoning seems like it would push Giannoni into rejecting determinacy even for basic arithmetic.

Returning to his response to Pollock, Giannoni wasn't merely rejecting (T) ad hoc. Instead he thinks Pollock's argument that conventionalists are committed to (T) is fallacious. Giannoni pointed to propositional systems studied by Curry to demonstrate that a theory T can be such that if T $\vdash A([\phi])$ then T $\vdash \phi$ while also being

such that $T \nvdash A([\phi]) \rightarrow \phi$.[21] Given Pollock's assumption that conventionalists equate truth and proof (an assumption that Giannoni accepts) this seems to block Pollock's argument for (T). Perhaps more to the point though, there are many theories that satisfy Giannoni's principle but not (T) for their own provability predicates, PA for instance (provided that it is ω-consistent).

Unlike Giannoni, I don't equate mathematical truth with M-provability for *any* given formal axiom system M, standardly conceived. I also reject Pollock's key assumption that M is, even if infinite itself, generated by finitely many syntactic conditions and rules of inference. As discussed at length in chapter 10, I think that the meaning-constituting rules of our mathematical language are both open-ended and infinitary. So I reject Pollock's key premise, that mathematical proof corresponds to M-proof for some specific formal theory M. And sophisticated conventionalists like Carnap also rejected it, long before Pollock's article was published, in fact. So I don't think that Pollock's argument gets off the ground when aimed at sophisticated versions of mathematical conventionalism, at least as he conceived it. But Pollock's argument can be extended.

The extension I have in mind makes Pollock's argument a version of Montague's paradox.[22] Instead of being concerned with a single axiomatic system like M, this version of the argument concerns our intuitive or informal notion of proof. As such, this version of the argument doesn't seem to have all that much to do with conventionalism. Here is Richard Grandy making a similar observation in an early review of Pollock's article:

> …the author must mean that we prove $[P_M([\phi]) \rightarrow \phi]$ informally not merely for any proof predicate of a formal system, but for the proof predicate of informal mathematics. But this would make classical mathematics inconsistent by the same argument. … Thus the conclusion drawn by Pollock concerning the conventionalist position, that we cannot both recognize the conventions and recognize that they are correct, seems equally applicable to the classical notion of proof.[23]

Although it is related to the paradox of the knower, it is fairly standard to see Montague's paradox as concerning our "informal", "intuitive", or "absolute" notion of provability when expressed with a predicate (as opposed to an operator).[24] If this is a problem for conventionalists, then it is a problem for everybody else too. Just as the liar paradox is a problem for everybody who accepts the notion of "truth". And as with the liar, there are a number of options for dealing with the paradox.

A fairly common suggestion tracing back to Myhill is that provability claims must be stratified into levels in the manner of the Tarski hierarchy for truth.[25] In 3.I, 4.VII, and 6.I I briefly appealed to a similar idea to brush aside worries about the inconsistency of the natural language notion of "validity" and related semantic paradoxes. All of these paradoxes are very similar, and in the present context, I think that all of them can mostly be ignored. Hannes Leitgeb puts a similar point nicely:

> This does not speak against informal provability in any way … this result does not even point to any substantial problem of formalizing informal provability in terms of a predicate … when we regard T and necessitation as valid for informal provability, we think primarily of *mathematical* instances … and maybe sentences or propositions which say something about the informal provability of such mathematical statements and maybe sentences or propositions that talk about the latter and the like, but not of self-referential or ungrounded statements such as "Informal Provability Liars".[26]

A detailed discussion would be distracting here. Feel free to substitute your own favored solution to natural language paradoxes of self-reference here. If that solution is at all adequate, it will preserve our ability to make the key generalizations about truth, validity, and informal provability that we really care about.

Those generalizations are all that the conventionalist needs in order to state their theory. To this point, it is important to keep in mind that the theory that is supposedly shown to be inconsistent by Montague's paradox (or the liar, or …) is not a mathematical theory,

but rather the English "theory" in which this book is written. When so interpreted, the status of (T) and (NEC) is much less clear. The sentences that are "informally provable" in some expansion of our mathematical language or "forced true" are the sentences of pure mathematics. As discussed in section II above, provability according to our conventional rules is itself a matter of fact. No claim needs to be made about informal provability claims themselves being true by convention in the needed sense. In any case, liar-like sentences can and should be mostly ignored when attempting to state and assess conventionalist principles.

Even if a conventionalist decided to do *nothing at all* to screen off liar-like sentences, all hope would not be lost. They could still apply a variant of 11.III's approach to the paradoxes of set theory. According to this approach, when the self-referential sentence "p" from above is added to our language, we find ourselves in an expansion of our previous language into which we cannot translate homophonically without contradiction. As with "set", "ordinal", "cardinal" and the like, the notion of being "provable in an expansion of our mathematical language" turns out to be indefinitely extensible.[27] As before, this wouldn't be worrisome in practice, since the conventionalist is concerned with pure mathematical claims and pure mathematical theories, not with quasi-empirical predicates that attempt to quantify over expansions of our language, either directly or indirectly.[28]

So Pollock's direct argument assumes that conventionalists identify mathematical truth with provability in some theory M, something I have not done here. And if this relativization to a theory is dropped from Pollock's argument, his challenge becomes identical to Montague's paradox. If this is a problem at all, it's a problem for everyone and is, in any case, no more or less troubling than the liar, the knower, and the other paradoxes of self-reference that we have been setting aside here.

V Gödel Against Conventionalism

In the early 1950s Gödel himself was asked to submit a paper for *The Library of Living Philosophers* volume on the philosophy of

Rudolf Carnap. But the volume appeared in 1963 without any contribution from Gödel.[29] It turns out that Gödel prepared six distinct drafts of a paper titled "Is Mathematics Syntax of Language?" but ultimately declined to publish any of them during his lifetime. Two drafts of the paper were finally published in 1995 in volume III of Gödel's posthumously published *Collected Works*.[30] In the paper, Gödel uses his second incompleteness theorem to argue against Carnap's mathematical conventionalism.

To set up his argument, Gödel defines "conventionalism" as the claim that mathematics somehow follows from the rules of syntax. He argues that if "syntax" is understood in anything like its standard sense, it must satisfy certain important constraints. For instance, he claims that syntactical rules cannot be infinitary but instead must be finitary, since any infinitary rules would not be "syntactic" in any standard sense. Carnap denied this in *Logical Syntax*, and I have denied it as well, but let's set this point aside for the moment. Gödel also argues for an even more controversial and involved constraint:

> Moreover, a rule about the truth of sentences can be called *syntactical* only if it is clear from its formulation, or if it can somehow be known beforehand, that it does not imply the truth or falsehood of any "factual" sentence (i.e., one whose truth, owing to the semantical rules of the language, depends on extralinguistic facts). This requirement not only follows from the concept of a convention about the use of symbols, but also from the fact that it is the lack of content of mathematics upon which its *a priori* admissibility in spite of strict empiricism is to be based. The requirement under discussion implies that the rules of syntax must be demonstrably consistent, since from an inconsistency *every* proposition follows, all factual propositions included.[31]

Gödel is here claiming that if we are to be conventionalists about mathematics, we must be able to demonstrate to ourselves that our syntactic rules for mathematics are consistent.

We can now set out Gödel's argument carefully (Gödel himself never sets out his argument in this explicit fashion, but I think this is a fair reconstruction). Again let M be our total mathematical theory:

1. If M is true by convention, then the consistency of M is demonstrable without significant mathematical assumptions (from Gödel's discussion above)
2. The consistency of M is not demonstrable in M or any weaker theory (from the constraints on theories and demonstrability implied by conventionalism together with Gödel's second incompleteness theorem)
3. *So*: If the consistency of M is demonstrable in theory T then T must have significant mathematical assumptions that go beyond M, in some sense (from 2, discussed below)
4. *So*: M is not true by convention (from 1,3 and the definition of M)

The idea is that if M's consistency must be demonstrated in a theory T with substantive mathematical assumptions more powerful than M's, then there is no sense in which M is true merely by the rules of syntax.

Gödel defends premise (3) as follows:

> In general the concepts and axioms occurring in the section of mathematics proved consistent need not all occur among, or be derivable from, those sufficient for a formalized consistency proof. … But the very fact that, if some of the essential concepts or axioms of the system are not contained in the system used for the consistency proof, others equally powerful (and equally problematic, if mathematical intuition is not trusted) must be present, reveals a principle which could perhaps be called the *non-eliminability of the mathematical content of an axiomatic system by the syntactical interpretation.*[32]

Assuming we ignore self-verifying theories and non-canonical provability predicates, Gödel's reasoning here is plausible.[33] It is true that (for example) Gentzen's consistency proof for PA can be carried out in theories that don't properly extend PA, but the theories have significant (and non-finitary) mathematical content by anyone's lights.[34] Gödel takes this point to refute any form of conventionalism and thereby open the door to Platonist theories of the kind he himself favored.[35]

Obviously, since my brand of conventionalism doesn't adhere to Gödel's finiteness constraint, it isn't in the target area of his argument. But even aside from this, I think Gödel's argument is mistaken. Its central failing is related to the translation mistake, yet again. As chapter 5's discussion of tonk and other inconsistent languages showed, in an inconsistent theory, there are no "factual" claims at all. A sentence ϕ in a language L is factual just in case neither it nor its negation is a theorem of L, but in an inconsistent language, every sentence is a theorem. So there is no true possibility of a "factual" sentence following from our conventions, because sentences themselves are not factual or non-factual independently of their place within a system of linguistic conventions. So in a very real sense, the danger that Gödel was attempting to head off with his premise (1) *does not even arise*.

Warren Goldfarb anticipated this point in his introduction to the published version of Gödel's paper:

> In this view of Carnap, there is no notion of "fact" or "empirical world" that is given prior to linguistic frameworks. The rules of a language, and hence mathematics, have to be in place before sense can be made of such notions. … [T]he notion of empirical fact is given *by way of* the distinction between what follows from the rules of a particular language and what does not. Thus, this reading of Carnap's view undercuts the very formulation of Gödel's argument.[36]

According to this reading, Gödel's argument is confused.

But even aside from this confusion, when taking things on his own terms, Gödel has not justified that the conventionality of M requires that M's consistency be demonstrable (premise 1). Why is it not enough that M simply *be consistent*? In ignoring this possibility, Gödel is assuming a brand of epistemological internalism that is implausible, in general, and anathema to the conventionalist, in particular (this type of internalism was rejected in chapter 6). So even if, pace my arguments, unrestricted inferentialism is false and consistency is a reasonable external constraint on truth by convention, demonstrable consistency need not be. On most

reasonable views, including my own unrestricted inferentialist view, premise 1 of Gödel's argument is false. Additionally, my discussion of the epistemology of consistency in 11.VII already showed that even internalists need to take extreme care when moving back and forth between claims about consistency and claims about consistency sentences.

Gödel's argument doesn't target the type of conventionalism I have developed here, embodies an implausible type of epistemological internalism that I have rejected, and seems misconceived on its own terms, even when the demands behind the argument are understood in an internalist fashion. I have already rejected all of Gödel's constraints on conventionalism, and think I've done so for good reasons. Given all of this, Gödel's argument fails to refute conventionalism, both on my terms and his own. There are other arguments that can be made to target mathematical conventionalism, but the most important and powerful have now been answered.[37]

≈≈≈≈≈≈≈≈

With this, my main development and defense of mathematical conventionalism is finished. Mathematical objects and mathematical truths, in different ways, are, like logical truths, reflections of the syntactic rules of our language. They have no deeper nature, and they answer to nothing external save for our varying needs—communicative, scientific, and the like. This combined philosophical theory of logic and mathematics is scientifically respectable and seems to me as comprehensive as one could hope. With my first-order presentation and defense finished, it is time to step back and view the theory, and my presentation and defense of it, in a broader context. This sort of thing is rarely done by proponents of major philosophical theories, at least in the light of day, but I think it needs to be done here, for all to see, given the sorry reputation of conventionalism in contemporary philosophy.

---

1 In Gödel (1931).

2 Rosser (1936) showed how to prove negation incompleteness without having to assume ω-consistency. An arithmetical theory T is ωinconsistent if for some open formula φ(x), for each numeral "n", T ⊢¬φ(n), but also T ⊢∃xφ(x), and ω-consistent otherwise.

3 I have abbreviated "S0" as "1" here and will continue to do so without comment below.

4 See Hilbert and Bernays (1939).

5 Though first stated in Gödel (1931), the second incompleteness theorem was first given a detailed proof in Hilbert and Bernays (1939).

6 Page 501 of Putnam (1994*b*)'s reprint of Putnam (1979).

7 Page 84 in Koellner (2009*a*).

8 From Ayer and Copleston (1957).

9 The point I am making here is not unrelated to Floyd and Putnam's (2000) reading of Wittgenstein's "notorious" paragraph on Gödel's theorem from the revised edition of his (1956). The main difference is that Floyd and Putnam seem forced to make a false claim about how the mathematical community would react to (for example) the ω-inconsistency of PA. Since I am concerned only with alternative practices, no descriptively false claim needs to be made to support the philosophical point.

10 Putnam's and Koellner's arguments are also discussed in my (2015*a*).

11 These points were made by Beth (1963) in a discussion of Carnap's *Logical Syntax* approach (see also 13.II).

12 See, for example, Quine (1940) and (1946) or Tarski (1933*a*).

13 In Kripke (1982) and Putnam (1980), respectively. In 10.IV I also noted that Putnam has used this type of argument against Field's approach to securing arithmetical determinacy.

14 Lewis (1984), page 225.

15 See Pollock (1967).

16 See Montague (1963) and Myhill (1960)—Pollock notes that his central theorem can be obtained from one of Montague's results.

17 Pollock (1967), page 241.

18 See Löb (1955).

19 Thanks to Scott Mutchnik for raising the question of what happens when a contradiction is substituted into Pollock's argument.

20 In Giannoni (1970).

21 His examples are from Curry (1963).

22 Thanks to Daniel Waxman for suggesting that I discuss this issue.

23 Grandy (1970)—I have updated the notation (in brackets) to match my discussion above.

24 See Dean (2014).

25 In Myhill (1960), see Tarski (1933*a*) for hierarchy approach to truth.

26 Leitgeb (2009), page 285.

[27] This is somewhat like the attempted super-paradox discussed at the end of my (2017*b*).

[28] Dummett (1963) also argues that "the class of intuitively acceptable proofs is an indefinitely extensible one".

[29] Schilpp (1963).

[30] Gödel (1953/9) in Feferman, Dawson Jr., Goldfarb, Parsons, and Solovay (1995).

[31] Feferman, Dawson Jr., Goldfarb, Parsons, and Solovay (1995), page 339. I have eliminated a footnote where Gödel notes that an inconsistency is also explosive in intuitionistic logic.

[32] Feferman, Dawson Jr., Goldfarb, Parsons, and Solovay (1995), page 345.

[33] See Willard (1993) and (2001) and Feferman (1960).

[34] See Gentzen (1936).

[35] See Gödel (1964).

[36] Feferman, Dawson Jr., Goldfarb, Parsons, and Solovay (1995), page 345.

[37] The main argument I have in mind is from Hauser and Woodin (2014), though their official target is "anti-realism" in the philosophy of mathematics. They note that any two large cardinal axioms are comparable in terms of their logical strength, by a comparison of their consistency sentences. They take this result to be highly unexpected on anti-realist views of mathematics, and thus to require a Platonist explanation. They spend most of their time aiming to undermine an alternative explanation of this widespread comparability, from Feferman (2007), saying that unless incomparability is explicitly built into the definitions, it should be expected. They give a "counterexample" to this claim, but admit that their counterexample is not *literally* a counterexample, since it does not actually involve two large infinity axioms. Even leaving this aside, I don't see that Platonism provides any explanation of comparability at all, so I think the entire line of thought is confused. It is certainly nothing that conventionalists need to worry about, even if they don't accept Feferman's response. On the general point, it is also perhaps worth noting the sociological point that even aside from Feferman, many other prominent logicians who have expressed opinions about the philosophy of set theory have argued that some form of anti-realism is the proper way to see things, given contemporary results and methods—see Hamkins (2012) and Shelah (2003) as well as the remarks on pages 479–482 of Kanamori (2003), a popular set theory textbook.

PART IV

REFLECTIONS

13

Old Slogans, New Dogmas

Back in 1.I I briefly traced the historical rise and fall of logical and mathematical conventionalism. And at various points throughout the book I've compared my approach to the views of previous conventionalists, most prominently Ayer and Carnap. But on the whole I have avoided most of the slogan formulations that historical conventionalists used when presenting their views. In this chapter, as a penance offering to the gods of philosophical pith, I discuss the canonical conventionalist slogans in light of the theory I've developed, to see which of them, if any, are still appropriate.

I will also discuss in some detail the two best books on conventionalism that I know of: Carnap's *The Logical Syntax of Language* and Giannoni's *Conventionalism in Logic*. Carnap's approach was more technical than philosophical, so Giannoni's is perhaps the only extended philosophical defense of either logical or mathematical conventionalism mounted before *Shadows of Syntax*. These quasi-historical discussions will lead into a quasi-sociological discussion of how it is that the rejection of conventionalism managed to become a dogma in contemporary philosophy.

I Conventionalist Slogans

In chapter 1 I noted that one of the major historical reasons for dissatisfaction with conventionalism was that most conventionalists relied on catchy, slogan statements of their views without ever giving a full and satisfying philosophical development and defense of those views. Ayer is the poster boy for this criticism, but it has some justice even when lodged against Carnap, probably the most sophisticated historical conventionalist (see the following section).

Parts II and III, against the background in part I, have now provided the needed development and defense. So let's turn back to the familiar slogans to see how they fare in the light of my theory.

Slogan 1: Logic and mathematics consist of tautologies.

The idea that logic consists of tautologies derives from Wittgenstein's *Tractatus Logico-Philosophicus*. In the *Tractatus* all meaningful sentences are analyzed as (possibly infinite) truth functions of fundamental atomic sentences and the logical truths are those sentences—the tautologies—that are true no matter the truth values of their atomic constituents.[1] Despite Wittgenstein's heroic efforts, the extension of this account to first-order logic, to say nothing of mathematics proper, is doomed to fail for various technical reasons.[2]

It is highly doubtful that first-order logical truths can be construed as tautologies in any plausible manner. Imagine trying to take a quantificational sentence like ⌜ $\forall x Fx$ ⌝ as equivalent to some infinitary conjunction, ⌜ $Fa \wedge Fb \wedge \ldots$ ⌝.[3] This equivalence is only plausible if every object has a name, which isn't something we should be forced to assume by our philosophy of logic. Even if every object has a name, if there are uncountably many objects, this conjunction will not just be infinitely long, it will be uncountably long. What is worse, even if we countenance uncountable conjunctions, in order for any conjunction, however long, to be equivalent to ⌜ $\forall x Fx$ ⌝, it would also need to include the information that *everything* was named in the conjunction. But this amounts to using universal quantification to analyze universal quantification, so the conceptual goodstanding of this approach is doubtful.

Aside from these quasi-technical problems, the conjunction analysis of quantification also seems to get things backward philosophically. Using a primitive uncountable conjunction operation to explain quantification is like using complex analysis to explain addition. These are all problems even at the level of first-order logic. When we consider applying slogan 1 to mathematics, things becomes even more doubtful.

In short, under its standard meaning, slogan 1 is false. Conventionalists who endorsed this slogan were either confused,

wrong, or, perhaps more likely, using “tautology” in a nonstandard and more expansive sense.

Slogan 2: Logical and mathematical truths are analytic/logical and mathematical truths are are conceptual truths.

In my framework, a natural definition of “analyticity” is:

Analyticity: A sentence ϕ is analytic in language *L* just in case its truth is fully explained by the automatic validity of the basic rules of *L*.

I have argued that logical and mathematical truths are analytic, in this sense (4.I, 11.I). In 3.III I also briefly discussed the evident fact that my account of logical truth plausibly generalizes to non-logical conceptual truths. There I gave inference rules for the word “bachelor”, including the following rule:

$$(BE)\quad \frac{\alpha \text{ is a bachelor}}{\alpha \text{ is a man}}$$

Using (*BE*), conditional proof, and universal introduction the following sentence will be provable from no premises:

$$\forall x(x \text{ is a bachelor} \rightarrow x \text{ is a man})$$

So the automatic validity of the basic rules—including the automatic validity of (*BE*)—shows that this sentence is necessarily true, in the strongest sense.

In short, slogan 2 is true. The logical and mathematical truths are analytic or conceptual truths. It is just that some analytic truths are neither logical nor mathematical. On my view though, all of these truths find their ultimate explanatory source in our linguistic conventions.

Slogan 3: Logic and mathematics consist of definitions/logical and mathematical truths are true by definition.

A natural reading of this slogan requires explicit definitions or acts of defining. But as I have stressed throughout the book, linguistic conventions are best seen as implicit linguistic rules (2.III). Stipulations or acts of defining are metasemanticaly idle (2.II). If I proclaim that I have defined "bachelor" to mean what "philosopher" means but then continued to use it in the old sense, never calling a woman a "bachelor" and so on, the term would still have its old meaning, despite my apparent re-definition (see also the discussion of slogan 4 immediately below).

In short, on the most natural reading, slogan 3 is false. But while it is probably best to avoid this slogan, the related idea that logical/mathematical truth is simply a byproduct of the way we use words is obviously central to my theory. And I agree that basic rules *implicitly* define logical and mathematical terminology. I just think that too much talk of "definitions" is likely to mislead.

Slogan 4: Logical and mathematical truths are true by declaration/stipulation.

As stressed in 2.II and when discussing the previous slogan, stipulations are metasemantically idle. What matters for metasemantics is always the underlying pattern of language use, not any stipulations or declarations we may offer, even if they are accompanied by shouting and a stomp of the foot.

I admit that stipulation and declaration are useful heuristics for thinking about these issues, but we should always keep their proper role in mind and not get carried away. Stipulation cannot be used to found truth. I worry that this is too often missed. In a book-length treatment of analyticity, Cory Juhl and Eric Loomis introduce the following "coordinative rule[s] concerning some stipulation sentence *s*":

> (Stip) Sentence *s* expresses some true proposition *p* (in our language L). Furthermore, the proposition *q*, *that s expresses a true proposition* (in L), is empirically indefeasible. No empirical evidence counts in favor of or against the truth of *q*.

(TStip) Sentence *s* expresses some true proposition *p*. Furthermore, the proposition *q*, *that s expresses some true proposition p*, is empirically indefeasible. Finally, proposition *p* is empirically indefeasible (no empirical evidence counts for or against the truth of *p*).[4]

Juhl and Loomis think this notion can do useful philosophical work, but in addition to being implausibly conceptually demanding, I think the focus on "stipulations" is misplaced. Something can only be empirically indefeasible if our use makes it so. To hit the long-dead horse one final time: metasemantics is not under the control of stipulation.

In short, slogan 4 is false, though it is sometimes useful for purposes of illustration.

Slogan 5: Logical and mathematical truths are true in virtue of meaning.

In the contemporary philosophical world, this is one of the more popular slogan statements of analyticity.[5] This is probably because of the central importance currently given to meaning and theories of meaning. If "truth in virtue of meaning" is used as a synonym for "analyticity", then slogan 5 stands or falls with slogan 2, which I have already endorsed.

The only worry I have about slogan 5 is that it is apt to suggest a misleading picture of meaning in which *meanings* are doing fundamental load-bearing work in metasemantics. In my view, any such picture is both scientifically dubious and metaphysically mysterious (1.IV). So the slogan is acceptable only when it is firmly kept in mind that meanings are themselves ultimately explained by way of our linguistic conventions, rather than vice-versa.

In short, slogan 5 is true on at least one plausible understanding, but potentially false and misleading on other natural ways of understanding it.

Slogan 6: Logical and mathematical truths are true by convention/according to the rules of language.

