

A PLEA
FOR
NATURAL
PHILOSOPHY

And Other Essays

PENELOPE MADDY

A PLEA FOR NATURAL PHILOSOPHY AND OTHER ESSAYS

A PLEA FOR NATURAL PHILOSOPHY
AND OTHER ESSAYS

Penelope Maddy

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 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 2022

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 Control Number: 2021947727
ISBN 978-0-19-750885-5

DOI: 10.1093/oso/9780197508855.001.0001

1 3 5 7 9 8 6 4 2

Printed by Integrated Books International, United States of America

For the Pals

CONTENTS

<i>Acknowledgments</i>	<i>ix</i>
Introduction	1
I. Method	
1. A Plea for Natural Philosophy	13
2. On the Question of Realism	49
II. Skepticism	
3. Hume and Reid	93
4. Moore's Hands	135
5. Wittgenstein on Hinges	148
III. Logic and language	
6. A Note on Truth and Reference	173
7. The Philosophy of Logic	188
8. A Second Philosophy of Logic	216
IV. Mathematics	
9. Psychology and the a priori Sciences	237
10. Do Numbers Exist?	253
11. Enhanced If-thenism	262
<i>References</i>	<i>295</i>
<i>Index</i>	<i>309</i>

ACKNOWLEDGMENTS

Many of my intellectual debts are acknowledged in footnotes to the individual papers—I'm most grateful to all those named there. Special thanks to Sorin Bangu, Adam Chin, Jeremy Heis, Alysha Kassam, Juliette Kennedy, Charles Leitz, Patricia Marino, Toby Meadows, Chris Mitsch, Stella Moon, Jeffrey Schatz, Evan Sommers, and Mark Wilson, for philosophical discussions and for their company during these strange times. Thanks also to Leitz and Mitsch again for editorial assistance, to Peter Ohlin for his support, and to David Malament, as ever, for everything.

Essays reprinted here—##3, 4, 7–10—appear with the kind permission of their original publishers, acknowledged below. Essays ##1, 2, 5, 6, and 11 are new to this volume.

Essay #3, 'Hume and Reid', appeared in *Analytic Philosophy* 52 (2011), pp. 2–34, under the title 'Naturalism and common sense'.

Essay #4, 'Moore's hands', is adapted from §III.2 of *What Do Philosophers Do? Skepticism and the Practice of Philosophy* (Oxford: Oxford University Press, 2014).

Essay #7, 'The philosophy of logic', appeared in the *Bulletin of Symbolic Logic* 18 (2012), pp. 481–504.

Essay #8, 'A second philosophy of logic', appeared in Penelope Rush, ed., *The Metaphysics of Logic*, (Cambridge: Cambridge University Press, 2014), pp. 93–108.

ACKNOWLEDGMENTS

Essay #9, 'Psychology and the a priori sciences', appeared in Sorin Bangu, ed., *Naturalizing Logico-Mathematical Knowledge: Approaches from Psychology and Cognitive Science* (New York: Routledge, 2018), pp. 15–29.

Essay #10, 'Do numbers exist?', appeared in in Gideon Rosen et al., eds., *Norton Introduction to Philosophy*, second edition, (New York: Norton, 2018), pp. 485–491.

Introduction

The essays in this collection, some old, some new, were all composed between 2010 and 2020. They all touch on Second Philosophy in one way or another, though this may not always be obvious, so I should begin with a few words about what that approach is supposed to be. Within the limits of contemporary taxonomies, Second Philosophy would have to count as a version of methodological naturalism, usually glossed as the assumption that the only reliable way of finding things out is to employ the methods of science.¹ The trouble with this, it seems to me, is that ‘scientific method’ is a moving target; any attempt to pin it down would artificially limit the crucial, ongoing processes of expansion and correction.² To get around this, I introduce a highly idealized inquirer called the Second Philosopher and characterize Second Philosophy as the answers she gives to traditionally philosophical questions.³ So how, then, do I describe her?

Most often, I’ve presented her as a perfectly straightforward inquirer: she begins with ordinary perceptual beliefs, gradually proceeds to generalizations and deliberate experimentation, eventually developing theories and various techniques of confirmation and disconfirmation, always circling back to critique and improve her body of belief and her store of methods as she goes. So far, so commonplace, but she’s also to be understood as limitless in curiosity, energy, and ingenuity, in time and resources, out to explore every aspect of

1. The contrast is with ontological or metaphysical naturalism, which claims that everything is physical, or material, or spatiotemporal, or some such thing. This sort of preemptive decision is not part of Second Philosophy, as the resulting position in the philosophy of mathematics demonstrates.

2. This is to take a stand in the old debate over the feasibility of a ‘demarcation criterion.’

3. In [2017a], the Second Philosopher is called the Plain Inquirer, to emphasize that she’s a more curious and meticulous extension of Austin’s Plain Man.

the world and our place in it. Of course, I've understood this to be no more than a suggestive sketch; more detail would emerge as she confronts various philosophical challenges and investigates various traditionally philosophical questions.

Perhaps inevitably, this sketchy and roundabout approach wasn't entirely successful. So, for example, I'm asked: with this emphasis on sense experience, aren't you just reviving empiricism? Well, the Second Philosopher's methods *are* broadly empirical, but there's no tabula rasa, no ideas, no sense data; the perceptual beliefs with which she begins are ordinary ones like 'there's a chair in front of me' when I see it there in good light. Then doesn't the Second Philosopher come with some rich cognitive mechanisms already in place? Well, yes, much as we all do, as the developmental literature shows, but of course she circles back and examines the reliability, quirks, and susceptibility to error in our various sensory systems and basic belief-forming mechanisms. For that matter, as her generalizing and theorizing show, she also comes equipped with inductive and deductive principles, not to mention memory and introspection, all of which she also eventually analyzes and assesses. Then what about testimony?—is she predisposed to trust others and only later moved to examine this predisposition critically or does she slowly build her trust based on track record arguments? Well, really, either way looks fine to me.⁴ Isn't she's too individualistic; isn't it the scientific community as a whole that comes to adopt a shared theory? Well, I've actually described her in places as recapitulating the history of science, for example, in the evolution of our attitudes toward mathematics in application,⁵ so sure, if it helps to think of her as an embodiment of the scientific community, then that works, too.

No doubt my interlocuters here have been frustrated by my imprecision and indifference,⁶ but my own nagging feeling has been that there's something wrong with this whole discussion. The analogy is ungenerous, but

4. For rhetorical purposes in [2011b], pp. 84–85, I characterize a case of testimony this way: a second-philosophical primatologist consults a second-philosophical botanist for help with her chimps' lack of appetite; when the new feed works, she wants to understand