The theories of logic and mathematics developed in parts II and III straightforwardly explain the sense in which logic is "true by convention" or "true according to the rules of language" or what have you. Of the standard slogans, this one is probably the least misleading and the most free from unwanted associations, which is why I have used versions of it throughout the book. I have also used a new slogan that I take to stand with this one.

Slogan 7: The source of logical and mathematical truth is linguistic convention.

In short, slogans 6 and 7 are true.

The final slogan is also new.

Slogan 8: Logical and mathematical truths are shadows of syntax/reflections of our linguistic rules.

Variants of this slogan have also been used throughout the book, including in giving the book its title. I think the metaphor here is powerful and suggestive. The full case for the truth of this slogan is found in the first three parts of the book. In brief, logical and mathematical facts are explained wholly in terms of syntactic rules of inference. Even in mathematics, mathematical truth is explained using the syntactic rules of inference we actually follow. Complex though they may be, the truths they generate are still shadows of syntax, still reflections of rules, now cast on a point at infinity.

The final scorecard shows that slogans 2, 6, 7 and 8 are upheld by my theory, 5 is a mixed case, and the others are false at worst and true but misleading at best. This situates my theory with respect to those conventionalists whose views didn't go far beyond the slogan phase. How does my view compare to the few conventionalist views that did go beyond slogans?

II Carnap's *The Logical Syntax of Language*

Carnap's *Logical Syntax of Language* is a strange and wonderful book. The book has generated a large, high-quality secondary literature as of late, so I can be fairly brief. The German edition of *Syntax* was published in 1934 and was followed by an English edition in 1937 that included several sections that had been left out of the German edition along with a number of corrections.[6]

In *Syntax* Carnap definitively broke with the *Tractatus*-inspired logical monism he had previously endorsed in favor of a radical logical pluralism embodied in the famous principle of tolerance.[7] Carnap's book is strange, in part, because it is a mix of technical and philosophical aims. Carnap was in the vanguard of mathematical logic in the early thirties.[8] He personally knew and discussed logical matters with both Gödel and Tarski. From the modern point of view, one of *Syntax*'s aims was to give an successful account of logical consequence and mathematical truth in light of Gödel's incompleteness theorems.[9] In his terminology, he was trying to characterize "analyticity"—logical/mathematical truth.

In the book Carnap introduces two specific languages, given the uninspired names of "Language I" and "Language II". Language I is meant to satisfy finitist and constructivist assumptions. From the modern point of view it is basically a version of primitive recursive arithmetic (PRA) with bounded quantification. Language II is meant to embody classical mathematics and is basically a version of the finite simple theory of types built over arithmetic.[10] Carnap defines a consequence relation for each of these languages and uses arithmetization to show that they can talk about their own syntax.

In doing this, Carnap makes an important distinction between "definite" and "indefinite" notions. In modern terminology, "definite" notions are recursive.[11] In particular, Carnap distinguishes between derivability (D) and consequence (C). In effect, the D-rules of a language are its recursive rules of proof, while the C-rules are its non-recursive rules. To provide a general definition of consequence for Language I Carnap uses a version of (ω) as a C-rule. But because Language I doesn't have unbounded quantification, the generality of the conclusion is expressed instead with a free variable. Rather than closing the axioms under (ω) in the modern way, Carnap

instead uses his rules to define a relation of direct consequence between sets of sentences of the language and considers finite sequences of sets standing in direct consequence relations, ending with the singleton of the conclusion. This seems to come to the same thing as the modern approach.[12] The "analytic" (mathematically true) sentences of Language I are the consequences of the null set.

For the much richer Language II Carnap gives a more complex account. Instead of starting with consequence and then defining "analytic" as he did for Language I, for Language II Carnap does the reverse, starting with "analytic" (and "contradictory") and then defining "consequence". It is often claimed that Carnap's Language II approach is a somewhat garbled version of Tarski's famous theory of truth.[13] It is garbled largely because Carnap doesn't offer a simple recursion on "analytic" in the manner of Tarski on "true".[14] Instead he gives various more or less complex rules of reduction and evaluation. For the logical and mathematical terminology of Language II (but not for the descriptive terminology), Carnap's account is essentially equivalent to Tarski's. A Tarski-style streamlining was suggested by Kleene in an early review and was accepted by Carnap.[15] With "analytic" and "contradictory" defined for classes of sentences of Language II, Carnap says (in modern notation) that a sentence ϕ is a consequence of a set of sentences Δ just in case $\Delta \cup \{\neg\phi\}$ is contradictory.[16]

In addition to these two specific languages, Carnap also attempts to offer a general characterization of analyticity. His general approach starts with an arbitrary consequence relation and appeals to a principled division between "logical" and "descriptive" vocabulary. For languages I and II, Carnap simply stipulated which vocabulary was "logical", but to give a general account he attempts to characterize these notions for arbitrary languages. Why does he have to do this? Largely because he allows his formal languages to include physical vocabulary and even physical rules of inference (P-rules) that are distinguished from the logical/mathematical rules (L-rules, which include both D-rules and C-rules). In including P-rules, Carnap was in danger of obliterating the distinction between physical laws and logical principles. Distinguishing "logical" and "descriptive"

expressions is part of an effort to set things right. In attempting to characterize the logical constants formally, Carnap was again—as with his accounts of consequence and analyticity—attempting to tackle a problem that would later be given a widely accepted mathematical solution by Tarski.[17]

Carnap's approach involved assuming that sentences composed wholly of logical/mathematical terminology were determinate (either analytic or contradictory), and then using this assumption to characterize the "logical" vocabulary in a language. Unfortunately Carnap's approach is technically flawed in both minor and major ways, as Saunders MacLane first pointed out in an early review.[18] Carnap eventually recognized that he hadn't provided a solution to this key problem, and continued to work toward it during his later "semantic" period, without much success:

> So far we have discussed the distinction between logical and descriptive expressions only in the form in which it appears when we have to do with a particular semantical system. … The problem is more difficult in the form it takes in *general semantics*. … A satisfactory solution is not yet known. … The distinction between logical and descriptive signs and the distinction between logical and factual truth belong to the most important problems of logical analysis. … This problem is very much in need of further investigation.[19]

I discussed what I think is distinctive of the logical notions in 3.III, but my approach appealed to features of the basic rules of our language. Carnap couldn't appeal to any such distinction, since he was trying to solve the problem for purely formal languages, and had also—as noted above—allowed P-rules among the linguistic rules.

This is one point where his characteristic blurring of the technical and the philosophical becomes a serious problem. If Carnap's point was merely that we can frame *physical theories* in something like Language II, that would be fine. But taken as philosophy, his claims about P-rules are deeply misleading. Schlick, the Vienna Circle's leader, was the first to protest about this:

> It is true that a sentence (a sign sequence) which, under the presuppositions of customary grammar, expresses a natural law can be made into a principle of language simply by stipulating it as a syntactical rule. But precisely by this device one changes the grammar and, consequently, interprets the sentence in an entirely new sense, or rather, one deprives the sentence of its original sense. It is then not a natural law anymore at all…[20]

In effect, Schlick is noting that *Carnap* was committing the translation mistake! If you take a sentence that currently expresses a law of nature and make it a linguistic rule, you thereby make it strongly valid and in the process deprive it of its status as a law of nature. This is similar to what happened with the stipulated ontological argument in 9.V. Arguably, Carnap was led into this misstep through the influence of other members of the Vienna Circle.[21] In any case, he never really managed to disentangle the logical and the factual in a way that finally satisfied him.

In *Logical Syntax* Carnap also discussed the ontology of mathematics. His position seemed to be that a mathematical referring claim in the material mode, like "five is a number," is misleading and should be replaced with a formal mode claim like "'five' is a number-word". Like my account of metaontology in chapter 9, Carnap's approach sees syntax as accounting for (or eliminating) ontology. But his *Syntax*-era metaontology is confused and confusing because it eschews reference and other semantic notions. Carnap's best and fullest statement of his mature metaontological views is in his later "Empiricism, Semantics, and Ontology", which I see as an important forerunner to my account (and the accounts of Putnam and Hirsch).[22]

It is at best unclear to what extent Carnap saw his proposals as applicable to natural language.[23] He doesn't do anything much to connect his proposals to our actual logical and mathematical practices, in sharp contrast to my discussion here.[24] From his pre-syntax days and throughout his trek through semantics, Carnap preferred formal statements framed in explicit metalanguages.[25] But as Beth pointed out in his contribution to the *Library of Living*

Philosophers volume on Carnap in the early sixties, *Logical Syntax* itself must be written in a syntactic metalanguage that is assumed to be standard, otherwise (for example) theories like Q+¬Con(Q) and Q will be seen as the very same theory.[26] In response, Carnap appealed to the supposed fact that natural language arithmetic is to be understood standardly, as is customary, so his syntactic metalanguages are, likewise, to be understood standardly.[27] This is pretty weak dialectically, but it is even more unsatisfying philosophically, at least without significant further supplementation.

What we need instead is an account of *how* we manage to have a *standard* grasp of syntax. I have attempted to provide this here, in a more or less Carnapian fashion (chapter 10, 11.I, 12.III). Carnap himself never offered anything much on this front. Perhaps he never meant to. But if this is the case then it is hard to see him as offering a philosophy of mathematics at all. Perhaps he was instead only proposing a quasi-formal toolkit for thinking about logic, mathematics, and the language of science? This dovetails with the recently popular idea that from *Logical Syntax* onward, Carnap was not really a logical or mathematical conventionalist, philosophical folklore be damned. I will discuss this recent trend in Carnap scholarship in 14.I, for now let us turn to the only major attempt to (partially) discharge the philosophical burdens of conventionalism of which I am aware.

III Giannoni's *Conventionalism in Logic*

To my knowledge, the only extended *philosophical* defense of logical conventionalism—before the present volume—is Carlo Giannoni's *Conventionalism in Logic* from 1971.[28] Giannoni's book is fairly short, and at this remove in time much of his discussion feels a bit quaint. Despite this, he makes a number of important distinctions that were rarely if ever made by previous conventionalists or their critics.

Crucially, Giannoni distinguishes between something like the three versions of conventionalism I distinguished in 1.II and correctly opts for the explanatory version, according to which logical truth and

validity hold “in virtue of” our linguistic rules. The only inadequacy in his account of this is that Giannoni didn’t have the expressive resources available to flesh out the “in virtue of” relation in a satisfactory way. Instead he ultimately opts for a quasi-epistemic reading in terms of “justification”, but to his credit he decries the inadequacy of the epistemic terminology:

> …in general I would say that the terminology of “justification” is unfortunate when applied to the statements themselves. I do not think we should ask what justifies a statement, but rather why is the statement true and why is it necessary?[29]

For this reason, despite the terminological differences I don’t think I’m coloring too far outside the lines when I say that Giannoni’s brand of conventionalism dovetails nicely with the explanatory conventionalism I have endorsed here.

Giannoni spends a significant amount of time attempting to clarify the nature of “linguistic rules” and their relation to other types of rules.[30] His account is that linguistic conventions are regularities of behavior from which prescriptions arise in a causal sense, as they often do from regularities. Linguistic rules are, in this way, the prescriptive aspects arising from these regularities. Giannoni sees entailment statements as *descriptions* of these regularities while rules concern their prescriptive function:

> In general, entailment statements DESCRIBE the linguistic regularity, while the rules of inference PRESCRIBE that one go on in the regular way.[31]

But while rules and entailments statements are obviously distinct, Giannoni’s characterization of entailment is mistaken. A description of a linguistic regularity would be an empirical claim like “speakers of English infer in such-and-such a manner”, but this clearly isn’t a statement of entailment or validity.

In effect, Giannoni has collapsed several distinct levels of analysis here. This is a major defect of his account, since while he seems committed to the crucial conventionalist thesis that linguistic rules

are automatically valid and even says things indicating this several times, he never fully explains why this is so on his account. Indeed, he can't really do so in an adequate manner, since for him entailment statements are *literally identical to* descriptive claims about how speakers behave.

As one particular illustration of this defect, when Giannoni comes to giving explicit characterizations of logical truth and similar notions, he relies upon a split picture. The truth and falsity of atomic sentences is given by way of some correspondence theory of truth and so is independent of our linguistic conventions, while by contrast, the truth and falsity of non-atomic sentences is given by their deductive relations to atomic sentences:

> If A is a nonatomic sentence, then A is TRUE if and only if either A is logically deducible from true premises, or only true conclusions are deducible from A together with other true statements as premises, or both.[32]

But this division seems to completely undermine conventionalism and its attendant logical pluralism (which Giannoni seems keen on endorsing). To illustrate: let α be a false atomic sentence and imagine that we adopt a package of rules R that are trivial (say they include the tonk rules) and also include the familiar rules for "and". Obviously, in this system, where β is a true atomic sentence, we have $\beta \vdash \alpha \wedge \alpha$ by the tonk rules, so "$\alpha \wedge \alpha$" is true since deducible from a true premise, and yet $\alpha \wedge \alpha \vdash \alpha$ by $\wedge$-elimination, and α is false while $\alpha \wedge \alpha$ is true, so the rule of $\wedge$-elimination is not valid, contrary to Giannoni's suggestions elsewhere that any rules of inference we adopt are automatically correct.

A version of "conventionalism" where our linguistic rules must meet some external standard and thus risk being invalid is not really a version of "conventionalism" at all, according to standard parlance (4.II). I don't think that Giannoni intended this seemingly anti-conventionalist result, but his discussion is difficult to follow at some points, so I'm not entirely sure. At the very least, this result *seems* to be in direct tension with a number of things that he says elsewhere in his book. I think the simplest explanation is that Giannoni was led

into a bit of muddle by his inadequate account of “entailment” as it relates to linguistic regularities.

Nonetheless, Giannoni’s discussion was, in several respects, an important philosophical advance over earlier treatments of logical conventionalism, including Carnap’s. Sadly, from what I am able to determine, I am perhaps the only person who ever read Giannoni’s book, some 45 years after its publication.[33]

IV New Dogmas

Conventionalism is not just rejected by contemporary philosophers, it is *vehemently* rejected. It is *despised* and *reviled*, to the point where anything that hints at conventionalism is immediately and thoughtlessly rejected. Theorists whose views come close to conventionalism often take great pains to distance themselves, lest they too be dubbed heretics. The rejection of conventionalism is now one of the central dogmas of contemporary philosophy.

What explains this state of affairs? Early in the book, in 1.I, I listed a few of the philosophical reasons for the historical failure of conventionalism. I still think that these were historically important, and an overestimation of the philosophical case against conventionalism likely led many philosophers to reject the view without any serious investigation. But I don’t think this can be the entire story. The situation is odd. For example, the master argument, discussed in 7.II, has gone almost unchallenged in the literature until very recently.[34] That is extremely rare in philosophy, for *any* argument.

In this book I think I’ve demonstrated that the arguments against conventionalism are either problematic on their own terms or aimed at implausible straw men. I hope my discussion convinces people to give conventionalism another look. But philosophical arguments only go so far when xit comes to influencing public opinion, no matter their power. There is also a major sociological reason that conventionalism has been widely rejected, one that goes beyond the lingering influence of Quine (which is itself a major philosophical, historical, and sociological factor).

That reason is that conventionalism seems to make things *too easy*. I mentioned this in the coda to chapter 3 when I compared unrestricted inferentialism's ability to slice through philosophical tangles to Alexander's famous hack through the Gordian knot. Conventionalism not only solves or dissolves many classical problems, it does so easily. Too easily, perhaps. I suspect that many philosophers harbor worries about this. Perhaps because they don't really *want* philosophical problems to be solved. I don't know. I am not sure how widespread this feeling is, but a number of philosophers have personally expressed to me, either directly or indirectly, a desire for philosophical problems to remain open and difficult, and for theories to be as technical and complicated as possible.

I don't share this desire. I am interested in finding out the truth and resolving the issues that have puzzled me for years. For me, this is not a game. It is deeply personal. The nature of logic and mathematics have long troubled me, but they trouble me no longer. The theory presented in this book has cured me, at least of these particular troubles. But this is confession, not argument.

≈≈≈≈≈≈≈

MY CONVENTIONALIST THEORY used various tools of contemporary philosophy to make the most detailed and sophisticated case possible. Positive cases for theories sometimes devolve into checklist-style approaches to answering objections, with silly objections given parity with serious objections. By contrast, I have tried to offer a sustained development of a positive theory, without adopting a misleading stance of impartiality concerning the strength of objections, or bogging my discussion down with too many criticisms of alternative approaches. In my happier moments, I think that the theory I've developed here is substantially correct and deeply illuminating. But this type of thought is not usually voiced in contemporary philosophy. Now that I have voiced it, it must itself be put under the microscope, so that my doubts and meta-doubts can grow from barely visible specks to dominant features of the formerly desert landscape stretching out before us.

[1] See Wittgenstein (1922).

[2] See Fogelin (1976), Geach (1981) and (1982), and Soames (1983) for discussion of some of the logical failings of the system of the *Tractatus*.

[3] This is similar to Wittgenstein's (1922) approach, though instead of conjunction he uses a generalized "neither-nor" operator, so that no other connectives are needed (Sheffer (1913) showed that neither-nor is truth-functionally complete). But Wittgenstein's approach is unable to make needed scope-distinctions and so can't represent both of the standard quantifiers without slight emendations.

[4] Quoted from Juhl and Loomis (2010), pages 218–219.

[5] It gives the title to Russell (2008), for example.

[6] Carnap (1934) and (1937).

[7] See Friedman (1997) and Awodey and Carus (2009) for discussion of the relationship between the *Tractatus* and *Logical Syntax*.

[8] As one illustration of this, *Logical Syntax* includes the first statement of the general diagonal lemma.

[9] Tarski (1936) credits Carnap with the first serious attempt to come up with a precise and satisfactory definition of "logical consequence".

[10] See Carnap (1931) for an introductory account of the theory of types.

[11] Koellner (2009*b*) notes that Carnap gives different and inequivalent characterizations of "definiteness", but understanding it as recursiveness seems to best capture his intentions.

[12] See de Rouilhan (2009) and the appendix to chapter 10.

[13] See Coffa (1987) and (1991).

[14] See Awodey (2007) and Coffa (1991).

[15] In Kleene (1939) and Carnap (1940).

[16] Carnap (1937), page 117—right after giving this definition Carnap notes (to use modern terminology) that this consequence relation is not compact.

[17] In Tarski (1986), which is the text of a lecture given in 1966.

[18] See MacLane (1938) and Awodey (2007) for discussion; Quine (1960*a*) and others later raised related problems. See Bonnay (2009) for general discussion and an attempt to patch a Carnap-style approach.

[19] Carnap (1942), pages 59, 242–243.

[20] From Schlick (1936), quoted from page 445 of Schlick (1979). The quoted bit is from a footnote Schlick added to the published version of a talk he delivered. Amusingly, Carnap heard the talk but didn't originally know who the target was until he wrote to Schlick—see the discussion on page 351 of Coffa (1991).

[21] This is argued for in chapter 18 of Coffa (1991); see Goldfarb (1997) for some pushback against Coffa's overall reading of Carnap.

[22] Carnap (1950)—see my (2016*c*) for an even closer parallel.

[23] Doubts occur as early as page 2 of his (1937).

[24] Though see his (1955).

[25] Carnap's "semantic" phase is represented in his (1939), (1942), (1943), and (1947). He was influenced here by Tarski (1933*a*).

[26] See Beth (1963) and Friedman (2016) for discussion.

[27] In Carnap (1963*b*).

[28] Giannoni (1971); Syverson (2002) purports to be a defense of conventionalism, but doesn't seem to contain a philosophically worked out discussion. While his book doesn't defend conventionalism about arithmetic, Giannoni did accept the extension, as can be gleaned from his (1970), which is reprinted as an appendix in the book.

[29] Giannoni (1971), page 78.

[30] His discussion of this issue is influenced by Black (1958) and Bowdery (1941).

[31] Giannoni (1971), page 57.

[32] Giannoni (1971), page 121.

[33] Even more sadly, Kolenda (1982) reports that Giannini died in 1981 at the age of 42. The echo of Russell (1903) on Frege is intentional.

[34] Russell (2008) is the only response I know of that predates my (2015*d*).

14

The Facts of the Matter

In my day-to-day life, I believe that conventionalism is true. Yet, like the logical positivists of yore, I harbor strong deflationary instincts about large swaths of philosophy. I suspect that a great many philosophical questions are either pseudo-questions or have no uniquely correct answers. These deflationary instincts threaten to destroy the foundation upon which the conventionalist cathedral has been built. So I owe the reader, and myself, an account of the metaphilosophical status of my convictions.

I'm also not blind to the bleak reality that I am basically the lone defender of logical and mathematical conventionalism in the last fifty years. If conventionalism is true, how have I alone been so lucky as to hit upon the truth? Do I put no stock at all in the disagreement amassed against me? Intellectual honesty demands that these concerns be squarely faced.

I Factualism in Metaphilosophy

I think that conventionalism is, at some fineness of grain, the uniquely true and correct theory of both logic and mathematics. If I didn't think this, I wouldn't have written this book. In fact, if I didn't think this, I wouldn't have *been able* to write this book. This is not to say that I don't have my doubts in reflective moments. These doubts will be discussed below. But first I want to try to justify my day-to-day attitude toward conventionalism.

My attitude here contrasts sharply with the attitudes that some of the logical positivists (seemingly) had about their own conventionalist theories of logic and mathematics. For a particularly clear illustration,

witness Reichenbach saying, in the course of giving a general treatment of induction and meaning:

> It would be entirely erroneous to ask: What is the true conception of meaning? or which conception *must* I choose? Such questions would be meaningless because meaning can only be determined by a definition. What we could do would be to propose the acceptance of this decision.[1]

This amounts to a non-factualism about the metasemantics that underwrites conventionalism. Obviously, if an inferentialist metasemantics is itself simply a mere matter of truly arbitrary choice, then there is no real sense in which conventionalism is the correct theory of logic and mathematics.

But just because some of the historical positivists flirted with a radical non-factualism about the very status of conventionalism does not mean that it was wise of them to do so. Alberto Coffa's commentary upon this strand of Reichenbach's thought is both amusing and devastating:

> If consistently pursued, these conventionalist remarks have a destructive effect upon Reichenbach's best work and, in the end, upon themselves. For example, Reichenbach's classic analysis of the conventionality of simultaneity would depend entirely on the adoption of a semantic convention that, on Reichenbach's own view, there is no *factual* reason to adopt. … the claim that "it is meaningless" to ask for a factual account of meaning … must be put forth either as a convention or as a statement of fact. In the former case, Reichenbach's wrath seems unjustified; in the latter, more likely alternative, the question is, Which facts of the matter has Reichenbach uncovered to establish the conventionality of meaning? This forgettable semantic doctrine was frequently forgotten by Reichenbach when he got down to the business of actual philosophizing, in which he was usually inspired by a rather extreme form of semantic intolerance.[2]

As Coffa notes, Reichenbach and other positivists pursued matters of meaning with a zeal that belies their professed metasemantic nonfactualism ("nonfactualism" is, somewhat confusingly, termed "conventionalism" about semantics in Coffa's quote). If they are *really* logical and mathematical conventionalists, then they also need to be metasemantic factualists. The positivists have traditionally been interpreted as conventionalists, despite their occasional self-undermining remarks; but the traditional view is starting to become passé in posh philosophical circles.

In the last several decades there has been a growing trend in favor of a radical reinterpretation of positivist philosophy according to which even Carnap, the canonical conventionalist of philosophical legend, was not really a conventionalist, didn't accept any notion of truth by convention, and wasn't even interested in providing an explanatory philosophical theory of logic and mathematics.[3] The reading was first argued for extensively by Thomas Ricketts:

> … Carnap rejects any notion of truth-in-virtue-of. Although Carnap aims to clarify the distinction between factual truths and formal or linguistic truths, his explication does not draw on any notion of truth, on a notion of something's making a statement true. … it is not apt to attribute to Carnap a conventionalist philosophy of mathematics.[4]

This new reading of Carnap has become quite influential in the secondary literature. It is now fairly common for interpreters to simply assume that Carnap wasn't *really* a conventionalist, thinking instead that he was merely offering proposals and not seriously advocating or endorsing any particular philosophical theory, conventionalist or otherwise.

I ended 13.II with a related suggestion, but when taken to an extreme, I start to find the new reading of Carnap implausible. It is worth quoting Hilary Putnam, who knew Carnap personally, when he replied to Ricketts by saying:

> Ricketts's Carnap, the Carnap who holds no doctrines but only asks for "clarification," ... is just not the Carnap I knew and loved.[5]

Obviously knowing someone personally isn't hermeneutic magic, but it does mean that you are plugged into many different high-quality channels of information that can aid in your interpretation of the person's work. In addition to Putnam, Quine certainly saw Carnap as a conventionalist, as is evident from his famous critical pieces on conventionalism, especially "Carnap on Logical Truth".[6] Those who knew Carnap personally and worked with him generally interpreted him as a conventionalist. And Carnap himself, in his late, retrospective remarks, always reaffirmed conventionalist doctrines in the strongest terms.[7] Perhaps this isn't completely decisive on its own, but it is certainly strong evidence that Carnap was a conventionalist, at least in some sense of the term.

The water is a bit muddy here because various related but distinct theses can be expressed with the same words. The issue is certainly not simply whether the sentence "logic and mathematics are true by convention" is itself true by convention. One could think that what we mean by "meaning" is ultimately a byproduct of our usage of the term, without thinking that there are alternatives that are *equally correct*. As we have seen at many points in this book (especially in 5.I and the rest of chapter 5), to say that something is a convention is not to say that it is arbitrary. If Carnap thought that mathematical truths in natural languages were analytic, then he was a conventionalist. I suspect that he did, so I suspect that he was. But Carnap's ambivalence toward natural languages, as opposed to formal languages, further muddies the waters.

When evaluating this issue it is also important to realize that the Carnapian principle of tolerance isn't in conflict with factualism about metasemantics. From *Syntax* to the end of his life, Carnap allowed that constructivists (for example) were free to craft and use their own restricted language if they so desired. But at the meta-level Carnap argued against constructivist restrictions. Here he is doing so in §43 of *Logical Syntax*:

> The lack of a method of resolution for indefinite terms has induced many logicians to reject these terms altogether, as meaningless (e.g., Poincaré, Brouwer, Wittgenstein, and Kaufmann). … To this it may be replied: it is true that we know of no method of searching for the answer, but we do know what form the discovery of the answer would take—that is to say, we know under what conditions we should say that the answer had been found. … Thus there exists the *possibility of the discovery of an answer*, and there appears to be no cogent reason for rejecting the question.[8]

This is certainly not the reply of "well, that is an alternative metasemantics under which tolerance is false!" As Coffa suggests, Carnap might have briefly flirted with both sides on this issue, or at least said many confusing—and perhaps confused—things. But when he did, he shouldn't have.[9] In his writings on logic and mathematics, as well as metaontology, Carnap allowed everyone the freedom to construct their own way of talking. Nobody disputes that. So there is a sense in which he would have regarded constructivism as a mere proposal, as good as any other (metaphysically speaking). But as the quote above illustrates, there is also a sense in which he rejected constructivist restrictions at the meta-level. I suspect his attitude toward tolerance was similar, and that this crucial ambiguity is responsible for the debate on this point.

Of course, it ultimately doesn't matter whether or not Carnap, Reichenbach, or any of the other logical positivists were *really* conventionalists about logic and mathematics. If the positivists accepted implausible metasemantic and metaphilosophical views that undercut their (apparent) conventionalism, then they did so to their own discredit. Metasemantics is a matter of fact. This isn't to say that what our utterances mean is always fully determinate, only that any semantic indeterminacy in a given utterance is itself a matter of fact. The factuality of metasemantics, at least in this sense, is fully compatible with a healthy amount of indeterminacy about meaning, reference, translation, and the like.

Metasemantic factualism is also compatible with a wide-ranging non-factualism about many other branches of philosophy, including

heavyweight metaphysics of the kind the Vienna Circle abhorred. The human activity of philosophy is not a *natural kind*. It is instead a wildly heterogenous collection of activities, in many cases held together only by contingent historical associations. Given this evident fact, there is no reason to expect a unified metaphilosophical analysis of the whole of philosophy. Because of this, a positivist-style non-factualism about much of philosophy is compatible with robust factualism about *some* branches of philosophy, including metasemantics. As I said elsewhere when discussing this issue: we must keep the ladder in place after we have ascended upon it.[10]

Of course, the factuality endorsed in metasemantics is only a kind of *scientific* realism, not any kind of *metaphysical* realism (1.IV). In general, the factual branches of philosophy are somewhat akin to branches of science, though they differ from science in various ways. In science, but not philosophy, we have rigid constraints on our theories in the form of experimental predictions and other direct causal anchors. In philosophy there are constraints too, but the constraints themselves can be called into question, and even when applicable they are more nebulous than they are in science. This doesn't mean that philosophical theories cannot be compared, nor does it mean that such comparisons can't show that some theories are objectively better than others. Theory choice is messy, but it isn't entirely subjective. I think this is beyond serious dispute, but several historically influential philosophers of science have denied it.[11] Subjectivism and non-factualism are even more tempting in the case of philosophy, but I think the temptation should be resisted. Despite all of the noise involved, it is reasonable to think that some branches of philosophy—including metasemantics—are, in some fashion or other, *matters of fact.*

But the factuality of metasemantics alone doesn't show that conventionalism is the uniquely correct philosophical theory of logic and mathematics. Nor does it even show us what it would truly mean for that to be the case.

II How Conventionalism Could Be Factually True

Given the evident factuality of metasemantics, there is a fact of the matter about the correct philosophical theory of logic and mathematics. I've endorsed and defended conventionalism/unrestricted inferentialism about these branches of discourse, but let me say a bit—from a general, high-level perspective—about the overall case I have made for these theories and why you should think the theory of the this book is right, true, and correct (this is where I lay it on a bit thick, so steel yourself—penance will be offered after the fact).

Conventionalism should be accepted because it is the best explanation of everything that needs to be explained in logic and mathematics. This is related to a standard form of non-deductive reasoning, called *abduction* or *inference to the best explanation* (IBE).[12] There is debate about the exact form of IBE-inferences, but it is widely accepted that IBE is very commonly used in science and philosophy.[13] The name "IBE" suggests that we have a given candidate pool of explanations and we use IBE to pick the "best" of the lot, but most proponents think that some condition of minimal goodness is also required in IBE-inferences. This constraint ensures that we can't use IBE-reasoning to justifiably accept an explanation unless that explanation is, in a vacuum, a decent enough explanation of the facts to be explained.

Explanations are judged according to various theoretical virtues: simplicity, non–ad hoc-ness, generality and explanatory scope, coherence with established facts, coherence with well established theories, fruitfulness, precision, intuitiveness, *et cetera*. There is no canonical list of theoretical virtues that is generally agreed to be complete, and comparisons on this score will involve a healthy helping of vagueness and context sensitivity. But there are still many cases where putative explanation *A* is so manifestly superior to putative explanation *B* that we can confidently accept *A* and reject *B*. In some situations one putative explanation is even manifestly superior to all other candidate explanations. I claim that we are in just such a situation with respect to conventionalism. Conventionalism, on nearly every front, is a better account of logical and mathematical facts than any of its philosophical competitors. Moreover, it is as good and as satisfying an explanation of these

facts as could reasonably be hoped for by philosophers. All of the various theoretical virtues listed above are satisfied by conventionalism to such a degree, and in such a way, that it is difficult to imagine another philosophical theory doing better. Let me briefly go through the listed virtues point-by-point, in light of the book's overall discussion, to illustrate this.

Conventionalism is unmatched in its *theoretical simplicity*. It approaches linguistic meaning with a minimum of theoretical posits by rooting meaning firmly in language use, appealing only to syntactic items that we are already independently committed to, along with simple and direct metasemantic principles.

Conventionalism is *not at all ad hoc*. This contrasts sharply with various forms of restricted inferentialism that make ad hoc restrictions on meaning-constituting rules, and are thereby forced into dealing with potential counterexamples with gerrymandering. Conventionalism's unrestricted approach doesn't require any special pleading in the face of supposed problem cases.

Conventionalism is *extremely general* and thus has unmatched *explanatory scope*. My picture of metasemantics applies to logic, arithmetic, set theory and the whole of mathematics. I also think that unrestricted inferentialism applies to language in general, beyond logical and mathematical discourse, though I admittedly haven't done much to establish that here. Even still, the general principles used here manage to explain both logic and mathematics uniformly and from an incredibly simple basis.

Conventionalism *perfectly coheres with a broadly naturalistic picture of the world*. In explaining logical and mathematical facts, some approaches are forced to posit spooky metaphysical objects as load-bearing features in our picture of the world. For conventionalists, logical and mathematical truth and knowledge are explained without any need for disreputable, anti-naturalist assumptions. This is particularly important, since a major goal of a philosophical theory of logic and mathematics, is (or should be) to show how these potentially puzzling subjects fit into the modern scientific picture of the world (1.IV).

Conventionalism is an *extremely fruitful theory*, naturally giving rise to various new ideas and approaches to puzzles in the general

philosophy of logic and mathematics. The basic idea is robust and flexible enough to give rise to important theories of tonk, rule-circularity, ontology, determinacy, mathematical paradoxes, and more. Whatever else may be, the fecundity of the theory is not in doubt.

Conventionalism is *extremely precise* in leading to *fine-grained* treatments of a wide variety of issues in the philosophy of logic and mathematics. I have waved my hands occasionally, as all philosophers do, but I have tried to do so only when gliding over irrelevant points. I've tried to find my way honestly over the most important obstacles, moving far beyond the stage of slogan and caricature on depressingly difficult issues like bad company, ontology, and determinacy.

Conventionalism provides an *intuitively appealing* picture of logical and mathematical facts. This is easy for philosophers to forget, but ask a non-philosopher about math and see what they say. More likely than not, you will get back some starkly proto-conventionalist material, about concepts and truth by definition. Also, unlike other philosophical theories, conventionalism treats logical and mathematical truth as being of a kind. This too, is an intuitively appealing result.

Taken together, these points show that conventionalism, as developed here, built upon a foundation of unrestricted metasemantic inferentialism, not only provides the best explanation currently available of logic and mathematics, but the explanation provided is so solid that *it is difficult to imagine a better one*. The overall discussion in this book provides a strong theoretical case that the theory I have offered is correct, I think. Obviously, many broadly similar theories could be offered, and various subtleties might be altered and wrinkles added, but in general, and perhaps unlike some of the canonical "conventionalists", I think that this book's conventionalist theory of logic and mathematics is *factually true*, *uniquely correct*, and that anyone who doesn't accept it *is making a philosophical mistake*.

At least, this is what I often think. But pride cometh before a fall.

III Philosophical Progress

Throughout the book I've considered and addressed a number of direct philosophical objections to inferentialism and conventionalism. But having just gone on record about the objective correctness of the theory, I now need to consider a couple of new objections. Not new first-order objections, but instead what we might call *meta*objections—objections that aren't based on the content of conventionalism, but rather on the broader context in which the debate over conventionalism is taking place. Two metaobjections will be considered, one in this section, and one in the next.

The first metaobjection is based on historical inertia. This inertia makes it difficult for philosophers to even consider taking conventionalism seriously. The principle form that it takes is as a concern about *philosophical progress*. Many philosophers think that conventionalism is one of the very few philosophical theories that has been decisively refuted. And since this theory was once dominant but has now been abandoned, it provides one of the central illustrations of the fact that, contrary to what many critics of philosophy have claimed, philosophy makes tangible progress over time.

This narrative of progress is one that many philosophers are quite invested in. Perhaps as a defense mechanism, these philosophers believe that just like science, though perhaps more slowly, philosophy has managed to add more and more to a secure store of knowledge through the ages, making real advances. The narrative has it that as part of this march of progress, flawed philosophical theories, like conventionalism, were rejected, just as flawed scientific theories, like phlogiston theory, were rejected. In fact, some may even think that nearly the entirety of philosophy's cache of knowledge is negative: the justified true belief theory of knowledge is false, Hilbert's program cannot succeed, conventionalism is false, *et cetera*.

Since conventionalism was very widely accepted in the relatively recent past but is now very widely rejected, even reopening the case of conventionalism *feels* like a major step backward to progress-obsessed philosophers. They think, and I agree, that top-notch

philosophy done today is done at a higher level of sophistication than previous philosophy, owing to greater background knowledge, conceptual sophistication, higher standards, and other factors. These higher standards and greater sophistication, it is thought, must have been used to find conventionalism wanting. So what appeared to be true to our less sophisticated philosophical ancestors (logical and mathematical conventionalism) can now be seen, from our superior epistemic situation, to be false. Thus runs the standard, anti-conventionalist narrative of progress. Perhaps because of an implicit acceptance of these points, there also seems to be a knee-jerk resistance to taking any defense of conventionalism seriously. Bring up conventionalism and watch your interlocutors suppress a shudder.

However prevalent such attitudes may be, it is important to note that they are mistaken. This is true *even if* it is accepted that philosophy has made progress. In fact, I personally think that philosophical progress has been made on many fronts. We stand now on the shoulders of giants. But this true claim about progress should not be confused with the false claim that philosophy has *invariably* made progress. Thinking like that is tantamount to engaging in Whig history. Global progress, on the whole, is compatible with many local fits and starts. Even “local” fits and starts that last for decades or centuries. Let us realize that the arc of philosophical history is long, but it bends toward truth.

A reexamination of conventionalism with sophisticated contemporary tools will certainly lead to further progress, in one direction or the other. To reject a view out of hand, rather than examining it and the arguments for it, is the least *philosophical* attitude that I can imagine. And wrapping oneself in the cloak of “progress” to justify such closed-mindedness is both theoretically question begging and morally contemptible. If the narrative of progress is used to shut down dissenting views before they are given a fair hearing, then actual progress will stop. Using the idea of progress to silence minority views is itself inimical to actual philosophical progress. If the dominant late nineteenth century idealists had been able to use claims of progress to shield their

views from criticism, we might have accepted a silly version of idealism unto the present day.[14]

Conventionalism can be reexamined, and even accepted (or re-accepted) by the philosophical world without undermining any plausible narrative of philosophical progress. If there is a case against my brand of conventionalism, then it must be made on its own terms and on the level. Unless of course, the second metaobjection is successful.

IV One Against Many

The second metaobjection is based on the fact that I am the lone contemporary defender of conventionalism in logic and mathematics. Surely, the metaobjection continues, this fact by itself means that we should think it very likely that I am wrong. Of course, I am not truly as isolated as I've just made it sound. In reality, the closet likely contains many conventionalists. More are emerging. And a number of other views—formalism, neo-logicism, neo-Fregeanism, procedural postulationalism, full-blooded Platonism, restricted inferentialism—agree with conventionalism about a great many things.[15] So I am not so isolated as it may seem, but let's leave that aside to develop the strongest metaobjection possible.

Related concerns have recently been much discussed in the epistemological literature, usually under the rubric of *peer disagreement*.[16] Our epistemic peers are those who are just as well placed as we are to make epistemic evaluations about a given subject matter. In the literature it is usually assumed in addition that peers have equally good evidence and evidence-processing skills in the relevant domains, but these requirements are probably too stringent to apply to any actual cases. At the very least, our (informal) epistemic peers with regard to some claim or subject matter stand in an epistemic position with respect to the claim or subject matter that is, globally, just as good as our own. According to this informal construal, there can be various fine-grained epistemic and evidential differences between epistemic peers.[17] The basic

idea is that our epistemic peers with respect to a claim or subject matter are, all things considered, just as likely as we are to be correct.

The central question about this that is examined in the philosophical literature concerns the epistemological relevance, if any, of disagreement between epistemic peers. Should the mere fact that an epistemic peer disagrees with you be evidence against your own opinion? Should you downgrade your confidence in a claim in light of peer disagreement? Some philosophers defend a *steadfast* position, according to which peer disagreement has little or no epistemic relevance.[18] Others defend a *conciliatory* position, according to which peer disagreement should lead to rational revisions in our credences.[19] There are also various subdivisions within each of these camps.

Obviously, the very idea of an epistemic peer, even in the undemanding sense, isn't very realistic.[20] So let's slacken the term even further, so that all philosophers working in the same area count as each other's epistemic peers. With this twice-weakened notion, we can say that I'm in the very worst situation that one can be in with respect to "peer" disagreement, for I am holding my opinion, alone, against *all* of my epistemic peers. I disagree with the philosophical community as a whole (keeping in mind the mild fiction we're operating in). Quite clearly, the situation of being the only person in your (epistemic) peer group maintaining that *p*, while your peers, one and all, maintain that not-*p* with a very high degree of confidence, should give anyone pause.

In this type of situation, for outsiders who have not evaluated any of the relevant first-order evidence for themselves it is usually rational to think that the consensus is correct (provided that a reasonably high opinion is held of the epistemic abilities of insiders). But this doesn't tell *me* what *I* should make of *my own* epistemic position with respect to my peers on the issue of conventionalism. We normally go along with a consensus because we think it is more likely to be correct than minority views, at least in established fields. But there have been many cases (in science, history, *et cetera*) where the consensus view has been wrong and an extreme minority view turned out to be correct. So it isn't obvious that insiders who

find themselves in an epistemic minority are rationally required to capitulate.

Should my own confidence in my conventionalist views be downgraded or erased on the basis of the sheer amount of disagreement I'm faced with? What about the virulence of the disagreement? Or the sheer amount of virulent disagreement? Downgrading is what conciliationist views advise, but as noted above, they are typically concerned with a more idealized and potentially more troubling type of "peer" disagreement. Conciliation in that sense doesn't entail conciliation in my actual epistemic position. Ultimately, I don't think conciliation is plausible in a case like mine.

When my philosophical peers disagree with me, they do so, one would hope, on the basis of a careful assessment of all of the relevant evidence and arguments, assessments which they can communicate to me. Given this, it would be silly to take the mere fact of their having made an assessment as epistemically relevant, *since I can evaluate and find wanting their assessment itself*.[21] Conciliation effectively counts the same evidence twice, once directly and once indirectly. Double counting evidence is epistemically improper. We can all agree that it would be silly for me to use my own opinion that *p* to bootstrap my way into a higher level of confidence that *p* (especially if the bootstrapping move can be iterated!), but counting the opinions of my peers as evidence isn't any better than the obviously problematic bootstrapping move.[22]

This is true at least in cases where we don't have reason to think our own assessments are compromised. The philosophical opinions I've expressed and defended here are very different than drunken claims made on a whim. They are instead the product of careful and extended contemplation. Partly for this reason, I'm not at all moved by the arguments of hardline conciliatory positions when applied to the case at hand.

In fact, I don't think that conciliationists can truly be serious if they mean their advice to apply to my case and similar cases. Conciliationism itself isn't close to universally accepted by epistemologists, so if disagreement makes it irrational for me to be a conventionalist, it will also make it irrational to be a conciliationist (though perhaps less irrational, due to the sheer numbers).[23]

Conciliationists could make completely ad hoc exceptions for their own views, but that would be silly. If their view suggests that I should give up conventionalism, then it should also suggest that they give up conciliationism (though perhaps not as strongly). Given this, some conciliationists might retreat to the weaker position of merely exploring and defending the view without truly endorsing it.[24] This is analogous to the retreat that some conventionalists blundered into, discussed and rejected in 14.I. In either case, we aren't left with a credible and sincere position claiming that it is irrational for me to be a conventionalist in my actual epistemic situation.

A possible response to this would be to admit that positions—including versions of conciliationism—can rationally be held in the face of widespread peer disagreement but to hold that, in the case of conventionalism, the extent and nature of the opposition is such that it is no longer rational to hold the view. It is of course true that the sheer numbers arrayed against me are a cause for concern. But it is not so clear that this should lead to a downgrade in confidence. I've already noted that the numbers matter, from the outside, but it is difficult to see why they should matter from the inside. Like any philosopher who is defending and developing a view, I think I have seen more and understood things better, including understanding the errors of my opponents together with the reasons for those errors.

Maybe some might interpret me here as claiming that, at least on the issue of conventionalism, *I have no epistemic peers*. Put that way, it sounds strikingly arrogant, but I think that's just an artifact of the verbal formulation being used at the moment. If a "peer" is defined as one whose disagreement about a claim should be evidential, for you, then charitable conciliationists are practically forced to view holders of unpopular positions—perhaps including themselves—as people who see themselves as having few epistemic peers with respect to the issue in question.

Stepping back: clearly it is unreasonable to think that endorsing an extremely unpopular view is, by virtue of its very unpopularity, *irrational*. This point is compatible with massive disagreement being epistemically relevant, in a weaker manner, provided of course that the proper downgrade in confidence can't be iterated so as to surpass any given limit. Many of the great advances in our

philosophical, mathematical, scientific, and historical understanding have come from people who argued against the conventional (so to speak) wisdom of the day. On the flipside, much crankish nonsense is produced by standing against the wisdom of the day. But the genius and crank are not distinguished from each other merely by their position with respect to their peers. Any view that would deem standing against the tide as, *thereby*, irrational, is itself irrational, epistemically speaking, and is an enemy of progress, practically speaking.

When defending an unpopular view you incur many serious epistemic obligations, but I have attempted to discharge these here. I've taken great care to honestly assess all of the relevant evidence, both old and new. A consensus against old and stale versions of conventionalism is *nothing to me*, for I too am part of that consensus, in a sense. Debate over the newly revitalized conventionalism only begins now. The debate cannot be ended before it begins by appeal to a non-existent consensus. And in the likely event that the consensus is unmoved by my case, having taken no notice of it, I will not be irrational in maintaining that my case *should* move philosophers and mathematicians. The debate over conventionalism, like all other serious debates, must proceed via discussion and evaluation of the relevant philosophical issues, rather than by counting heads or allowing the currents of sociological inertia to carry us along.

V Doubts

Neither of these metaobjections trouble me greatly, *theoretically*, at least. I don't think that revitalizing an old view involves rejecting philosophical progress, nor do I think that lone thinkers standing against the tide of opinion are, thereby, being irrational. But this is from the head; what about the heart? Even aside from any arguments, the overall situation I'm in makes me uneasy, and unease causes doubts.

Up to this point, this chapter has been an exercise in confidence. As an outsider, I've taken a position of strength. I've claimed that this

book's theory of logic and mathematics is correct and that virtually everyone else who has thought about these matters is wrong. Perhaps this confidence was required of me in order to see this project through to the end. Perhaps overall inquiry works best at finding truth when the individual inquirers are all *over*committed. In any case, nothing makes academics as suspicious as bold confidence.

I recognize this, of course. But I have aimed to state things boldly in the service of intellectual honesty and clarity of expression, not out of delusional grandiosity or thoughtlessness. And I am not just going through the motions either. I have written this chapter as part of a serious effort to answer my own very real doubts about my epistemic situation. Psychologically, at least, my success has been equivocal. These doubts are cracks in the metaphorical foundations. I have attempted to repair them here. At the moment they are repaired to my satisfaction, but I keep looking back at them, half expecting the cracks to have widened.

Many canonical philosophers have expressed extreme confidence, but a few of them also expressed doubts for all to see. In the conclusion to book I of the *Treatise*, Hume wrote memorably about the external doubts occasioned by his epistemic solitude, adding that "When I turn my eye inward, I find nothing but doubt and ignorance." Most of the time the arguments of this chapter convince me. There is a fact of the matter; this is it. But doubts linger, about both of these points. Doubts that are harder to shake at night, with head to pillow, than they are in the light of day, when the work is going well.

Most troublingly, when these doubts are answered, more insidious doubts tend to replace them. Doubts about substance. Doubts about value. Doubts about the point of philosophy, and about the worth of this work *even if everything I've said here is true*. So despite this chapter's theoretical pushback against anticipated metaobjections, I am beset by doubts. Doubts about my views, about my defense of them, about their status, and about the importance and worth of this book even if all of my other doubts are baseless.

I DREAMT MYSELF an architect, long ago, finishing a grand cathedral. Day and night, from week to month, year by year, I labor in solitude. Chiseling the smallest features on gargoyles and grotesques.

Features far too small to be noticed by human eyes looking up to the sky. I tell those who will listen that I labor only to honor the glory of God. I tell others, I tell myself. But I lie. I lie to others, I lie to myself. In my heart I desire the praise of men. In my heart I doubt God's glory. At last, it is finished. I feel joy and apprehension in equal measure. A group of worldly notables is to inspect the cathedral on the morrow. But in the night a fire starts. I know not how. Rivals? A flaw in my design? My craft? I do not know. The fire burns hot, engulfing my creation entirely. The flames devour me as well. I die, alone. Come the morning, the fire has been quelled, the cathedral destroyed, the gargoyles shattered. A crowd gathers as the notables inspect the rubble. The foremost among them casually remarks to his fellows that it was obviously the carelessness of the architect that caused the fire and that shoddy workmanship showed throughout the ruin. To this, the others murmur thoughtless agreement. Nobody spoke of the cathedral or its architect ever again. All was forgotten as they strode out through the ashes.

---

1 Reichenbach (1938), page 41.

2 Coffa (1991), page 318.

3 See Ebbs (2001) and (2011), Goldfarb (1995) and (1997), and Ricketts (1994).

4 Ricketts (1994), page 177.

5 Clark and Hale (1994), page 281. In a qualified sense, Putnam agrees with Ricketts that Carnap assumed no language-independent notion of "fact" and no substantive notion of "making true", but he persuasively argues that despite this Carnap was, pace Ricketts, both an empiricist and a conventionalist.

6 Quine (1960*a*), see also his (1936) and (1951*a*). For an attempt to argue both that Carnap didn't accept truth by convention and that Quine knew this, see Ebbs (2011).

7 At many points in Carnap (1963*a*), for example.

8 Carnap (1937), page 161.

9 See the discussion in chapter 17 of Coffa (1991).

10 At the close of my (2016*a*), which deals with this general issue.

11 For influential writings in this tradition, see Kuhn (1962), Lakatos (1970), and Feyerabend (1975). For a diagnosis of the common error made by these philosophers, see Stove (1982) and (2004). Putnam (1974), ostensibly a

reaction to Popper's work, such as his (1934), contains an attractive general account of the different forms that scientific questions and disputes have. Putnam's approach, at a higher level of generality, might map over to much of philosophy as well.

12 See Harman (1965) for the introduction of this terminology and Lipton (2004) for extensive discussion.

13 Here I take no stand on the exact logical form of IBE-inferences. The points I am making will stand on any plausible approach.

14 I have in mind here the prominence of so-called *British* idealism in the late 19th century. Canonical works in this tradition include Bradley (1893) and McTaggart (1908).

15 For example see Balaguer (1995), Fine (2005), Hale (1987), Rayo (2013), Weir (2010), and Wright (1983) for views that have many important things in common with my approach to mathematics.

16 For overviews see Christensen (2009) and Matheson (2015).

17 Matheson (2015) calls something like this the "broad" notion of an epistemic peer, in contrast to the "narrow" notion previously canvased.

18 See Enoch (2010), Kelly (2005), and Wedgwood (2007).

19 See Christensen (2007), Elga (2007), and Kelly (2010).

20 A similar thought is pursued in King (2011).

21 This point is somewhat related to the view of Bergmann (2009).

22 Arguments like this are discussed as problems for conciliationism by Kelly (2010).

23 See Elga (2010).

24 Jon Matheson suggested this possibility to me in conversation; see also his (2014).

Bibliography

Achinstein, Peter. (1983). *The Nature of Explanation*. Oxford: Oxford University Press.

Anderson, A. and Belnap, N. (1975). *Entailment: The Logic of Relevance and Necessity: Volume 1.* Princeton: Princeton University Press.

Antonelli, Aldo G. (2013). "On the General Interpretation of First-Order Quantifiers." *Review of Symbolic Logic* 6(4): 637–658.

Armstrong, D.M. (1968). *A Materialistic Theory of the Mind.* London: RKP.

Audi, Robert. (1972). "The Concept of Believing." *Personalist* 53: 43–62.

Audi, Robert. (1994). "Dispositional Beliefs and Dispositions to Believe." *Noûs* 28: 419–434.

Awodey, Steve. (2007). "Carnap's Quest for Analyticity: The *Studies in Semantics*." In Michael Friedman and Richard Creath (eds.), *The Cambridge Companion to Carnap*. Cambridge: Cambridge University Press: 226–247.

Awodey, Steve. (2010). *Category Theory*. Second Edition. Oxford: Oxford University Press.

Awodey, Steve and A.W. Carus. (2009). "From Wittgenstein's Prison to the Boundless Ocean: Carnap's Dream of Logical Syntax." In Pierre Wagner (ed.), *Carnap's Logical Syntax of Language*. New York: Palgrave Macmillan: 79–106.

Ayer, Alfred Jules. (1936). "Truth by Convention: A Symposium." *Analysis* 4(2/3): 17–22.

Ayer, Alfred Jules. (1946). *Language, Truth and Logic*. Second edition. London: Victor Gollancz Ltd.

Ayer, Alfred Jules (ed.). (1959). *Logical Positivism.* New York: The Free Press.

Ayer, Alfred Jules. (1982). *Philosophy in the Twentieth Century.* New York: Random House.

Ayer, Alfred Jules. (1992). "Reply to Franciso Miró Quesada." In Lewis Edwin Hahn (ed.), *The Philosophy of A.J. Ayer*. La Salle: Open Court: 478–488.

Ayer, Alfred Jules and Frederick Copleston. (1957). "Logical Positivism—A Debate." In Paul Edwards and Arthur Pap (eds.), *A Modern Introduction to Philosophy: Readings from Classical and Contemporary Sources*. New York: The Free Press of Glencoe, Inc.:586–618.

Azzouni, Jody. (2004). *Deflating Existential Consequence: A Case for Nominalism*. New York: Oxford University Press.

Baker, G.P. and P.M.S. Hacker. (2009). *Wittgenstein: Rules, Grammar, and Necessity—Volume 2 of An Analytical Commentary on the Philosophical*

Investigations: Essays and Exegesis of §§185–242. Second Edition. West Sussex: Wiley-Blackwell.
Balaguer, Mark. (1995). “A Platonist Epistemology.” *Synthese* 103: 303–325.
Balaguer, Mark. (1998). *Platonism and Anti-Platonism in Mathematics.* New York: Oxford University Press.
Bar-Hillel, Yehoshua. (1954). “Logical Syntax and Semantics.” *Language* 30: 230–237.
Beall, J.C. and Greg Restall. (2006). *Logical Pluralism.* New York: Oxford University Press.
Belnap, Nuel. (1962). “Tonk, Plonk and Plink.” *Analysis* 22(6): 130–134.
Bell, John and Moshé Machover. (1977). *A Course in Mathematical Logic*. Amsterdam: Elsevier Science, B.V.
Benacerraf, Paul. (1962). “Tasks, Super-Tasks, and the Modern Eleatics.” *Journal of Philosophy* 59: 765–784.
Benacerraf, Paul. (1965). “What Numbers Could Not Be.” *Philosophical Review* 74(1): 47–73.
Benacerraf, Paul. (1973). “Mathematical Truth.” *Journal of Philosophy* 70: 661–680.
Benacerraf, Paul. (1981). “Frege: The Last Logicist.” *Midwest Studies in Philosophy* 6(1): 17–36.
Benacerraf, Paul. (1996). “Recantation *or* Any Old *ω*-sequence Would Do After All.” *Philosophia Mathematica* 4(2): 184–189.
Benacerraf, Paul. (1998). “What Mathematical Truth Could Not Be—I.” In Matthias Schirn (ed.), *The Philosophy of Mathematics Today*. Oxford: Clarendon Press: 33–76.
Benacerraf, Paul and Putnam, Hilary (eds.). (1964). *Philosophy of Mathematics: Selected Readings*. Englewood Cliffs: Prentice-Hall.
Benacerraf, Paul and Hilary Putnam. (1983). “Introduction.” in Paul Benacerraf and Hilary Putnam (eds.), *Philosophy of Mathematics: Selected Readings (2nd edition)*. Cambridge: Cambridge University Press.
Ben-Menahem, Yemima. (2006). *Conventionalism.* Cambridge: Cambridge University Press.
Bendall, Kent. (1979). “Negation as a Sign of Negative Judgement.” *Notre Dame Journal of Formal Logic* 20(1): 68–76.
Bennett, Jonathan. (1961). “On Being Forced to a Conclusion.” *Proceedings of the Aristotelian Society* suppl. vol. 35: 15–34.
Bergmann, Michael. (2009). “Rational Disagreement after Full Disclosure.” *Episteme: A Journal of Social Philosophy* 6(3): 336–353.
Berko, J. (1958). “The Child’s Learning of English Morphology.” *Word* 14: 150–177.
Bermudez, Jose Luis. (1995). “Syntax, Semantics and Levels of Explanation.” *Philosophical Quarterly* 45(180): 361–367.
Berry, Sharon. (2014). “Malament–Hogarth Machines and Tait’s Axiomatic Conception of Mathematics.” *Erkenntnis* 79(4): 893–907.

Berry, Sharon. (2015). “Chalmers, Quantifier Variance, and Mathematicians’ Freedom.” In Alessandro Torza (ed.), *Quantifiers, Quantifiers, and Quantifiers: Themes in Logic, Metaphysics, and Language*. Dordrecht: Synthese Library Volume 373: Springer: 191–219.

Berry, Sharon. (2017). “Physical/Metaphysical Possibility and Determinate Number Theory.”

Besson, Corine. (2012). “Logical Knowledge and Ordinary Reasoning.” *Philosophical Studies* 158: 59–82.

Beth, E.W. (1963). “Carnap’s Views on the Advantages of Constructed Systems over Natural Languages in the Philosophy of Science.” In P. Schilpp (ed.), *The Philosophy of Rudolf Carnap*. La Salle: Open Court: 469–502.

Birch, Jonathan. (2014). “Propositional Content in Signalling Systems.” *Philosophical Studies* 171(3): 493–512.

Black, Max. (1936). “Truth by Convention: A Symposium” *Analysis* 4(2/3): 28–32.

Black, Max. (1954). “Inductive Support of Inductive Rules.” In *Problems of Analysis*. Ithaca: Cornell University Press.

Black, Max. (1958). “Notes on the Meaning of ‘Rule’.” *Theoria* 24: 107–136, 139–161.

Blackburn, Simon. (1984*a*). *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Clarendon Press.

Blackburn, Simon. (1984*b*). “The Individual Strikes Back.” *Synthese* 58: 281–302.

Blackburn, Simon. (1987). “Morals and Modals.” In Graham Macdonald and Crispin Wright (eds.), *Face, Science, and Morality* Oxford: Blackwell.

Blanchette, Patricia A. (2001). “Logical Consequence.” In Lou Goble (ed.), *The Blackwell Guide to Philosophical Logic*. Oxford: Blackwell: 115–135.

Block, Ned. (1978). “Troubles with Functionalism.” *Minnesota Studies in the Philosophy of Science* 9: 261–325.

Block, Ned. (1986). “Advertisement for a Semantics for Psychology.” In Peter French, Theodore Uehling Jr., and Howard Wettstein. (eds.), *Midwest Studies in Philosophy* 10. University of Minnesota Press: Minneapolis.

Block, Ned. (1997). “Anti-Reductionism Slaps Back.” *Philosophical Perspectives 11: Mind, Causation, and World*: 107–132.

Boghossian, Paul. (1989). “The Rule-Following Considerations.” *Mind* 98(392): 507–549.

Boghossian, Paul. (1990). “Naturalizing Content.” In Barry Loewer and Georges Rey (eds.), *Meaning in Mind: Fodor and His Critics.* Oxford: Blackwell: 65–86.

Boghossian, Paul. (1996). “Analyticity Reconsidered.” *Noûs* 30: 360–391.

Boghossian, Paul. (2000). “Knowledge of Logic.” In Paul Boghossian and Christopher Peacocke (eds.), *New Essays on the Apriori.* Oxford: Oxford University Press: 229–254.

Boghossian, Paul. (2001). “How Are Objective Epistemic Reasons Possible?” *Philosophical Studies* 106: 1–40.

Boghossian, Paul. (2003). “Blind Reasoning.” *Proceedings of the Aristotelian Society, Supplementary* 77: 225–248.
Boghossian, Paul. (2006). *Fear of Knowledge: Against Relativism and Constructivism.* Oxford: Clarendon Press.
Boghossian, Paul. (2008). *Content and Justification: Philosophical Papers.* Oxford: Oxford University Press.
Boghossian, Paul. (2014). “What Is Inference?” *Philosophical Studies* 169(1): 1–18.
Bonevac, Daniel A. (1982). *Reduction in the Abstract Sciences*. Indianapolis: Hackett.
BonJour, Laurence. (1998). *In Defense of Pure Reason*. Cambridge: Cambridge University Press.
Bonnay, Denis. (2009). “Carnap’s Criterion of Logicality.” In Pierre Wagner (ed.), *Carnap’s Logical Syntax of Language*. New York: Palgrave Macmillan: 147–164.
Boolos, George. (1971). “The Iterative Conception of Set.” *Journal of Philosophy* 68(8): 215–231.
Boolos, George. (1984). “To Be Is to Be a Value of a Variable (or to Be Some Values of Some Variables).” *Journal of Philosophy* 81: 430–449.
Boolos, George. (1985). “Nominalist Platonism.” *Philosophical Review* 94: 327–344.
Boolos, George. (2000). “Must We Believe in Set Theory?” In Gila Sher and Richard Tieszen (eds.), *Between Logic and Intuition.* Cambridge: Cambridge University Press: 257–268.
Boolos, George and Richard C. Jeffrey. (1989). *Computability and Logic. Third edition*. Cambridge: Cambridge University Press.
Bostock, David. (1997). *Intermediate Logic*. Oxford: Clarendon Press.
Bourget, David and David J. Chalmers. (2014). “What Do Philosophers Believe?” *Philosophical Studies* 170(3): 465–500.
Bowdery, George. (1941). “Conventions and Norms.” *Philosophy of Science* 8: 493–505.
Bradley, F.H. (1893). *Appearance and Reality*. London: S. Sonnenschein.
Brandom, Robert. (1994). *Making It Explicit.* Cambridge: Harvard University Press.
Brandom, Robert. (2000). *Articulating Reasons: An Introduction to Inferentialism.* Cambridge: Harvard University Press.
Britton, Karl. (1947). “Are Necessary Truths True by Convention?” *Proceedings of the Aristotelian Society, Supplementary Volumes 21: Explanation in History and Philosophy*: 78–133.
Broome, John. (2014*a*). “Comments on Boghossian.” *Philosophical Studies* 169(1): 19–25.
Broome, John. (2014*b*). “Normativity in Reasoning.” *Pacific Philosophical Quarterly* 95(4): 622–633.
Buldt, Bernd. (2004). “On RC 102-43-14.” In Steve Awodey and Carsten Klein (eds). *Carnap Brought Home: The View from Jena.* Chicago and LaSalle: Open

Court: 225–246.
Burge, Tyler. (1979). “Individualism and the Mental.” *Midwest Studies in Philosophy* 4: 73–121.
Burge, Tyler. (1993). “Content Preservation.” *Philosophical Review* 102(4): 457–488.
Burgess, Alexis G. and John P. Burgess. (2011). *Truth*. Princeton: Princeton University Press.
Burgess, John P. (1981). “Relevance: A Fallacy?” *Notre Dame Journal of Formal Logic* 22: 76–84.
Burgess, John P. (2003). “Numbers and Ideas.” *Richmond Journal of Philosophy* 1: 12–17.
Burgess, John P. (2004). “Quine, Analyticity and Philosophy of Mathematics.” *Philosophical Quarterly* 54(214): 38–55.
Burgess, John P. (2005*a*). “Translating Names.” *Analysis* 65: 96–204.
Burgess, John P. (2005*b*). “On Anti-Anti-realism.” *Facta Philosophica* 7(2): 145–165.
Burgess, John P. (2013). *Kripke*. Cambridge: Polity Press.
Button, Tim and Sean Walsh. (2015) “Ideas and Results in Model Theory: Reference, Realism, Structure and Categoricity.” arXiv:1501.00472, 2015.
Byrne, R. M. J. (1989). “Suppressing Valid Inferences with Conditionals.” *Cognition* 31: 1–21.
Cao, Rosa, and Jared Warren. (2019). “Neural Decoding, the Atlantis Machine, and Zombies.”
Carnap, Rudolf. (1964). “The Logicist Foundations of Mathematics.” In Paul Benacerraf and Hilary Putnam Hilary (eds.), *Philosophy of Mathematics: Selected Readings*. Englewood Cliffs: Prentice-Hall. (Original work published 1934.) [94] Carnap, Rudolf. (1934). *Logische Syntax der Sprache.* [The Logical Syntax of Language] Vienna: Springer.
Carnap, Rudolf. (1937). *The Logical Syntax of Language.* Routledge and Kegan Paul: London.
Carnap, Rudolf. (1939). “Foundations of Logic and Mathematics.” In *International Encyclopedia of Unified Science* I(3). Chicago: University of Chicago Press.
Carnap, Rudolf. (1940). “On the Term ‘Analytic’ in Logical Syntax by S. C. Kleene.” *Journal of Symbolic Logic* 5(4): 157–158.
Carnap, Rudolf. (1942). *Introduction to Semantics.* Cambridge: Harvard University Press.
Carnap, Rudolf. (1943). *Formalization of Logic.* Cambridge: Harvard University Press.
Carnap, Rudolf. (1947). *Meaning and Necessity: A Study in Semantics and Modal Logic.* Chicago: University of Chicago Press.
Carnap, Rudolf. (1950). “Empiricism, Semantics, and Ontology.” *Revue Internaitonale de Philosophie* 4: 20–40.
Carnap, Rudolf. (1952). “Meaning Postulates.” *Philosophical Studies* 3: 65–73.

Carnap, Rudolf. (1955). “Meaning and Synonymy in Natural Language.” *Philosophical Studies* 7: 33–47.
Carnap, Rudolf. (1958). “Beobachtungssprache und theoretische Sprache.” [Observational Language and Theoretical Language] *Dialectica* 12: 236–248.
Carnap, Rudolf. (1963*a*). “Intellectual Autobiography.” In Paul Arthur Schilpp (ed.), *The Philosophy of Rudolf Carnap.* La Salle: Open Court: 3–84.
Carnap, Rudolf. (1963*b*). “Replies and Systematic Expositions.” In Schilpp (ed.), *The Philosophy of Rudolf Carnap.* La Salle: Open Court: 859–1013.
Carroll, Lewis. (1895). “What the Tortoise Said to Achilles.” *Mind* 4: 278–280.
Chalmers, David J. (1995). “Explaining Consciousness: The Hard Problem.” *Journal of Consciousness Studies* 2(3): 200–219.
Chalmers, David J. (1996). *The Conscious Mind: In Search of a Fundamental Theory.* Oxford: Oxford University Press.
Chalmers, David J. (2002). “Does Conceivability Entail Possibility?” In Tamar Gendler and John Hawthorne (eds.), *Conceivability and Possibility*. Oxford: Oxford University Press: 145–200.
Chalmers, David J. (2012). *Constructing the World.* Oxford: Oxford University Press.
Chalmers, David J., David Manley, and Ryan Wasserman (eds.). (2009). *Metametaphysics: New Essays on the Foundations of Ontology.* Oxford: Oxford University Press.
Charlesworth, M.J. (trans.) (1965). *St. Anselm’s Proslogion.* Oxford: Oxford University Press.
Chihara, Charles. (1977). “Wittgenstein’s Analysis of the Paradoxes in His Lectures on the Foundations of Mathematics.” *Philosophical Review* 86(3): 365–381.
Chihara, Charles. (1979). “The Semantic Paradoxes: A Diagnostic Investigation.” *Philosophical Review* 88: 590–618.
Chomsky, Noam. (1955). “Logical Syntax and Semantics: Their Linguistic Relevance.” *Language* 31(1): 36–45.
Chomsky, Noam. (1957). *Syntactic Structures.* The Hague: Mouton.
Chomsky, Noam. (1959). “A Review of B.F. Skinner’s *Verbal Behavior*.” *Language* 35(1): 26–58.
Chomsky, Noam. (1965). *Aspects of the Theory of Syntax.* Cambridge: The MIT Press.
Chomsky, Noam and Morris Halle. (1968). *The Sound Pattern of English.* New York: Harper and Row.
Chow, Timothy Y. (2019). “The Consistency of Arithmetic.” *Mathematical Intelligencer* 41(1): 22–30.
Christensen, David. (2007). “Epistemology of Disagreement: The Good News.” *Philosophical Review* 116(2): 187–218.
Christensen, David. (2009). “Disagreement as Evidence: The Epistemology of Controversy.” *Philosophy Compass* 4(5): 756–767.

Church, Alonzo. (1932). “A Set of Postulates for the Foundation of Logic.” *Annals of Mathematics. Series* 2. 33(2): 346–366.
Clark, Peter and Bob Hale (eds.). (1994). *Reading Putnam.* Oxford: Blackwell.
Coffa, J. Alberto. (1986). “From Geometry to Tolerance: Sources of Conventionalism in Ninteenth-Century Geometry.” In Robert G. Colodny (ed.), *From Quarks to Quasars: Philosophical Problems of Modern Physics* Pittsburgh: University of Pittsburgh Press.
Coffa, J. Alberto. (1987). “Carnap, Tarski, and the Search for Truth.” *Noûs* 21(4): 547–572.
Coffa, J. Alberto. (1991). *The Semantic Tradition from Kant to Carnap: To the Vienna Station.* New York: Cambridge University Press.
Colyvan, Mark. (2006). “Naturalism and the Paradox of Revisability.” *Pacific Philosophical Quarterly* 87(1): 1–11.
Cook, Roy. (2005). “What’s Wrong With Tonk(?)” *Journal of Philosophical Logic* 34: 217–226.
Cosmides, Leda and John Tooby. (1992). “Cognitive adaptations for social exchange.” In J. Barkow, L. Cosmides, and J. Tooby (Eds.) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press: 163–228.
Cottingham, John, Stoothoff, Robert, and Murdoch, Dugald (eds.). (1984). *The Philosophical Writings of Descartes: Volume II.* Cambridge: Cambridge University Press.
Curry, Haskell. (1963). *Foundations of Mathematical Logic*. New York: McGraw-Hill.
Daly, C. (2013). *Philosophy of Language: An Introduction.* London: Bloomsbury Academic.
Davidson, Donald. (1967). “Truth and Meaning.” *Synthese* 17: 304–323.
Davidson, Donald. (1974). “On the Very Idea of a Conceptual Scheme.” *Proceedings and Addresses of the American Philosophical Association* 47: 5–20.
Davidson, Donald. (1984). *Inquiries into Truth and Interpretation.* Oxford: Clarendon.
Davies, E.B. (2001). “Building Infinite Machines.” *British Journal for the Philosophy of Science* 52(4): 671–682.
Davies, Martin and Lloyd Humberstone. (1980). “Two Notions of Necessity.” *Philosophical Studies* 38: 1–30.
Davies, P.S., J.H. Fetzer, and T.R. Foster. (1995). “Logical Reasoning and Domain Specificity.” *Biology and Philosophy* 10(1): 1–37.
Davis, Philip J. and Reuben Hersh. (1981). *The Mathematical Experience*. Cambridge: Birkhauser.
Dean, Walter. (2014). “Montague’s Paradox, Informal Provability, and Explicit Modal Logic.” *Notre Dame Journal of Formal Logic* 55(2): 157–196.

de Bouvère, K. (1965). “Synonymous theories.” In J. W. Addison, L. Henkin, and A. Tarski (eds.), *The Theory of Models: Proceedings of the 1963 International Symposium at Berkeley*. Amsterdam: North Holland: 402–406.
Dedekind, Richard. (1888). *Was sind und was sollen die Zahlen?* [What are Numbers and What are they Good For?] Braunschweig: Vieweg.
Dennett, Daniel C. (1987). *The Intentional Stance.* Cambridge: The MIT Press.
Densmore, Dana (ed.). (2002). *Euclid’s Elements*. Santa Fe: Green Lion Press.
Detlefsen, Michael. (1986). *Hilbert’s Program*. Dordrecht: Reidel.
de Rouilhan, Philippe. (2009). “Carnap on Logical Consequence for Languages I and II.” In P. Wagner (ed.), *Carnap’s Logical Syntax of Language*. Palgrave Macmillan: 121–146.
Diamond, Cora (ed.). (1976). *Wittgenstein’s Lectures on the Foundations of Mathematics: Cambridge 1939.* Chicago: The University of Chicago Press.
Dogramaci, Sinan. (2010). “Knowledge of Validity.” *Noûs* 44(3): 403–432.
Dorr, Cian. (2005). “What We Disagree about When We Disagree about Ontology”, In M. Kalderon (ed.), *Fictionalism in Metaphysics*. Oxford: Oxford University Press: 234–286.
Dorr, Cian. (2014). “Quantifier Variance and the Collapse Theorems.” *The Monist* 97: 503–570.
Došen, Kosta. (1989). “Logical Constants as Punctuation Marks.” *Notre Dame Journal of Formal Logic* 30(3): 363–381.
Dreier, James. (2004). “Meta-Ethics and the Problem of Creeping Minimalism.” *Philosophical Perspectives* 18(1): 23–44.
Dummett, Michael. (1956). “Nominalism.” *Philosophical Review* 65: 491–501.
Dummett, Michael. (1959). “Wittgenstein’s Philosophy of Mathematics.” *Philosophical Review* 68: 324–348.
Dummett, Michael. (1963). “The Philosophical Significance of Gödel’s Theorem.” *Ratio* v: 140–155.
Dummett, Michael. (1973*a*). “The Philosophical Basis of Intuitionistic Logic.” Reprinted in Dummett (1978).
Dummett, Michael. (1973*b*). “The Justification of Deduction.” British Academy Lecture. Reprinted in Dummett (1978).
Dummett, Michael. (1973*c*). *Frege: Philosophy of Language*. London: Duckworth.
Dummett, Michael. (1975). “Wang’s Paradox.” *Synthese* 30: 301–324.
Dummett, Michael. (1978). *Truth and Other Enigmas.* London: Duckworth.
Dummett, Michael. (1991*a*). *The Logical Basis of Metaphysics*. Cambridge: Harvard University Press.
Dummett, Michael. (1991*b*). *Frege: Philosophy of Mathematics*. Cambridge: Harvard University Press.
Dummit, D.S. and R.M. Foote. (1991). *Abstract Algebra*. Upper Saddle River: Prentice Hall.
Earman, John and John D. Norton. (1993). “Forever Is a Day: Supertasks in Pitowsky and Malament-Hogarth Spacetimes.” *Philosophy of Science* 60(1): 22–

42.
Earman, John and John D. Norton. (1996). “Infinite Pains: The Trouble with Supertasks.” In Adam Morton and Stephen P. Stich (eds.), *Benacerraf and His Critics*. Oxford: Blackwell: 231–259.
Ebbs, Gary. (2001). “Carnap’s Logical Syntax.” In Richard Gaskin (ed.), *Grammar in Early Twentieth-Century Philosophy*. London: Routledge: 218–237.
Ebbs, Gary. (2011). “Carnap and Quine on Truth by Convention.” *Mind* 120 (478): 193–237.
Edgington, Dorothy. (1997). “Vagueness by Degrees.” In Rosanna Keefe and Peter Smith (eds.), *Vagueness: A Reader.* Cambridge: The MIT Press: 294–316.
Einheuser, Iris. (2003). *Conventionalism*. PhD diss., Massachusetts Institute of Technology.
Einheuser, Iris (2006). “Counterconventional Conditionals,” *Philosophical Studies* 127: 459–482.
Eklund, Matti. (2006). “Neo-Fregean Ontology.” *Philosophical Perspectives* 20(1): 95–121.
Eklund, Matti. (2009). “Carnap and Ontological Pluralism.” In D.J. Chalmers, D. Manley, and R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.
Elga, Adam. (2007). “Reflection and Disagreement.” *Noûs* 41(3): 478–502.
Elga, Adam. (2010). “How to Disagree About How to Disagree.” In R. Feldman and T. Warfield (eds.), *Disagreement*. New York: Oxford University Press: 175–186.
Enderton, Herbert. (2001). *A Mathematical Introduction to Logic*: Second edition. San Diego: Harcourt Academic Press.
Enoch, David. (2010). “Not Just a Truthometer: Taking Oneself Seriously (But Not Too Seriously) in Cases of Peer Disagreement.” *Mind* 119(476): 953–997.
Evans, Gareth. (1979). “Reference and Contingency.” *The Monist* 62(2): 178–213.
Ewing. A.C. (1940). “The Linguistic Theory of “*A priori*” Propositions.” *Proceedings of the Aristotelian Society* 40: 207–244.
Feferman, Solomon. (1960). “Arithmetization of Metamathematics in a General Setting.” *Fundamenta Mathematicae* 49: 35–92.
Feferman, Solomon. (1991). “Reflecting on Incompleteness.” *Journal of Symbolic Logic* 56: 1–49.
Feferman, Solomon. (2000). “Does Reductive Proof Theory Have a Viable Rationale?” *Erkenntnis* 53: 63–96.
Feferman, Solomon. (2007). “Responses to the 5 Questions.” In V.F. Hendricks and H. Leitgeb (eds.), *Philosophy of Mathematics: 5 Questions*. Copenhagen: Automatic Press.
Feferman, S., J.D. Dawson, W. Goldfarb, C. Parsons, and R. Solovay (eds.) (1995). *Kurt Gödel: Collected Works Volume III: Unpublished Essays and Lectures*. Oxford: Oxford University Press.
Felleisen, Matthias. (1991). “On the Expressive Power of Programming Languages.” *Science of Computer Programming* 17(1–3): 35–75.

Feyerabend, Paul. (1975). *Against Method.* New Left Books.
Field, Hartry. (1977). “Logic, Meaning, and Conceptual Role.” *Journal of Philosophy* 74: 379–409.
Field, Hartry. (1984*a*). “Is Mathematical Knowledge Just Logical Knowledge?” *Philosophical Review* 93(4): 509–552.
Field, Hartry. (1984*b*). “Platonism for Cheap? Crispin Wright on Frege’s Context Principle.” *Canadian Journal of Philosophy* 14:637–62.
Field, Hartry. (1984*c*). “Can We Dispensex with Spacetime?” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* (2): 33–90.
Field, Hartry. (1986). “The Deflationary Conception of Truth.” In Graham Macdonald and Crispin Wright (eds.), *Fact, Science, and Morality: Essays on A.J. Ayer’s* Language, Truth, and Logic. Oxford: Blackwell: 55–117.
Field, Hartry. (1994*a*). “Deflationist Views of Meaning and Content.” *Mind* 103: 249–284.
Field, Hartry. (1994*b*). “Are Our Mathematical and Logical Concepts Highly Indeterminate?” In P. French, T. Uehling, and H. Wettstein (eds.), *Midwest Studies in Philosophy* 19: 391–429.
Field, Hartry. (1998*a*). “Do We Have a Determinate Conception of Finiteness and Natural Number?” In Matthias Schirn (ed.), *The Philosophy of Mathematics Today.* Oxford: Oxford University Press: 99–130.
Field, Hartry. (1998*b*). “Which Undecidable Mathematical Sentences Have Determinate Truth Values?” In H.G. Dales and G. Olivari (eds.), *Truth in Mathematics.* Oxford: Oxford University Press: 291–310.
Field, Hartry. (2001). *Truth and the Absence of Fact.* Oxford: Oxford University Press.
Field, Hartry. (2005). “Recent Debates on the *A priori*.” In Tamar Gendler and John Hawthorne (eds.), *Oxford Studies in Epistemology: Volume 1*. Oxford University Press: 69–88.
Field, Hartry. (2009). “What Is the Normative Role of Logic?” *Proceedings of the Aristotelian Society Supplementary Volume* 83(1): 251–268.
Field, Hartry. (2017). “Egocentric Content.” *Noûs* 51(3): 521–546.
Fine, Kit. (1975). “Vagueness, Truth and Logic.” *Synthese* 30(3–4): 265–300.
Fine, Kit. (1994). “Essence and Modality.” *Philosophical Perspectives* 8: 1–16.
Fine, Kit. (2001). “The Question of Realism.” *Philosopher’s Imprint* 1: 1–30.
Fine, Kit. (2005). “Our Knowledge of Mathematical Objects.” In Tamar Gendler and John Hawthorne (eds.), *Oxford Studies in Epistemology: Volume 1.* Oxford: Oxford University Press: 89–110.
Fine, Kit. (2006). “Relatively Unrestricted Quantification.” In A. Rayo and G. Uzquiano (eds.), *Absolute Generality*. Oxford: Oxford University Press: 20–44.
Fine, Kit. (2009). “The Question of Ontology.” In D. Chalmers, D. Manley, and R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.

Fine, Kit. (2012). “Guide to Ground.” In Fabrice Correia and Benjamin Schnieder (eds.), *Metaphysical Grounding*. Cambridge: Cambridge University Press: 37–80.
Finn, Suki and Otávio Bueno. (2018). “Quantifier Variance Dissolved.” *Royal Institute of Philosophy Supplement* 82: 289–307.
Floyd, Juliet and Hilary Putnam. (2000). “A Note on Wittgenstein’s ‘Notorious Paragraph’ about the Godel Theorem.” *Journal of Philosophy* 97(11): 624–632.
Fodor, Jerry A. (1974). “The Disunity of Science as a Working Hypothesis.” *Synthese* 28(2): 97–115.
Fodor, Jerry A. (1978). “Propositional Attitudes.” *The Monist* 61: 573–591.
Fodor, Jerry A. (1990). *A Theory of Content and Other Essays.* Cambridge: The MIT Press.
Fodor, Jerry A. and Lepore, Ernest. (1992). *Holism: A Shopper’s Guide.* Oxford: Blackwell.
Fogelin, Robert. (1976). *Wittgenstein*. London and New York: Routledge.
Forbes, Graeme. (1984). “Scepticisim and Semantic Knowledge.” *Proceedings of the Aristotelian Society 1993–4*: 223–237.
Forster, Michael N. (2004). *Wittgenstein on the Arbitrariness of Grammar.* Princeton: Princeton University Press.
Franzén, Torkel. (2004). *Inexhaustibility: A Non-Exhaustive Treatment*. Wellesley: A K Peters, Ltd.
Frege, Gottlob. (1879). *Begriffsschrift: eine der arithmetischen nachgebildete Formelsprache des reinen Denkens.* [Concept Writing] Verlag: Halle.
Frege, Gottlob. (1884). *Die Grundlagen der Arithmetik: Eine logisch-mathematische Untersuchung uber deb Begriff der Zahl*. Breslau: Wilhelm Koebner. Trans. J.L. Austin. *Foundations of Arithmetic*. (1950). Oxford: Blackwell.
Frege, Gottlob. (1892). “Uber Sinn und Bedeutung.” [On Sense and Reference] *Zeitschrift für Philosophie und Philosophische Kritik* 100: 25–50.
Frege, Gottlob. (1893). *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*. [Basic Laws of Arithmetic] Volume 1. Jena: Pohle.
Frege, Gottlob. (1903). *Grundgesetze der Arithmetik, begriffsschriftlich abgeleitet*. [Basic Laws of Arithmetic] Volume 2. Jena: Pohle.
Frege, Gottlob. (1960). “Negation.” In P. Geach and M. Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Oxford University Press.
French, Rohan. (2019). “Notational Variance and Its Variants.” *Topoi* 38(2): 321–331.
Friedman, Michael. (1974). “Explanation and Scientific Understanding.” *The Journal of Philosophy* 71: 1–19.
Friedman, Michael. (1997). “Carnap and Wittgenstein’s *Tractatus*.” In W. Tait (ed.), *Early Analytic Philosophy*. La Salle: Open Court. 19–36.

Friedman, Michael. (1999). *Reconsidering Logical Positivism.* Cambridge: Cambridge University Press.
Friedman, Michael. (2016). “From Intuition to Tolerance: The Development of Carnap’s Philosophy of Mathematics.” In C. Posey and O. Rechter (eds.), *Kant’s Philosophy of Mathematics, Volume II: Reception and Influence after Kant.* Cambridge: Cambridge University Press.
Garson, James W. (2013). *What Logics Mean: From Proof Theory to Model-Theoretic Semantics.* Cambridge: Cambridge University Press.
Gasking, D.A.T. (1940). “Mathematics and the World.” *Australiasian Journal of Psychology and Philosophy* 18(2): 97–116.
Geach, Peter. (1962). *Reference and Generality*. Ithaca: Cornell University Press.
Geach, Peter. (1965). “Assertion.” *Philosophical Review* 74(4): 449–465.
Geach, Peter. (1975). “Names and Identity.” In Samuel Guttenplan (ed.), *Mind and Language*. Oxford: Clarendon Press: 139–158.
Geach Peter. (1981). “Wittgenstein’s Operator ‘N’.” *Analysis* 41.
Geach, Peter. (1982). “More on Wittgenstein’s Operator ‘N’.” *Analysis* 42.
Gelman, Susan A. (2003). *The Essential Child: Origins of Essentialism in Everyday Thought*. New York: Oxford University Press.
Gentzen, Gerhard. (1934). “Untersuchungen über das logische Schliessen.” [Studies on Logical Closure] *Mathematische Zeitschrift* 39: 176–210, 405–431.
Gentzen, Gerhard (1936). “Die Widerspruchsfreiheit der reinen Zahlentheorie.” [The Consistency of Pure Number Theory] *Mathematische Annalen* 112: 493–565.
Gettier, Edmund. (1963). “Is Justified True Belief Knowledge?” *Analysis* 23: 121–123.
Giannoni, Carlo. (1970). “A Note on the Impossibility Proof of the Conventionalist Thesis.” *Philosophical Studies* 21(4): 61–64.
Giannoni, Carlo. (1971). *Conventionalism in Logic: A Study in the Linguistic Foundation of Logical Reasoning.* Mouton: The Hague.
Gibbard, Allan. (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment.* Cambridge: Harvard University Press.
Ginsborg, Hannah. (2011). “Primitive Normativity and Skepticism about Rules.” *Journal of Philosophy* 108(5): 227–254.
Glivenko, V. (1929). “Sur quelques points de la logique de M. Brouwer.” [On Some Points of Mr. Brouwer’s Logic] *Bulletins de la Classe des Sciences* 15(5): 183–188.
Glock, Hans-Johann. (2003). “The Linguistic Doctrine Revisited.” *Grazer Philosophische Studien* 66: 143–170.
Glock, Hans-Johann. (2008). “Necessity and Language: In Defense of Conventionalism.” *Philosophical Investigations* 31(1): 24–47.
Glymour, Clark. (1970). “Theoretical Realism and Theoretical Equivalence.” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Association* 275–288.

Gödel, Kurt. (1929). *Über die Vollständigkeit des Logikkalküls*. [About the Completeness of the Logical Calculus] PhD thesis, University of Vienna.
Gödel, Kurt. (1931). "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I." [On Formally Undecidable Propositions of Principia Mathematic and Related Systems I] *Montashefte für Mathematik und Physik* 38: 173–198.
Gödel, Kurt. (1933). "Zur intuitionistischen Arithmetik und Zahlentheorie." [On Intuitionistic Arithmetic and Number Theory] *Ergebnisse eines mathematischen Kolloquiums* 4: 34–38.
Gödel, Kurt. (1953/9). "Is Mathematics Syntax of Language?" In S. Feferman, J.D. Dawson, W. Goldfarb, C. Parson, and R. Solovay (eds.), *Kurt Gödel: Collected Works, Volume III: Unpublished Essays and Lectures*. Oxford: Oxford University Press.
Gödel, Kurt. (1964/1983). "What Is Cantor's Continuum Problem?" (revised edition). In P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics: Selectd Readings*. Second edition. Cambridge University Press: 470–485.
Goldfarb, Warren. (1995). "Introductory Note to Gödel (1953/9)." In S. Feferman, J.D. Dawson, W. Goldfarb, C. Parson, and R. Solovay (eds.), *Kurt Gödel: Collected Works, Volume III: Unpublished Essays and Lectures.* Oxford: Oxford University Press: 324–333.
Goldfarb, Warren. (1997). "Semantics in Carnap: A Rejoinder to Alberto Coffa." *Philosophical Topics* 25: 51–66.
Goodman, Nelson and W.V. Quine. (1947). "Steps Toward a Constructive Nominalism." *Journal of Symbolic Logic* 12(4): 105–12.
Grandy, Richard. (1970). "Review of Pollock." *Journal of Symbolic Logic* 35(4): 576.
Grattan-Guinness, Ivor. (2008). "Solving Wigner's Mystery: The Reasonable (Though Perhaps Limited) Effectiveness of Mathematics in the Natural Sciences." *The Mathematical Intelligencer* 30(3): 7–17.
Grover, Dorothy, Joseph Camp, and Nuel Belnap. (1975). "A Prosentential Theory of Truth." *Philosophical Studies* 27: 73–125.
Guardo, Andrea. (2012). "Rule-Following, Ideal Conditions, and Finkish Dispositions." *Philosophical Studies* 157(2): 195–209.
Haack, Susan. (1974). *Deviant Logic, Fuzzy Logic: Beyond the Formalism.* Chicago: The University of Chicago Press.
Haack, Susan. (1976). "The Justification of Deduction." *Mind* 85: 112–119.
Hacking, Ian. (1979). "What Is Logic?" *Journal of Philosophy* 76(6): 285–319.
Hahn, Hans. (1933). "Logic, Mathematics, and Knowledge of Nature." In A.J. Ayer (ed.) *Logical Positivism*. New York: The Free Press: 147–164.
Hale, Bob. (1987). *Abstract Objects*. Oxford: Basil Blackwell.
Hale, Bob. (1994). "Singular Terms (2)." In Brian McGuiness and Gianluigi Oliveri (eds.), *The Philosophy of Michael Dummett*. Dordrecht: Kluwer Academic Publishers: 17–44.

Hale, Bob. (1996). “Singular Terms (1).” In Matthias Schirn (ed.), *Frege: Importance and Legacy*. New York: Walter de Gruyter: 438–457.
Hale, Bob. (2007). “Neo-Fregeanism and Quantifier Variance.” *Proceedings of the Aristotelian Society* 107(1.3): 375–385.
Hale, Bob and Crispin Wright (2000). “Implicit Definition and the *a priori*.” In Paul Boghossian and Christopher Peacocke (eds.), *New Essays on the a Priori*. Oxford: Oxford University Press 286–319.
Hale, Bob and Crispin Wright. (2001). *The Reason’s Proper Study*. Oxford: Oxford University Press.
Hale, Bob and Wright, Crispin. (2009). “The Metaontology of Abstraction.” In D.J. Chalmers, D. Manley, and R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press: 178–212.
Hamkins, Joel David. (2012). “The Set-Theoretic Multiverse.” *Review of Symbolic Logic* 3: 416–449.
Hamkins, Joel David and Ruizhi Yang. (2014). “Satisfaction In Not Absolute.” *Review of Symbolic Logic*: 1–34.
Hanson, William H. (1989). “Two Kinds of Deviance.” *History and Philosophy of Logic* 10(1): 15–28.
Hanson, William H. (1997). “The Concept of Logical Consequence.” *Philosophical Review* 106(3): 365–409.
Harman, Gilbert. (1965). “The Inference to the Best Explanation.” *Philosophical Review* 74(1): 88–95.
Harman, Gilbert. (1973). *Thought*. Princeton: Princeton University Press.
Harman, Gilbert. (1982). “Conceptual Role Semantics.” *Notre Dame Journal of Formal Logic* 23: 242–256.
Harman, Gilbert. (1986*a*). “The Meanings of Logical Constants.” In LePore (ed.), *Truth and Interpretation:Perspectives on the Philosophy of Donald Davidson*: New York: Blackwell: 125–134.
Harman, Gilbert. (1986*b*). *Change in View: Principles of Reasoning*. Cambridge: The MIT Press.
Harris, J. H. (1982). “What’s So Logical about the “Logical” Axioms’?” *Studia Logica* 41: 159–171.
Hauser, Kai and W. Hugh Woodin. (2014). “Strong Axioms of Infinity and the Debate about Realism.” *Journal of Philosophy* 111(8): 397–419.
Hawley, Katherine. (2007). “Neo-Fregeanism and Quantifier Variance.” *Aristotelian Society Supplementary Volume* 81(1): 233–249.
Hawthorne, John. (2006). “Plenitude, Convention, and Ontology.” In *Metaphysical Essays*. Oxford: Oxford University Press: 53–70.
Haze, Tristan. (2017). “On Warren’s ‘The Possibility of Truth by Convention’.” Blog post, May 2, 2017, http://sprachlogik.blogspot.com/2017/05/on-warrens-possibility-of- truth-by.html.
Hellman, Geoffrey. (1985). “Determination and Logical Truth.” *Journal of Philosophy* 82(11): 607–616.

Hellman, Geoffrey. (1986). “Logical Truth by Linguistic Convention.” In Lewis Hahn and Paul Arthur Schilpp (eds.), *The Philosophy of W.V. Quine*. La Salle: Open Court: 189–205.
Hellman, Geoffrey. (1989). *Mathematics Without Numbers.* Oxford: Oxford University Press.
Hempel, Carl G. (1945). “On the Nature of Mathematical Truth.” *American Mathematical Monthly* 52: 543–556.
Hempel, Carl G. (1965). *Aspects of Scientific Explanation*. New York: Free Press.
Hempel, Carl G. and Oppenheim, Paul. (1948). “Studies in the Logic of Explanation.” *Philosophy of Science* 15: 135–175.
Henkin, Leon. (1950). “Completeness in the Theory of Types.” *Journal of Symbolic Logic* 15(2): 81–91.
Hersh, Reuben. (1997). *What Is Mathematics, Really?* Oxford: Oxford University Press.
Hilbert, David. (1926). “Über das Unendliche.” [About the Infinite] *Mathematische Annalen* 95: 161–190.
Hilbert, David. (1931*a*). “Die Grundlegung der elementaren Zahlenlehre.” [The Foundation of Elementary Number Theory] *Mathematische Annalen* 104: 485–494.
Hilbert, David. (1931*b*). “Beweis des Tertium non datur.” [Proof of Excluded Middle] *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen aus dem Jahre 1931, Mathematische-Physikalische Klasse*: 120–125.
Hilbert, David and Paul Bernays. (1939). *Grundlagen der Mathmatik, Volume II.* [Foundations of Mathematics] Berlin: Springer.
Hirsch, Eli. (2002). “Quantifier Variance and Realism.” *Philosophical Issues* 12: 51–73.
Hirsch, Eli. (2008). “Language, Ontology, and Structure.” *Noûs* 42: 509–528.
Hirsch, Eli. (2011). *Quantifier Variance and Realism: Essays in Meta-ontology.* New York: Oxford University Press:.
Hirsch, Eli and Jared Warren. (2019). “Quantifier Variance and the Demand for a Semantics.” *Philosophy and Phenomenological Research* 98(3): 592–605.
Hirsch, Eli and Jared Warren. (forthcoming). “Quantifier Variance.” In *The Routledge Handbook of Philosophy of Relativism*.
Hofweber, Thomas. (2009). “Ambitious, Yet Modest, Metaphysics.” In D.J. Chalmers, D. Manley, and R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press: 260–289.
Hogarth, Mark. (2004). “Deciding Arithmetic Using SAD Computers.” *British Journal for the Philosophy of Science* 55(4): 681–691.
Horn, Laurence R. (1989). *A Natural History of Negation*. Chicago: University of Chicago Press.
Horwich, Paul. (1990). *Truth*. Oxford: Oxford University Press.
Horwich, Paul. (1998*a*). *Meaning*. Oxford: Oxford University Press.
Horwich, Paul. (1998*b*). *Truth*. 2nd edition. Oxford: Clarendon Press.

Horwich, Paul. (2005). *Reflections on Meaning*. Oxford: Clarendon Press.
Humberstone, Lloyd. (2000*a*). “The Revival of Rejective Negation.” *Journal of Philosophical Logic* 29: 331–381.
Humberstone, Lloyd. (2000*b*). “Contra-Classical Logics.” *Australasian Journal of Philosophy* 78: 438–474.
Hume, David. (1739). *A Treatise of Human Nature*. London: John Noon.
Hume, David. (1779). *Dialogues Concerning Natural Religion*. London.
Incurvati, Luca and Peter Smith. (2010). “Rejection and Valuations.” *Analysis* 70(1): 3–10.
Jackendoff, Ray. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
Jackson, Frank. (1998). *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. Oxford: Oxford University Press.
Jackson, Frank, Graham Oppy, and Michael Smith. (1994). “Minimalism and Truth Aptness.” *Mind* 103(411): 287–302.
Jaśkowski, S. (1934). “On the Rules of Supposition in Formal Logic” as reprinted in S. McCall (editor), *Polish Logic 1920–1939*. Oxford: Oxford University Press: 232–258.
Johnston, Mark. (1992). “How to Speak of the Colors.” *Philosophical Studies* 68: 221–263.
Juhl, Cory and Eric Loomis. (2010). *Analyticity.* New York: Routledge.
Kanamori, Akihiro. (2003). *The Higher Infinite: Large Cardinals in Set Theory from Their Beginnings.* New York: Springer-Verlag.
Kant, Immanuel. (1787/1965). *The Critique of Pure Reason.* Kemp Smith, Norman (trans.). New York: St. Martin’s Press.
Kaplan, David. (1989). “Demonstratives.” In Joseph Almog, John Perry, and Howard Wettstein (eds.), *Themes From Kaplan.* Oxford University Press: 481–564.
Kaye, Richard. (1991). *Models of Peano Arithmetic*. Oxford: Clarendon Press.
Kelly, Thomas. (2005). “The Epistemic Significance of Disagreement.” In Tamar Gendler and John Hawthorne (eds.), *Oxford Studies in Epistemology: Volume 1*. Oxford: Oxford University Press: 167–196.
Kelly, Thomas. (2010). “Peer Disagreement and Higher-Order Evidence.” In R. Feldman and T. Warfield (eds.), *Disagreement.* Oxford University Press: New York: 111–174.
King, Nathan L. (2011). “Disagreement: What’s the Problem? Or a Good Peer is Hard to Find.” *Philosophy and Phenomenological Research* 85(2): 249–272.
Kitcher, Philip. (1989). “Explanatory Unification and the Causal Structure of the World.” In P. Kitcher and W.P. Salmon (eds.), *Minnesota Studies in the Philosophy of Science* XIII. Minneapolis: Minnesota University Press: 71–91.
Kleene, Stephen. (1939). “The Logical Syntax of Language by Rudolf Carnap.” *Journal of Symbolic Logic* 4(2): 82–87.

Kleene, Stephen. (1952). *Introduction to Metamathematics.* Amsterdam: North-Holland.
Kleene, Stephen and J. Barkley Rosser. (1935). “The Inconsistency of Certain Formal Logics.” *The Annals of Mathematics* 36(3): 630.
Kneale, William. (1956). “The Province of Logic.” In H.D. Lewis (ed.), *Contemporary British Philosophy*. London: Allen and Unwin: 237–262.
Kocurek, Alexander. (2018). *What Can You Say? Measuring the Expressive Power of Languages*. PhD diss., University of California, Berkeley.
Koellner, Peter. (2009*a*). “Truth in Mathematics: The Question of Pluralism.” In *New Waves in Philosophy of Mathematics*. New York: Palgrave Macmillan: 80–116.
Koellner, Peter. (2009*b*). “Carnap on the Foundations of Logic and Mathematics.”
Kolenda, Konstantin. (1982). “Carlo Borromeo Giannoni 1939–1981.” *Proceedings and Addresses of the American Philosophical Association* 55(4): 497.
Kripke, Saul A. (1976). “Is There a Problem about Substitutional Quantification?” In Gareth Evans and John McDowell (eds.), *Truth and Meaning.* Oxford: Clarendon Press: 325–419.
Kripke, Saul A. (1979). “A Puzzle about Belief.” In A. Margalit (ed.), *Meaning and Use*. Dordrecht: Reidel: 239–283.
Kripke, Saul A. (1980). *Naming and Necessity.* Cambridge: Harvard University Press.
Kripke, Saul A. (1982). *Wittgenstein on Rules and Private Language.* Cambridge: Harvard University Press.
Kroedel, Thomas. (2012). “Implicit Definition and the Application of Logic.” *Philosophical Studies* 158: 131–148.
Kuhn, Thomas. (1962). *The Structure of Scientific Revolutions*. Chicago: The University of Chicago Press.
Kunen, Kenneth. (1971). “Elementary Embeddings and Infinitary Combinatorics.” *The Journal of Symbolic Logic* 36(3): 407–413.
Kunen, Kenneth. (1980). *Set Theory: An Introduction to Independence Proofs*. Amsterdam: Elsevier.
Kusch, Martin. (2006). *A Sceptical Guide to Meaning and Rules: Defending Kripke’s Wittgenstein.* Ithaca: McGill-Queen’s University Press.
Lakatos, Imre. (1970). “Falsification and the Methodology of Scientific Research Programs.” In I. Lakatos and A. Musgrave (eds.), *Criticism and the Growth of Knowledge.* Cambridge: Cambridge University Press: 91–196.
Lakatos, Imre. (1976). *Proofs and Refutations: The Logic of Mathematical Discovery*. New York: Cambridge University Press.
Lakoff, George and Raphael E. Núñez. (2000). *Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being*. New York: Basic Books.
Lavine, Shaughan. (1994). *Understanding the Infinite*. Cambridge: Harvard University Press.

Lavine, Shaughan. (2006). “Something about Everything: Universal Quantification in the Universal Sense of Universal Quantification.” In A. Rayo and G. Uzquiano (eds.), *Absolute Generality.* Oxford: Oxford University Press: 98–148.
Lavine, Shaughan. (2001). *Skolem Was Wrong.* Unpublished manuscript.
Leitgeb, Hannes. (2009). “On Formal and Informal Provability.” In Otávio Bueno and Øystein Linnebo (eds.), *New Waves in Philosophy of Mathematics*. New York: Palgrave-Macmillan: 263–299.
Lemmon, E.J. (1965). *Beginning Logic.* Cambridge: Hackett.
Levine, Joseph. (1983). “Materialism and Qualia: The Explanatory Gap.” *Pacific Philosophical Quarterly* 64: 354–361.
Lewis, C.I. (1946). *An Analysis of Knowledge and Valuation*. La Salle: Open Court.
Lewis, David. (1968). “Counterpart Theory and Quantified Modal Logic.” *Journal of Philosophy* 65: 113–126.
Lewis, David. (1969). *Convention: A Philosophical Study.* Cambridge: Harvard University Press.
Lewis, David. (1970). “How to Define Theoretical Terms.” *Journal of Philosophy* 67: 427–446.
Lewis, David. (1974). “Radical Interpretation.” *Synthese* 23: 331–344.
Lewis, David. (1980). “Mad Pain and Martian Pain.” In Ned Block (ed.), *Readings in the Philosophy of Psychology*. Cambridge: Harvard University Press: 216–222.
Lewis. David. (1984). “Putnam’s Paradox.” *Australasian Journal of Philosophy* 62: 221–236.
Lewis, David. (1991). *Parts of Classes.* Blackwell: Oxford.
Lewis, David. (1996). “Elusive Knowledge.” *Australasian Journal of Philosophy* 74: 549–567.
Lewis, David. (2001). “Forget about the ‘Correspondence Theory of Truth’.” *Analysis* 61(4): 280–289.
Lewy, Casimir. (1976). *Meaning and Modality.* Cambridge: Cambridge University Press.
Linaza, José. (1984). “Piaget’s Marbles: The Study of Children’s Games and Their Knowledge of Rules.” *Oxford Review of Education* 10(3): 271–274.
Lindström, Per. (1966). “First Order Predicate Logic with Generalized Quantifiers.” *Theoria* 32(3): 186–195.
Linnebo, Øystein. (2018). *Thin Objects: An Abstractionist Account.* Oxford: Oxford University Press.
Lipton, Peter. (2004). *Inference to the Best Explanation.* Second edition. New York: Routledge.
Löb, M.H. (1955). “Solution of a Problem of Leon Henkin.” *Journal of Symbolic Logic* 20: 115–118.
Loewer, Barry. (2009). “Why Is There Anything Except Physics?” *Synthese* 170(2): 217–233.
Lucas, J.R. (1961). “Minds, Machines and Gödel.” *Philosophy* 36: 112–127.

Łukasiewicz, J. (1957). *Aristotle's Syllogistic*. Oxford: Oxford University Press.
Mackie, J.L. (1977). *Ethics: Inventing Right and Wrong*. Penguin Books.
MacLane, Saunders. (1938). "Carnap on Logical Syntax." *Bulletin of the American Mathematical Society* 44(3): 171–176.
MacLane, Saunders. (1986). *Mathematics, Form and Function*. Springer-Verlag.
Malcolm, Norman. (1940). "Are Necessary Propositions Really Verbal?" *Mind* 49(194): 189–203.
Malmgren, Anna-Sara. (2018). "Varieties of Inference?" *Philosophical Issues: Philosophy of Logic and Inferential Reasoning* 28(1): 1–34.
Malmgren, Anna-Sara. (2019). "Availability, Goodness, and Argument Structure." Unpublished paper.
Marcus, Ruth B. (1990). "Some Revisionary Proposals about Belief and Believing." *Philosophy and Phenomenological Research* 50: 132–153.
Martin, C.B. (1994). "Dispositions and Conditionals." *Philosophical Quarterly* 44: 1–8.
Mates, Benson. (1965). *Elementary Logic.* New York: Oxford University Press.
Matheson, Jonathan. (2014). "Are Conciliatory Views of Disagreement Self-Defeating?" *Social Epistemology* 29(2): 145–159.
Matheson, Jonathan. (2015). "Disagreement and Epistemic Peers." *Oxford Handbooks Online*.
McCawley, James D. (1981). *Everything that Linguists Have Always Wanted to Know about Logic* (*but were ashamed to ask).* Chicago: University of Chicago Press.
McDaniel, Kris. (2010). "A Return to the Analogy of Being." *Philosophy and Phenomenological Research* 81(3): 688–717.
McDowell, John. (1997). "Another Pleas for Modesty." In R. Heck (ed.), *Language, Thought, and Logic: Essays in Honour of Michael Dummett.* Oxford: Clarendon Press: 105–130.
McGee, Vann. (1985). "A Counterexample to Modus Ponens." *Journal of Philosophy* 82(9): 462–471.
McGee, Vann. (1991). "We Turing Machines Aren't Expected-Utility Maximizers (Even Ideally)." *Philosophical Studies* 64(1): 115–123.
McGee, Vann. (1996). "Logical Operations." *Journal of Philosophical Logic* 25(6): 567–580.
McGee, Vann. (1997). "How We Learn Mathematical Language." *Philosophical Review* 106: 35–68.
McGee, Vann. (2000). "Everything." In G. Sher and R. Tieszen (eds.), *Between Logic and Intuition.* Cambridge: Cambridge University Press.
McGee, Vann. (2006). "There's a Rule for Everything." In A. Rayo and G. Uzquiano (eds.), *Absolute Generality.* Oxford: Oxford University Press.
McGinn, Colin. (1989). "Can We Solve the Mind-Body Problem?" *Mind* 98: 349–366.

McGinn, Colin. (2000). *Logical Properties: Identity, Existence, Predication, Necessity, Truth.* Oxford: Oxford University Press.
McTaggart, J.M.E. (1908). “The Unreality of Time.” *Mind* 17(68): 457–474.
Mehlberg, Henryk. (1958). *The Reach of Science*. Toronto: University of Toronto Press.
Merrill, G.H. (1980). “The Model-Theoretic Argument Against Realism.” *Philosophy of Science* 47: 69–81.
Midgley, G.C.J. (1959). “Linguistic Rules.” *Proceedings of the Aristotelian Society* 59(1): 271–290.
Miller, Alexander. (2003). “Does ‘Belief Holism’ Show that Reductive Dispositionalism about Content Could Not Be True?” *Supplement to the Proceedings of the Aristotelian Society* 77: 73–90.
Montague, Richard. (1963). “Syntactical Treatments of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability.” *Acta Philosophica Fennica* 16: 153–167.
Morris, Errol. (2018). *The Ashtray (Or the Man Who Denied Reality)*. Chicago: University of Chicago Press.
Morton, Adam. (1973). “Denying the Doctrine and Changing the Subject.” *Journal of Philosophy* 70(15): 503–510.
Moss, Sarah. (2012). “Solving the Color Incompatibility Problem.” *Journal of Philosophical Logic* 41(5): 841–851.
Mostowski, Andrzej. (1957). “On a Generalization of Quantifiers.” *Fundamenta Mathematicae* 44(2): 12–36.
Murzi, Julien and Hjortland, Ole T. (2009). “Inferentialism and the Categoricity Problem: Reply to Raatikainen.” *Analysis* 69(3): 480–488.
Murzi, Julien and Steinberger, Florian. (2013). “Is Logical Knowledge Dispositional?” *Philosophical Studies* 166: 165–183.
Myhill, John. (1960). “Some Remarks on the Notion of Proof.” *Journal of Philosophy* 57: 461–471.
Nagel, Ernest. (1944). “Logic Without Ontology.” In Y.H. Krikorian (ed.), *Naturalism and the Human Spirit*. New York: Columbia University Press: 210–241.
Nagel, Ernest. (1961). *The Structure of Science*. London: Routledge and Kegan Paul.
Nagel, Thomas. (1974). “What Is it Like to Be a Bat?” *Philosophical Review* 83(4): 435–450.
Nelson, Edward. (1986). *Predicative Arithmetic*. Princeton: Princeton University Press.
Neta, Ram. (2013). “What Is an Inference?” *Philosophical Issues* 23(1): 388–407.
Newman, M.H. (1928). “Mr. Russell’s ‘Causal Theory of Perception’” *Mind* 37(146): 137–148.
Oaksford, M. (2005). “Reasoning.” In N. Braisby and M. Gellatly (eds), *Cognitive Psychology*. Oxford: Oxford University Press: 418–455.

Odifreddi, Piergiorgio. (1989). *Classical Recursion Theory: The Theory of Functions and Sets of Natural Numbers*. Amsterdam: Elsevier.

Oppenheim, Paul and Hilary Putnam. (1958) “Unity of Science as a Working Hypothesis.” In H. Feigl, M. Scriven, and G. Maxwell (eds), *Minnesota Studies in the Philosophy of Science, Volume 2: Concepts, Theories, and the Mind–Body Problem*. Minneapolis: University of Minnesota Press: 3–36.

Paoli, Francesco. (2003). “Quine and Slater on Paraconsistency and Deviance.” *Journal of Philosophical Logic* 32: 531–548.

Pap, Arthur. (1958). *Semantics and Necessary Truth*. New Haven: Yale University Press.

Parisi. Andrew. (2017). “What’s in a Name: A Response to Ramsey.” *Thought* 6(4): 1–10.

Parsons, Charles. (2001). “Communication and the Uniqueness of the Natural Numbers.” *The Proceedings of the First Seminar of the Philosophy of Mathematics in Iran, Shahid University.*

Parsons, Charles. (2008). *Mathematical Thought and Its Objects*. Cambridge: Cambridge University Press.

Parsons, Terence. (1980). *Nonexistent Objects.* New Haven: Yale University Press.

Parsons, Terence. (1982). “Are There Nonexistent Objects?” *American Philosophical Quarterly* 19(4): 365–371.

Paseau, A.C. (2019). “A Measure of Inferential-Role Preservation.” *Synthese* 196(7): 2621–2642.

Pauling, Linus. (1939). *The Nature of the Chemical Bond and the Structure of Molecules and Crystals*. Ithaca: Cornell University Press.

Peacocke, Christopher. (1987). “Understanding Logical Constants: A Realist’s Account.” *Proceedings of the British Academy*: 73: 153–200.

Peacocke, Christopher. (1992). *A Study of Concepts*. Cambridge: The MIT Press.

Peacocke, Christopher. (1993). “Proof and Truth.” In J. Haldane and C. Wright (eds.), *Reality, Representation, and Projection.* Oxford: Oxford University Press.

Peacocke, Christopher. (2004). *The Realm of Reason*. Oxford: Oxford University Press.

Peano, Giuseppe. (1889). *Arithmetices principia, nova methodo exposita*. [Arithmetical Principles of the New Method Explained] Reprinted in English translation in van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931.* Cambridge: Harvard University Press.

Peirce, Charles Sanders. (1878). “How to Make Our Ideas Clear.” *Popular Science Monthly* 12 (Jan.): 286–302.

Penrose, Roger. (1989). *The Emperor’s New Mind*. New York: Oxford University Press.

Penrose, Roger. (1994). *Shadows of the Mind*. Oxford: Oxford University Press.

Peregrin, Jaroslav. (2014). *Inferentialism: Why Rules Matter*. London: Palgrave Macmillan.

Peregrin, Jaroslav. (2018). “Is Inferentialism Circular?” Analysis 78(3): 450–454.
Pinker, Steven. (1999). *Words and Rules: The Ingredients of Language.* Basic Books.
Plantinga, Alvin. (1974). *The Nature of Necessity.* Oxford: Clarendon Press.
Poincaré, Henri. (1902). *La Science et l’Hypothèse*. [Science and Hypothesis] Paris: Flammarion [439] Pollock, John. (1967). “Mathematical Proof.” *American Philosophical Quarterly* 4: 238–244.
Popper, Karl. (1947). “New Foundations for Logic.” *Mind* 56(223): 193–235.
Popper, Karl. (1934). *Logik der Forschung*. [The Logic of Science] Vienna: Julius Springer Verlag.
Potter, Michael. (2000). *Reason’s Nearest Kin: Philosophies of Arithmetic from Kant to Carnap*. Oxford: Oxford University Press.
Prawitz, Dag. (1965). *Natural Deduction: A Proof-Theoretical Study.* Stockholm: Almqvist and Wiksell.
Prawitz, Dag. (1971). “Ideas and Results in Proof Theory.” J.E. Fenstad (ed.), *Proceedings of the Second Scandinavian Logic Symposium*. Amsterdam: North Holland: 235–308.
Price, H.H. (1969). *Belief*. London: Allen and Unwin.
Price, Huw. (1983). “Sense, Assertion, Dummett and Denial.” *Mind* 92: 174–188.
Price, Huw. (1990). “Why “Not”?” *Mind* 99: 221–238.
Priest, Graham. (1979). “The Logic of Paradox.” *Journal of Philosophical Logic* 8: 219–241.
Priest, Graham. (1987). *In Contradiction.* Martinus Nijhoff: Dordrecht.
Priest, Graham. (1995). *Beyond the Limits of Thought.* Cambridge University Press: Cambridge.
Priest, Graham. (1998). “What’s So Bad about Contradictions?” *Journal of Philosophy* 95: 410–426.
Priest, Graham. (2006). *Doubt Truth to Be a Liar.* Oxford: Oxford University Press.
Prior, A. N. (1960). “The Runabout Inference-Ticket.” *Analysis* 21: 2.
Putnam, Hilary. (1960). “Minds and Machines.” In Sidney Hook (ed.), *Dimensions of Mind*. New York: New York University Press: 57–80.
Putnam, Hilary. (1962). “The Analytic and the Synthetic.” In H. Feigl and G. Maxwell (eds.), *Minnesota Studies in the Philosophy of Science* III. Minneapolis: University of Minnesota Press: 33–69.
Putnam, Hilary. (1967). “Mathematics without Foundations.” *Journal of Philosophy* 64(1): 5–22.
Putnam, Hilary. (1971). *Philosophy of Logic*. London: George Allen and Unwin.
Putnam, Hillary. (1974). “The ‘Corroboration’ of Theories.” In Paul A. Schilipp (ed.), *The Philosophy of Karl Popper*. La Salle: Open Court: 221–240.
Putnam, Hilary. (1975). “The Meaning of ‘Meaning.’” In K. Gunderson (ed.), *Language, Mind and Knowledge. Vol. 7 of Minnesota Studies in the Philosophy of Science*. Minneapolis: University of Minnesota Press.

Putnam, Hilary. (1978). *Meaning and the Moral Sciences*. London: Routledge and Kegan Paul.
Putnam, Hilary. (1979) “Philosophy of Mathematics: Why Nothing Works.” In *Words and Life*. Cambridge: Harvard University Press.
Putnam, Hilary. (1980). “Models and Reality.” *Journal of Symbolic Logic* 45(3): 464–482.
Putnam, Hilary. (1981). *Reason, Truth and History.* Cambridge: Cambridge University Press.
Putnam, Hilary. (1983). “Equivalence” in Putnam, Hilary. *Realism and Reason: Philosophical Papers volume 3.* Cambridge: Cambridge University Press.
Putnam, Hilary. (1992). “Is it Necessary that Water is H O?” In Lewis Edwin Hahn (ed.), *The Philosophy of A.J. Ayer*. La Salle: Open Court: 429–454.
Putnam, Hilary. (1994*a*). “Reply to Blackburn.” In Clark and Hale (eds.), *Reading Putnam.* Oxford: Blackwell: 242–254.
Putnam, Hilary. (1994*b*). *Words and Life*. Cambridge: Harvard University Press.
Putnam, Hilary. (2004). *Ethics without Ontology*. Cambridge: Harvard University Press.
Putnam, Hilary. (2015*a*). “Intellectual Autobiography.” In R.E. Auxier, D.R. Anderson, and L.E. Hahn (eds.), *The Philosophy of Hilary Putnam*. Illinois: Open Court: 1–110.
Putnam, Hilary. (2015*b*). “Reply to Hartry Field.” In R.E. Auxier, D.R. Anderson, and L.E. Hahn (eds.), *The Philosophy of Hilary Putnam*. La Salle: Open Court: 173–179.
Quine, W.V. (1936). “Truth by Convention.” In *Philosophical Essays for Alfred North Whitehead*. Longmans, Green, and Company, Inc.: New York: 90–124.
Quine, W.V. (1937). “New Foundations for Mathematical Logic.” *American Mathematical Monthly* 44: 70–80.
Quine, W.V. (1940). *Mathematical Logic*. Cambridge: Harvard University Press.
Quine, W.V. (1946). “Concatenation as a Basis for Arithmetic.” *Journal of Symbolic Logic* 11(4): 105–114.
Quine, W.V. (1948). “On What There Is.” *Review of Metaphysics* 2(5): 21–36.
Quine, W.V. (1951). “Two Dogmas of Empiricism.” *Philosophical Review* 60: 20–43.
Quine, W.V. (1953). “Three Grades of Modal Involvement.” *Proceedings of the XIth International Congress of Philosophy* 14: 65–81.
Quine, W.V. (1958). “Speaking of Objects.” *Proceedings and Addresses of the American Philosophical Association* 31: 5–22.
Quine, W.V. (1960*a*). “Carnap and Logical Truth.” *Synthese* 12: 350–374.
Quine, W. V. O. (1960*b*). *Word and Object*. Cambridge: The MIT Press.
Quine, W.V. (1967). “Variables Explained Away.” *Proceedings of the American Philosophical Society* 104(3): 343–347.
Quine, W.V. (1968). “Ontological Relativity.” *Journal of Philosophy* 65(7): 185–212.

Quine, W.V. (1970*a*). “Methodological Reflections on Current Linguistic Theory.” *Synthese* 21(3/4): 386–398.
Quine, W.V. (1970*b*). *Philosophy of Logic*. Englewood Cliffs: Prentice Hall.
Quine, W.V. and Ullian, J.S. (1978). *The Web of Belief* Second edition. New York: McGraw-Hill.
Raatikainen, Panu. (2005). “On Horwich’s Way Out.” *Analysis* 65(3): 175–177.
Raatikainen, Panu. (2008). “On Rules of Inference and The Meanings of Logical Constants.” *Analysis* 68(4): 282–287.
Ramsey, Frank P. (1925). “Universals.” *Mind* 34: 401–417.
Ramsey, Frank P. (1927). “Facts and Propositions.” *Aristotelian Society Supplementary* 7: 153–170.
Ramsey, Frank. (1978). “Theories” In D.H. Mellor (ed.), *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*. Routledge and Kegan Paul: London: 101–125.
Rawls, John. (1955). “Two Concepts of Rules.” *Philosophical Review* 64(1): 3–32.
Rayo, Agustín (2013). *The Construction of Logical Space*. Oxford: Oxford University Press.
Read, Stephen. (1988). *Relevant Logic: A Philosophical Examination of Inference.* Oxford: Basil Blackwell.
Read, Stephen. (2000). “Harmony and Autonomy in Classical Logic.” *Journal of Philosophical Logic* 29: 123–154.
Read, Stephen. (2010). “General-Elimination Harmony and the Meaning of the Logical Constants.” *J urnal of Philosophical Logic* 39(5): 557–576.
Reichenbach, Hans. (1920). *Relativitätstheorie und Erkenntnis Apriori.* [Relativity Theory and *A priori* Knowledge] Berlin: Springer.
Reichenbach, Hans. (1938). *Experience and Prediction*. Chicago: University of Chicago Press.
Reichenbach, Hans. (1951). *The Rise of Scientific Philosophy*. Berkeley: University of California Press.
Reinhardt, W. N. (1967). *Topics in the Metamathematics of Set Theory*. PhD diss., University of California, Berkeley.
Resnik, Michael D. (1997). *Mathematics as a Science of Patterns.* New York: Oxford University Press.
Restall, Greg. (2000). *An Introduction to Substructural Logics.* New York: Routledge.
Restall, Greg. (2002). “Carnap’s Tolerance, Meaning, and Logical Pluralism.” *Journal of Philosophy* 99(8): 426–443.
Restall, Greg. (2005). “Multiple Conclusions.”In P. Hajek, L. Valdez-Villanueva, and D. Westerstahl (eds.), *Proceedings of the 12th International Congress of Logic, Methodology and Philosophy of Science (King’s College)*.
Restall, Greg. (2008). “Proof Theory and Meaning: On Second Order Logic.” In Michael Pelis (ed.), *Logica 2007 Yearbook*. Filosofia: 157–170.

Ricketts, Thomas. (1994). “Carnap’s Principle of Tolerance, Empiricism, and Conventionalism.” In P. Clark and B. Hale (eds.), *Reading Putnam*. Oxford: Blackwell: 176–200.
Ripley, David. (2011). “Negation, Denial, and Rejection.” *Philosophy Compass* 6(9): 622–629.
Rivest, R., A. Shamir, and L. Adleman. (1978). “A Method for Obtaining Digital Signatures and Public-Key Cryptosystems.” *Communications of the ACM* 21(2): 120–126.
Robinson, R. (1952). “An Essentially Undecidable Axiom System.” *Proceedings of the International Congress of Mathematicians, Cambridge, Mass., 1950* Vol I: 729–730. Providence, R.I.
Rosen, Gideon. (1993). “The Refutation of Nominalism (?)” *Philosophical Topics* 21: 149–186.
Rosen, Gideon. (2010). “Metaphysical Dependence: Grounding and Reduction.” In Bob Hale and Aviv Hoffmann (eds.), *Modality: Metaphysics, Logic, and Epistemology*. Oxford: Oxford University Press: 109–136.
Rossberg, Marcus. (2008). “On the Logic of Quantifier Variance.” Unpublished paper.
Rosser, J. Barkley. (1936). “Extensions of Some Theorems of Gödel and Church.” *Journal of Symbolic Logic* 1: 230–235.
Rosser, J. Barkley. (1942). “The Burali-Forti Paradox.” *Journal of Symbolic Logic* 7: 11–17.
Rumfitt, Ian. (1997). “The Categoricity Problem and Truth-Value Gaps.” *Analysis* 57(4): 223–235.
Rumfitt, Ian. (2000). “‘Yes’ and ‘No’.” *Mind* 109: 781–823.
Russell, Bertrand. (1903). *The Principles of Mathematics*. London: Cambridge University Press.
Russell, Bertrand. (1906). “On Some Difficulties in the Theory of Transfinite Numbers and Order Types.” *Proceedings of the London Mathematics Society* 4: 29–53.
Russell, Bertrand. (1908). “Mathematical Logic as Based on the Theory of Types.” *American Journal of Mathematics* 30: 222–262.
Russell, Bertrand. (1919). “The Philosophy of Logical Atomism.” *The Monist* 29(1): 32–63.
Russell, Bertrand. (1921). *The Analysis of Mind*. London: George Allen and Unwin.
Russell, Bertrand. (1945). *A History of Western Philosophy*. New York: Simon and Schuster.
Russell, Gillian. (2008). *Truth in Virtue of Meaning.* Oxford: Oxford University Press.
Russell, Jeffrey Sanford. (2010). “Comments on Aaron Segal’s ‘Identity Variance’.” Unpublished paper.
Ryle, Gilbert. (1949). *The Concept of Mind*. London: Penguin.

Salmon, Wesley C. (ed.). (1970). *Zeno's Paradoxes*. New York: The Bobbs-Merrill Company, Inc.
Salmon, Wesley C. (1989). "Four Decades of Scientific Explanation." In P. Kitcher and W.P. Salmon (eds.), *Minnesota Studies in the Philosophy of Science* XIII. Minneapolis: Minnesota University Press: 3–219.
Sarukkai, Sundar. (2005). "Revisiting the 'Unreasonable Effectiveness' of Mathematics." *Current Science* 88.3(10): 415–423.
Schaffer, Jonathan. (2009). "On What Grounds What." In David Manley, David J. Chalmers, and Ryan Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press, 347–383.
Schiffer, Stephen. (1996). "Language-Created Language-Independent Entities." *Philosophical Topics* 24(1): 149–167.
Schilpp, Paul (ed.). (1963). *The Philosophy of Rudolf Carnap: Library of Living Philosophers.* La Salle: Open Court.
Schlick, Moritz. (1936). "Are Natural Laws Conventions?" in Schlick (1979): 437–445.
Schlick, Moritz. (1979). *Philosophical Papers: Volume 2*. Dordrecht: Reidel.
Schroyens, W. and W. Schaeken. (2003). "A Critique of Oaksford, Chater, and Larkin's (2000) Conditional Probability Model of Conditional Reasoning." *Journal of Experimental Psychology: Learning, Memory and Cognition* 29: 140–149.
Schwitzgebel, Eric. (2013). "A Dispositional Approach to Attitudes: Thinking Outside the Belief Box." In N. Nottelman (ed.), *New Essays on Belief*. New York: Palgrave Macmillan, 75–99.
Searle, John. (1969). *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: Cambridge University Press.
Searle, John. (1980). "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3(3): 417–457.
Searle, John. (1992). *The Rediscovery of the Mind*. Cambridge: The MIT Press.
Searle, John. (1995). *The Construction of Social Reality.* New York: The Free Press.
Segal, Aaron. (2010). "Identity Variance." Unpublished paper.
Segerberg, Krister. (1982). *Classical Propositional Operators: An Exercise in the Foundations of Logic*. Oxford: Oxford University Press.
Sellars, Wilfrid. (1953). "Inference and Meaning." *Mind* 62(247): 313–338.
Shapiro, Stewart. (1991). *Foundations without Foundationalism: A Case for Second-Order Logic*. Oxford: Clarendon Press.
Shapiro, Stewart. (1997). *Philosophy of Mathematics: Structure and Ontology.* Oxford: Oxford University Press.
Shea, Nicholas, Peter Godfrey-Smith, and Rosa Cao. (2017). "Content in Simple Signalling Systems." *British Journal for the Philosophy of Science* 69(4): 1009–1035.
Sheffer, H.M. (1913). "A Set of Five Independent Postulates for Boolean Algebras." *Transactions of the American Mathematical Society* 14: 481–488.

Shelah, Saharon. (2003). “Logical Dreams.” *Bulletin of the AMS* 40(2): 203–228.
Shoenfield, Joseph. (1967). *Mathematical Logic*. Reading: Addison-Webley.
Shoesmith, D.J. and Timothy Smiley (1978). *Multiple Conclusion Logic.* Cambridge: Cambridge University Press.
Shogenji, Tomoji. (1993). “Modest Scepticism about Rule-Following.” *Australasian Journal of Philosophy* 71(4): 486–500.
Sidelle, Alan. (1989). *Necessity, Essence, and Individuation: A Defense of Conventionalism*. Ithaca: Cornell University Press.
Sidelle, Alan. (2009). “Conventionalism and the Contingency of Conventions.” *Noûs* 43(2): 224–241.
Sider, Theodore. (2003). “Reductive Theories of Modality.” In M.J. Loux and D.W. Zimmerman (eds.), *Oxford Handbook of Metaphysics*. Oxford: Oxford University Press.
Sider, Theodore. (2007). “Neo-Fregeanism and Quantifier Variance.” *Aristotelian Society, Supplementary Volume* 81: 201–232.
Sider, Theodore. (2009). “Ontological Realism.” In D.J. Chalmers, D. Manley, and R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press: 384–423.
Sider, Theodore. (2011). *Writing the Book of the World.* Oxford: Oxford University Press.
Skolem, Thoraf. (1922). “Some Remarks on Axiomatized Set Theory.” In van J. Heijenoort (ed.), *From Frege to Gödel: A Sourcebook in Mathematical Logic, 1879–1931*. Cambridge: Harvard University Press: 290–301.
Skolem, Thoraf. (1923). “The Foundations of Elementary Arithmetic Established by Means of the Recursive Mode of Thought, without the Use of Apparent Variables Ranging over Infinite Domains.” In van J. Heijenoort (ed.), *From Frege to Gödel: A Sourcebook in Mathematical Logic, 1879–1931.* Cambridge: Harvard University Press: 302–333.
Skolem, Thoraf. (1934). “Über die Nicht-charakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen.” [Abotu the non-characterizability of the Number Series by Means of Finite or Countably Infinite Statements with Exclusively Numerical Variables] *Fundamenta Mathematicae* 23: 150–161.
Skorupski, John. (1993). “Anti-Realism, Inference, and the Logical Constants.” In J. Haldane and C. Wright (eds.), *Reality, Representation, and Projection.* Oxford: Oxford University Press: 133–164.
Skyrms, Brian (2000). “The Evolution of Inference.” In Tim Kohler and George Gumerman (eds.), *Dynamics of Human and Primate Societies*. New York: Oxford University Press.
Skyrms, Brian. (2010). *Signals: Evolution, Learning, and Information*. Oxford: Oxford University Press.
Smiley, Timothy. (1996). “Rejection.” *Analysis* 56: 1–9.
Smith, A.D. (1977). “Dispositional Properties.” *Mind* 86: 439–445.

Smith, Peter. (2013). *An Introduction to Gödel's Theorems.* Cambridge: Cambridge University Press.
Smullyan, Arthur. (1948). "Modality and Description." *Journal of Symbolic Logic* 13(1): 31–37.
Soames, Scott. (1983). "Generality, Truth Functions, and Expressive Capacity in the *Tractatus*." *Philosophical Review* 92(4).
Soames, Scott. (2003). *Philosophical Analysis in the Twentieth Century: Volume 1: The Dawn of Analysis*. Princeton: Princeton University Press.
Sober, Elliot. (2000). "Quine's Two Dogmas." *Proceedings of the Aristotelian Society*, Supplementary 74: 237–280.
Stalnaker, Robert. (1984). *Inquiry*. Cambridge: The MIT Press.
Steiner, Mark. (1978). "Mathematical Explanation." *Philosophical Studies* 34(2): 135–151.
Steiner, Mark. (1998). *The Applicability of Mathematics as a Philosophical Problem*. Cambridge: Harvard University Press.
Stewart, Ian. (1989). *Galois Theory*. Second edition. Boca Raton: Chapman and Hall/CRC.
Stove, David C. (1982). *Popper and After*. Oxford: Pergamon Press.
Stove, David C. (2004). "Cole Porter and Karl Popper: The Jazz Age in the Philosophy of Science." In Anthony O'Hear (ed.), *Karl Popper: Critical Assessments of Leading Philosophers: Volume 2: Philosophy of Science 1*. London: Routledge: 39–58.
Strawson, Peter. (1974). *Subject and Predicate in Logical Grammar*. Farnham: Ashgate.
Strevens, Michael. (2008). *Depth: An Account of Scientific Explanation*. Cambridge: Harvard University Press.
Stroud, Barry. (1965). "Wittgenstein and Logical Necessity." *Philosophical Review* 74(4): 504–518.
Syverson, Paul. (2002). *Logic, Convention, and Common Knowledge: A Conventionalist Account of Logic*. Stanford: CSLI Publications.
Tait. W.W. (2001). "Beyond the Axioms: The Question of Objectivity in Mathematics." *Philosophia Mathematica* 9(1): 21–36.
Tarski, Alfred. (1933*a*). *Pojecie Prawdy w Jezykach Nauk Dedukcyjnych*. Prace Towarzystwa Naukowego Warszawskiego, Wyzdial III matematycznofizycznych, no. 34. Warsaw.
Tarski, Alfred. (1933*b*). "Some Observations on the Concepts of ω-Consistency and ω-Completeness." In Tarski (1956). *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Oxford: Oxford University Press: 279–295.
Tarski, Alfred. (1936). "On the Concept of Logical Consequence." In Tarski (1956).
Tarski, Alfred. (1956). *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Oxford: Oxford University Press.
Tarski, Alfred. (1986) "What Are Logical Notions?" (1966 lecture) in J. Corcoran (ed.), *History and Philosophy of Logic* 7: 143–154.

Tarski, Alfred, Andrzej Mostowski, and Raphael Robinson. (1953). *Undecidable Theories*. Amsterdam: North Holland.
Tennant, Neil. (1978). *Natural Logic*. Edinburgh: Edinburgh University Press.
Tennnant, Neil. (1997). *Anti-Realism and Logic: Truth as Eternal*. Oxford: Oxford University Press.
Tennant, Neil. (2002). *The Taming of the Tru*e. Oxford: Oxford University Press.
Thomasson, Amie. (2009). "Answerable and Unanswerable Questions." In D.J. Chalmers, D. Manley, and R. Wasserman (eds.), *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press: 444–471.
Thomson, James. (1954). "Tasks and Super-Tasks." *Analysis* 15: 1–13.
Topey, Brett. (2019). "Linguistic Convention and Worldly Fact: Prospects for a Naturalist Theory of the *A Priori*." *Philosophical Studies* 176(7): 1725–1752.
Troelstra, A.S. and H. Schwichtenberg. (2000). *Basic Proof Theory*. Second edition. Cambridge: Cambridge University Press.
Turner, Jason. (2010). "Ontological Pluralism." *Journal of Philosophy* 107(1): 5–34.
Turner, Jason. (2012). "Logic and Ontological Pluralism." *Journal of Philosophical Logic* 41(2): 419–448.
Unger, Peter. (2014). *Empty Ideas: A Critique of Analytic Philosophy.* New York: Oxford University Press.
Unger, Peter. (2017*a*). "Being Concretely Informative is One Thing, Being Concretely Substantial is Quite Another." Unpublished paper.
Unger, Peter. (2017*b*). "Emptiness Is Not Necessity, Whatever, if Anything, the Latter Might Be." Unpublished paper.
Van Bentham, Johan. (2008). "Logic and Reasoning: Do the Facts Matter?" *Studia Logica* 88(1): 67–84.
Van Fraassen, Bas. (1966). "Singular Terms, Truth-Value Gaps, and Free Logic." *Journal of Philosophy* 63(17): 481–495.
Van Fraassen, Bas. (1980). *The Scientific Image*. Oxford: Clarendon Press.
Van Inwagen, Peter. (1990). *Material Beings*. Ithaca: Cornell University Press.
Vandoulakis, Ioannis M. (2015). "On A.A. Markov's Attitude towards Brouwer's Intuitionism." *Philosophia Scientiæ* 19(1): 143–158.
Waismann, Friedrich. (1968). *How I See Philosophy*. NewYork: Macmillan.
Walmsley, James. (2002). "Categoricity and Indefinite Extensibility." P*roceedings of the Aristotelian Society* 102(3): 217–235.
Warren, Jared. (2015*a*). "Conventionalism, Consistency, and Consistency Sentences." *Synthese* 192(5): 1351–1371.
Warren, Jared. (2015*b*). "Talking with Tonkers." *Philosophers' Imprint* 15(24): 1–24.
Warren, Jared. (2015*c*). "Quantifier Variance and the Collapse Argument." *Philosophical Quarterly* 65(259): 241–253.
Warren, Jared. (2015*d*). "The Possibility of Truth by Convention." *Philosophical Quarterly* 65(258): 84–93.
Warren, Jared. (2016*a*). "Trapping the Metasemantic Metaphilosophical Deflationist?" *Metaphilosophy* 47(1): 108–121.

Warren, Jared. (2016*b*). “Sider on the Epistemology of Structure.” *Philosophical Studies* 173(9): 2417–2435.
Warren, Jared. (2016*c*). “Internal and External Questions Revisited.” *Journal of Philosophy* 113(4): 177–209.
Warren, Jared. (2017*a*). “Revisiting Quine on Truth by Convention.” *Journal of Philosophical Logic* 46(2): 119–139.
Warren, Jared. (2017*b*). “Quantifier Variance and Indefinite Extensibility.” *Philosophical Review* 126(1): 81–122.
Warren, Jared (2017*c*). “Epistemology versus Non-Causal Realism.” *Synthese* 194(5): 1643–1662.
Warren, Jared. (2018). “Change of Logic, Change of Meaning.” *Philosophy and Phenomenological Research* 96(2): 421–442.
Warren, Jared. (forthcoming *a*). “Killing Kripkenstein’s Monster.” *Noûs.*
Warren, Jared. (forthcoming *b*). “Ontological Commitment and Ontological Commitments.” *Philosophical Studies*.
Warren, Jared and Daniel Waxman. (forthcoming *a*). “A Metasemantic Challenge for Mathematical Determinacy.” *Synthese*.
Warren, Jared and Daniel Waxman. (forthcoming *b*). “Supertasks and Arithmetical Truth.”
Wason, P.C. (1966). “Reasoning.” In B.M. Foss (Ed.), *New Horizons in Psychology 1*. Harmondsworth: Penguin.
Wason, P.C. (1968). “Reasoning about a Rule.” *Quarterly Journal of Experimental Psychology* 20(3): 273–281.
Wason, P.C. and D. Shapiro. (1971). “Natural and Contrived Experience in a Reasoning Problem.” *Quarterly Journal of Experimental Psychology* 23: 63–71.
Waxman, Daniel. (2019*a*) “Is Mathematics Unreasonably Effective?” Unpublished paper.
Waxman, Daniel. (2019*b*). “Did Gentzen Prove the Consistency of Arithmetic?” Unpublished paper.
Waxman, Daniel. (2019*c*). “Imagining the Infinite.” Unpublished paper.
Wedgwood, Ralph. (2007). *The Nature of Normativity*. New York: Oxford University Press.
Weir, Alan. (1986). “Classical Harmony.” *Notre Dame Journal of Formal Logic* 27(4):459–482.
Weir, Alan. (2010). *Truth through Proof: A Formalist Foundation for Mathematics.* Oxford: Clarendon Press.
Weston, Thomas. (1976). “Kreisel, the Continuum Hypothesis and Second Order Set Theory.” *Journal of Philosophical Logic* 5(2): 281–298.
Weyl, Herman. (1949). *Philosophy of Mathematics and Natural Science.* Princeton: Princeton University Press.
Whiteley, C.H. (1936). “Truth by Convention: A Symposium” *Analysis* 4(2/3): 22–28.

Wigglesworth, John. (2017). “Logical Anti-Exceptionalism and Theoretical Equivalence.” *Analysis* 77(4): 759–767.
Wigner, Eugene. (1960). “The Unreasonable Effectiveness of Mathematics in the Natural Sciences.” *Communications on Pure and Applied Mathematics* 13: 1–14.
Willard, Dan. (1993). “Self-Verifying Axiom Systems.” *KGC 1993: Computational Logic and Proof Theory*: 325–336.
Willard, Dan. (2001). “Self-Verifying Axiom Systems, the Incompleteness Theorem and Related Reflection Principles.” *Journal of Symbolic Logic* 66(2): 536–596.
Williamson, Timothy. (1990). *Identity and Discrimination*. Oxford: Basil Blackwell.
Williamson, Timothy. (1994). *Vagueness*. Routledge: London and New York.
Williamson, Timothy. (2006*a*). “Absolute Identity and Absolute Generality.” In A. Rayo and G. Uzquiano (eds.), *Absolute Generality.* Oxford: Oxford University Press.
Williamson, Timothy. (2006*b*). “Conceptual Truth.” *Aristotelian Society Supplementary Volume* 80(1): 1–41.
Williamson, Timothy. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell Publishing.
Williamson, Timothy. (2013). “How Deep Is the Distinction Between *A Priori* and *A Posteriori* Knowledge?” In Albert Casullo and Joshua C. Thurow (eds.), *The A Priori in Philosophy*. Oxford: Oxford University Press: 291–312.
Wilson, Jessica. (2014). “No Work for a Theory of Grounding.” *Inquiry* 57(5–6): 535–579.
Wilson, Mark. (1982). “Predicate Meets Property.” *Philosophical Review* 91(4): 549–589.
Wittgenstein, Ludwig. (1922). *Tractatus Logico-Philosophicus*. English translation by C.K. Ogden and F.P. Ramsey. Routledge and Kegan Paul: London.
Wittgenstein, Ludwig. (1929). “Some Remarks on Logical Form.” *Proceedings of the Aristotelian Society, Supplementary Volume 9: Knowledge, Experience and Realism*: 162–171.
Wittgenstein, Ludwig. (1953). *Philosophical Investigations.* Englewood Cliffs: Prentice Hall.
Wittgenstein, Ludwig. (1956). *Remarks on the Foundations of Mathematics.* New York: Macmillan.
Wittgenstein, Ludwig. (1974). *Philosophical Grammar*. Oxford: Basil Blackwell.
Woods, Jack. (2014). “Logical Indefinites.” *Logique Et Analyse: Special Issue* 227: 277–307.
Woods, Jack. (2016). “Assertion, Denial, Content, and Form.” *Synthese* 193(6): 1667–1680.
Woods, Jack. (2018). “Intertranslatability, Theoretical Equivalence, and Perversion.” *Thought* 7(1): 58–68.
Woods, Jack. (2019). “Logical Partisanhood.” Philosophical Studies 176(5): 1203–1224.

Woods, Jack. (forthcoming). “The Frege-Geach Problem” In Tristram McPherson and David Plunkett (eds.), *The Routledge Handbook of Metaethics*.
Woodward, Jim. (2011). “Causes, Conditions, and the Pragmatics of Causal Explanation.” In Morgan (ed.), *Philosophy of Science Matters: The Philosophy of Peter Achinstein*. Oxford: Oxford University Press: 247–257.
Wright, Crispin. (1980). *Wittgenstein on the Foundations of Mathematics.* London: Duckworth.
Wright, Crispin. (1983). *Frege’s Conception of Numbers as Objects.* Aberdeen: Aberdeen University Press.
Wright, Crispin. (1985). “Facts and Certainty.” *Proceedings of the British Academy* 7: 429–472.
Wright, Crispin. (1992). *Truth and Objectivity.* Cambridge: Cambridge University Press.
Wright, Crispin. (2000). “Cogency and Question-Begging: Some Reflections on McKinsey’s Paradox and Putnam’s Proof.” *Philosophical Issues* 10: 140–163.
Wright, Crispin. (2001). “On Basic Logical Knowledge: Reflections on Paul Boghossian’s ‘How Are Objective Epistemic Reasons Possible?’” *Philosophical Studies* 106: 41–85.
Wright, Crispin. (2004). “Intuition, Entitlement, and the Epistemology of Logical Laws.” *Dialectica* 58(1): 155–175.
Wright, Crispin. (2014). “Comment on Paul Boghossian, ‘What is Inference?’” *Philosophical Studies* 169(1): 27–37.
Yablo, Stephen (1992). “Review of Alan Sidelle’s *Necessity, Essence, and Individuation: A Defense of Conventionalism.*” *Philosophical Review*, 101(4): 878–881.
Yablo, Stephen. (1993). “Is Conceivability a Guide to Possibility?” *Philosophy and Phenomenological Research* 53 (1): 1–42.
Zangwill, Nick. (2015). “Logic as Metaphysics.” *Journal of Philosophy* 112(10): 517–550.
Zangwill, Nick. (2019). “Logic and Inference Again.” Unpublished paper.
Zermelo, Ernst. (1930). “Über Stufen der Quantifikation und die Logik des Unendlichen.” [About Levels of Quantification and the Logic of the Infinite] *Jahresbericht der Deutschen Mathematiker-Vereinigung (Angelegenheiten)* 31: 85–88.

Name Index

Achinstein, Peter 12
Anderson, Alan 150
Antonelli, Aldo 85
Armstrong, David 36
Audi, Rober 36
Awodey, Steve 230, 326, 327, 328
Ayer, A.J. 4, 7–9, 34, 72, 108, 188, 197–198, 203, 304, 321
Azzouni, Jody 210

Baker, G.P. 5
Balaguer, Mark 177, 290, 343
Bar-Hillel, Yehoshua 11
Beall, J.C. 138
Bell, John 271
Belnap, Nuel 10, 90, 108, 150
Benacerraf, Paul 4–5, 181, 197–198, 205, 215, 236, 264, 267, 280, 290
Ben-Menahem, Yemima 8
Bendall, Kent 84
Bennett, Jonathan 188, 190
Bergmann, Michael 345
Berko, Jean 26–27
Bermudez, Jose Luis 15
Bernays, Paul 297, 302, 303, 310
Berry, Sharon 251–255, 260, 267, 288
Besson, Corine 45
Beth, E.W. 306, 329
Birch, Jonathan 22
Black, Max 157, 187, 330
Blackburn, Simon 9, 41
Blanchette, Patricia A. 187–188
Block, Ned 16, 36, 73, 92
Boghossian, Paul 17, 25, 33, 34, 37, 38, 42, 43, 44, 48, 51, 56, 63, 120–121, 157–166, 169, 173–176, 184, 228
Bonevac, Daniel A. 148
BonJour, Laurence 19, 153

Bonnay, Denis 328
Boolos, George 97, 144, 210, 213, 290,
Bostock, David 79 Bourget, David 55
Bowdery, George 330
Bradley, F.H. 343
Brandom, Robert 33, 48, 51, 74, 78, 88, 92, 131, 168, 218, 234
Britton, Karl 5
Broome, John 25, 35, 48, 116
Bueno, Otávio 231
Buldt, Bernd 263
Burge, Tyler 130, 161, 224, 284
Burgess, Alexis G. 134
Burgess, John P. 36, 101, 107, 134, 144, 150, 201, 283
Button, Tim 244
Byrne, R.M.J. 28

Camp, Joseph 108
Cao, Rosa 22, 234
Carnap, Rudolf 5, 9, 11, 23, 24, 28, 30, 33, 35–36, 62–64, 72, 78, 80–85, 88, 90, 95, 98, 106, 114, 126, 127–129, 133, 135, 151–152,
164, 165, 198, 203, 204, 209, 214, 215, 217, 227, 229, 263, 265, 270, 278, 280, 281–284, 304, 306, 311, 313–316, 321, 325–330, 332,
335–337
Carroll, Lewis 50, 183
Carus, A.W. 326
Chalmers, David J. 17, 36, 55, 61, 215, 258, 265
Chihara, Charles 134
Chomsky, Noam 11, 26, 28, 49, 184
Chow, Timothy Y. 299
Christensen, David 343, 344
Church, Alonzo 19, 205, 240, 289, 296
Clark, Peter 336
Coffa, J. Alberto 3, 5, 6, 8, 165, 263, 265, 327–329, 335–337
Colyvan, Mark 150
Cook, Roy 90, 136
Copleston, Fredrick 188, 304
Cosmides, Leda 28
Cottingham, John 202
Curry, Haskell 134, 311

Daly, Christopher 14
Davidson, Donald 35, 91–92, 111, 127
Davies, E.B. 253, 267

Davies, Martin 102,
Davies, P.S. 28
Davis, Philip J. 201
Dawson Jr., John D. 202, 313–316
Dean, Walter 312
de Bouvère, K. 145
Dedekind, Richard 199, 247–250, 272
Dennett, Daniel C. 35, 91, 92
Densmore, Dana 209 Detlefsen, Michael 297 de Rouilhan, Philippe 327 Diamond, Cora 134, 186
Dogramaci, Sinan 167
Dorr, Cian 226–227, 230–232,
Došen, Kosta 88
Dreier, Jamie 112
Dummett, Michael 63, 90, 151, 157, 167, 185–189, 198, 204, 213, 218, 220, 239, 262, 286, 288, 313
Dummit, D.S. 237

Earman, John 253, 267,
Ebbs, Gary 9, 11, 335–336 Edgington, Dorothy 165
Einheuser, Iris 3, 114, 172,
Eklund, Matti 204, 214, 223, 224, 231
Elga, Adam 344–345
Enderton, Herbert 97, 107, 218
Enoch, David 344
Evans, Gareth 102
Ewing, A.C. 173

Feferman, Solomon 146, 202, 242, 297, 313–316
Felleisen, Matthias 147
Feyerabend, Paul 130, 338
Field, Hartry 18, 58, 92, 108–111, 116, 154, 155, 203, 205, 232, 242, 249–255, 262, 290, 307
Fine, Kit 16, 34, 151, 198, 203, 207, 210, 241, 246, 287, 343
Finn, Suki 231
Floyd, Juliet 305,
Fodor, Jerry A. 16, 35, 42, 43, 56, 66
Fogelin, Robert 322
Foote, R.M. 237
Forbes, Graeme 42
Forster, Michael N. 42
Franzén, Torkel 295
Frege, Gottlob 8, 30, 31, 36, 201, 213, 235, 280, 284–285, 289, 332

French, Rohan 145
Friedman, Michael 6, 9, 12, 88, 326, 347

Garson, James W. 79–81, 91
Gasking, D.A.T. 9
Geach, Peter 8, 30, 218, 322
Gelman, Susan A. 87
Gentzen, Gerhard 29, 62–63, 146, 298–299, 315
Gettier, Edmund 39
Giannoni, Carlo 3, 9, 310–311, 321, 330–332
Gibbard, Alan 9
Ginsborg, Hannah 42
Glivenko, V. 135, 146
Glock, Hans-Johann 3, 9
Glymour, Clark 145
Gödel, Kurt 19, 98, 107–108, 146, 154, 197–198, 202, 205, 207–208, 246, 264, 278, 295–297, 301–305, 310, 313–316, 326
Godfrey-Smith, Peter 22
Goldfarb, Warren 202, 313–316, 329, 335
Goodman, Nelson 203
Grandy, Richard 311
Grattan-Guinness, Ivor 293
Grover, Dorothy 108
Guardo, Andrea 42

Haack, Susan 88, 146, 157
Hacker, P.M.S. 5
Hacking, Ian 63, 76, 84, 91
Hale, Bob 198, 204, 214, 218, 220, 228, 336, 343
Hamkins, Joel David 262, 316
Hanson, William H. 77, 150
Harman, Gilbert 31, 45, 68, 73, 92, 111, 116, 142, 339
Harris, J.H. 228, 230, 242
Hauser, Kai 316
Hawley, Katherine 204, 214
Hawthorne, John 231
Haze, Tristan 176
Hellman, Geoffrey 3, 9, 11, 24, 103, 107, 236
Hempel, Carl G. 4–5, 15, 88
Henkin, Leon 244
Hersh, Reuben 201
Hilbert, David 8, 263, 284, 297, 302, 303, 310, 342
Hirsch, Eli 24, 148, 204, 209, 214–215, 219–221, 223–228, 231, 329

Hjortland, Ole T. 84, 98
Hofweber, Thomas 16
Hogarth, Mark 267
Horn, Laurence R. 234
Horwich, Paul 92, 108–111, 223, 241
Humberstone, Lloyd 98, 102, 128
Hume, David 34, 201, 202, 235, 347

Incurvati, Luca 81, 84

Jackendoff, Ray 26,
Jackson, Frank 61, 110 Jeffrey, Richard C. 97
Johnston, Mark 39
Juhl, Cory 324

Kanamori, Akihiro 316
Kant, Immanuel 9, 17, 202–203, 221, 232–235, 237, 257
Kaplan, David 6, 148, 225
Kaye, Richard 251
Kelly, Thomas 344–345
King, Nathan L. 344
Kitcher, Philip 12
Kleene, Stephen 84, 289, 327
Kneale, William 62
Kocurek, Alexander 225
Koellner, Peter 301, 304, 306, 309, 326
Kolenda, Konstantin 332
Kripke, Saul A. 6, 34, 36, 40–43, 47, 56, 60, 67, 91, 93, 119, 130, 166, 189, 223, 307
Kroedel, Thomas 84, 137
Kuhn, Thomas 130–131, 338
Kunen, Kenneth 248, 272, 289
Kusch, Martin 41

Lakatos, Imre 130, 201, 338
Lakoff, George 201
Lavine, Shaughan 32, 242, 245, 247, 250
Leitgeb, Hannes 312
Lemmon, E.J. 156
Lepore, Ernest 66
Levine, Joseph 17
Lewis, C.I. 173

Lewis, David 16, 21–23, 36, 38, 57, 72, 91, 92, 104, 115, 140, 160, 204, 217, 224, 254, 290, 308
Lewy, Casimir 173
Linaza, José 46
Lindström, Per 85
Linnebo, Øystein 198
Lipton, Peter 13
Löb, M.H. 310
Loewer, Barry 16
Loomis, Eric 324
Lucas, J.R. 17, 34
Łukasiewicz, J. 30

Machover, Moshé 271
Mackie, J.L. 17
MacLane, Saunders 291, 328
Malcolm, Norman 5, 173
Malmgren, Anna-Sara 39, 162
Marcus, Ruth B. 35
Martin, C.B. 39
Mates, Benson 156
Matheson, Jonathan 343, 345
McCawley, James D. 78, 80
McDaniel, Kris 224
McDowell, John 111
McGee, Vann 11, 19, 32, 75, 110, 192–194, 205, 242–243, 250–254, 256, 260, 273
McGinn, Colin 17, 203, 210
McTaggart, J.M.E. 343
Mehlberg, Henryk 162, 207, 241
Merrill, G.H. 57, 254
Midgley, G.C.J. 50, 119
Miller, Alexander 43
Montague, Richard 59, 309–313
Morris, Errol 130
Morton, Adam 128
Moss, Sarah 114
Mostowski, Andrzej 85, 108, 145, 225
Murdoch, Dugald 202
Murzi, Julien 45, 84, 98
Mutchnik, Scott 310
Myhill, John 312

Nagel, Ernest 4, 16
Nagel, Thomas 17
Nelson, Edward 148, 298
Neta, Ram 37
Newman, M.H. 293
Norton, John D. 253, 267
Núñez, Raphael E. 201

Oaksford, M. 28
Odifreddi, Piergiorgio 19
Oppenheim, Paul 15, 16
Oppy, Graham 110

Paoli, Francesco 88
Pap, Arthur 6, 173
Parisi, Andrew 218
Parsons, Charles 32, 199, 202, 242, 247, 249, 250, 252–253, 255, 257–258, 262, 313–316
Parsons, Terence 203, 210
Paseau, A.C. 141
Pauling, Linus 13
Peacocke, Christopher 27, 63, 66, 90, 161
Peano, Giuseppe 199
Peirce, Charles Sanders 260
Penrose, Roger 17, 34
Peregrin, Jaraslov 78, 120
Pinker, Steven 26
Plantinga, Alvin 202
Poincaré, Henri 8, 144, 337
Pollock, John 198, 301, 309–313
Popper, Karl 62, 338
Potter, Michael 263
Prawitz, Dag 90
Price, H.H. 36
Price, Huw 31
Priest, Graham 31–32, 126, 139, 285
Prior, A.N. 72, 93
Putnam, Hilary 4, 5, 16, 33, 34, 36, 56, 57, 61, 73, 88, 104, 114, 130, 148, 187, 204, 205, 209, 214, 215, 219–221, 223–227, 228, 231, 236, 244, 253, 259, 260, 267, 284, 290, 293, 301, 303–309, 329, 336, 338

Quine, W.V. 6, 9, 13, 18, 21, 23, 24, 26, 34, 49, 50, 91, 92, 96, 108, 127, 128, 142, 144, 181–185, 186, 189, 191, 192, 203, 209–212,

224, 232, 235, 281–283, 289, 303, 307, 328, 333, 336

Raatikainen, Panu 78, 265
Ramsey, Frank P. 36, 108, 114, 204, 217, 218
Rawls, John 50, 119
Rayo, Agustín 198, 343
Read, Stephen 90, 91, 150,
Reichenbach, Hans 5, 9, 304, 334–337
Reinhardt, W. N. 289
Resnik, Michael D. 236
Restall, Greg 71, 78, 80, 88, 138, 244
Ricketts, Thomas 11, 335–336
Ripley, David 30
Rivest, R. 292
Robinson, Raphael M. 108, 145, 154, 225
Rosen, Gideon 16, 18, 212
Rossberg, Marcus 228
Rosser, J. Barkley 205, 207, 208, 289, 302
Rumfitt, Ian 31, 80, 81, 83, 84, 111
Russell, Bertrand 24, 35, 59, 114, 245, 280, 284–286, 288, 289, 332
Russell, Gillian 324
Russell, Jeffrey Sanford 221
Ryle, Gilbert 34

Salmon, Wesley C. 15, 267
Sarukkai, Sundar 291
Schaffer, Jonathan 16, 233
Schiffer, Stephen 214
Schilpp, Paul Arthur 313
Schlick, Moritz 328
Schroyens, W. 28
Schaeken, W. 28
Schwichtenberg, H. 79
Schwitzgebel, Eric. 35
Searle, John 34, 36, 50, 77, 119, 228
Segal, Aaron 221
Segerberg, Krister 145
Sellars, Wilfrid 33, 74, 92, 131
Shamir, A. 292
Shapiro, D. 28
Shapiro, Stewart 107, 218, 236, 242, 244
Shea, Nicholas 22
Sheffer, H.M. 322

Shelah, Saharon 316
Shoenfield, Joseph 242
Shoesmith, D.J. 80
Shogenji, Tomoji 41
Sidelle, Alan 3, 6, 61, 173
Sider, Theodore 16, 105, 151, 174–175, 181, 184, 204, 205, 213, 214, 226, 227, 230, 254
Skolem, Thoraf 32, 85, 199, 271, 293
Skorupski, John 90
Skyrms, Brian 22–23
Smiley, Timothy 30–31, 80–81, 98
Smith, A.D. 39
Smith, Michael 110
Smith, Peter 81, 84, 264
Smullyan, Arthur 235
Soames, Scott 181, 322
Sober, Elliot 173
Solovay, Robert N. 202, 313–316
Stalnaker, Robert 148
Steinberger, Florian 45
Steiner, Mark 13, 293
Stewart, Ian 180
Stoothoff, Robert 202
Stove, David C. 338
Strawson, Peter 218
Strevens, Michael 12
Stroud, Barry 186
Syverson, Paul 3, 22, 330

Tait, W.W. 267
Tarski, Alfred 75, 108, 145, 154, 155, 215, 225, 263–265, 307, 312, 326, 327, 329
Tennant, Neil 63, 90, 199, 278
Thomasson, Amie 214
Thomson, James 182, 264
Tooby, John 28
Topey, Brett 176
Troelstra, A.S. 79
Turner, Jason 224, 229

Unger, Peter 112, 178, 179

Van Bentham, Johan 28
Van Fraassen, Bas 12, 15, 162, 207, 241

Van Inwagen, Peter 220
Vandoulakis, Ioannis M. 19

Waismann, Friedrich. 8
Walmsley, James 248
Walsh, Sean 244
Wason, P.C. 28
Waxman, Daniel 18, 205, 207, 241, 258, 267, 293, 299, 311
Wedgwood, Ralph 344
Weir, Alan 91, 265, 343
Weston, Thomas 244, 273
Weyl, Herman 267
Whiteley, C.H. 187
Wigglesworth, John 145–146
Wigner, Eugene 291–292
Willard, Dan 297, 315
Williamson, Timothy 32, 34, 81, 192–194, 207, 211, 241, 242, 283
Wilson, Jessica 16
Wilson, Mark 187
Wittgenstein, Ludwig 3, 5, 9, 40, 41, 62, 63, 88, 92, 103, 114, 134, 185–187, 190, 280, 304, 305, 321–322, 337
Woodin, W. Hugh 316
Woods, Jack 8, 75, 88, 107, 145, 146, 148, 150
Woodward, Jim 15
Wright, Crispin 34, 37, 51, 125, 157, 159, 161, 186, 198, 204, 213, 214, 220, 228, 232, 343

Yablo, Stephen 173, 175, 258
Yang, Ruizhi 262

Zangwill, Nick 120, 121
Zermelo, Ernst 262, 272, 284, 285

Subject Index

Abstraction Principles 201, 204, 211–212
Acceptance 29–32, 33–37, 48–49
Anti-Overspill Argument
  Berry's 253–255
  Field-style 252–253
  McGee's 250–252
Applications of Mathematics 291–294
Arithmetization of Syntax 296–298, 301–303

Bad Company 157, 166–169
Basic Rules 66–72
Basic Understanding 86–87, 279–280
Bivalence 107, 206–207, 261–262
Bilateralism 30–32, 78–86, 132–133
Bilateralist Arithmetic 208

Carnap's Problem 78–86
  Solution for Connectives 78–84
  Solution for Quantifiers 84–86
Categoricity 239–242, 244, 247–263, 271–273
Charity 91–94, 125–131
Circularity Objection
  Against Logical Conventionalism 185–192
  Against Mathematical Conventionalism 306–309
Collapse Argument
  Against Variance 228–230
  Dorr's 230–232
  For Categoricity 247–250
Communicative Generality Argument 255–261
Completeness Theorem 98–99, 106–108, 122–124
Comprehension Principles
  In Second-Order Logic 218–219
  In Set Theory 284–288
Consistency Sentence 296–299, 303–306
Contingency Argument 171–173

Conventionalism
  Argument for Logical… 95–102
  Argument for Mathematical… 276–281
  Descriptive Version 7–8
  Explanatory Version 9–11
  Identity Version 8–9
Conventionalism in Logic 330–332

Deflationism 108–111
Derivative Rules 66–72
Determinacy
  Arithmetical 239–242, 261–263
  Mathematical 204–207
  Operators 207
Disagreement 149–152
Dispositionalism
  About Acceptance and Rejection 33–37
  About Inference 37–40
  About Inference Rule-Following 43–47
  General theory 48–51
  Kripkenstein versus 40–43
Dispositions
  Complex 41–42, 71
  Composite 41–42, 71
  Simple 41

Epistemology
  Of Consistency 296–299
  Of Logic 153–169
  Of Mathematics 294–296
Equivalence 84, 129, 143–144, 145–149, 225–227
Explanation 11–16
Explicit Conventionalism 181–183
Externalism
  Epistemological/Normative 116, 166
  Semantic 130, 254, 259

Facts 111–113, 114, 173–181
Freedom in Mathematics 288–291
Functionalism 36–38

Generalized Quantifiers 85–86
Gödel's Argument 313–316

Gödel-Rosser Sentence 205, 207, 208, 302

Heavyweight Metaphysics 16, 143, 145, 151, 176, 178–179, 213, 224, 226
Holism 66–69, 129, 131–137, 141
Hyperintensionality 14–16, 143, 176–180

Identity Criteria 210–212
Incompleteness Theorems
  First 106–108, 197, 204–207, 208, 301–303
  Second 296–299, 301–303, 303–306, 313–316
Indefinite Extensibility 284–288
Inference 37–40
Inference Rules 25–33, 43–51
Inference Rule-following
Inferentialism 55–94, 197–201

Kant's Challenge 202–204, 232–235
Knowledge
  Conditions on Arguments 157–162
  Of Consistency 296–299
  Of Logical Validity 153–170
Kripkenstein 40–43

Lazy Gauss 159–160
Logical Inferentialism Principle 56
Logical Syntax of Language 325–330
Logicality Principle 76, 280

Martians 36, 38, 100, 106, 122, 139–140, 224, 258–259
Master Argument 173–181
Metaontology 209–238
Metaphilosophy 334–343
Meanings Are Cheap (MAC) 64
Meaning Entitlement Connection (MEC) 163
Meaning Validity Connection (MVC) 58
Metadeflationism 215–216, 219
Montague's Paradox 309–313

Naturalism 16–20
Necessity 101–102
Neo-Fregeanism 198, 201, 203–204, 211, 213–214, 220, 235
Normativity
  Of Logic 115–118

Omega Rule 263–270, 274–275, 276–281
Ontological Commitment 209–214
Open-endedness 32–33, 242–247, 270–273
Oracle Example 282

Peano Rules 199–200
Peer Disagreement 343–346
Philosophical Progress 341–343
Pluralism
  Logical 125–152
  Mathematical 239–242, 303–306
  Name 229
  Predicate 229, 231
  Quantifier 221–227
Plutonians 304–308
Pollock's Argument 309–313
Principle of Tolerance 106, 151–152, 326, 337
Putnam-Koellner Argument 303–306

Quantifier Variance
  Modest 226–227, 229, 231–232, 286–287
  Strong 226
Quine's Master Argument 183
Quine's Regress Argument 183
Quine's Super-task Argument 182

Reference 213
Rejection 30–32, 33–37
Relative Interpretability 108, 145, 225
Robinson Arithmetic 107–108, 148, 154, 205, 263, 274, 295, 302
Role Priority
  Name 217
  Predicate 217
  Quantifier 216
Rule-following 25–29, 40–43, 43–47
Russell's Paradox 284–285

Semantic Counterparts 137–141
Set Theory
  Naïve 284–288
  ZF 218, 272–273, 285, 287–288, 290, 297–299
Simple Sam 160
Slogans 321–325

Social Conventions 21–23
Stipulations 23–25
Structural Rules 69–72, 81–82
Super-tasks 182, 253, 264–265, 267–270

T-Schema 108
Tonk 93–94, 125–137
Translation Mistake 94, 125–131, 167–168, 173, 305, 315, 328
True Arithmetic 263, 269, 270, 271, 274–275, 278, 299
Truth
  Logical 95–102
  Mathematical 276–281
Truthmaking 178–179

Understanding 86–87, 279–280
Unrestricted Inferentialism 63–72

Validity
  Bilateralist 82, 98, 122–124
  Strong 60–61, 69–70, 101, 154, 236
Variance Theory 286
Venusians 268

Wug test
  Modified 27
  Original 26

ZF-Rules 272–273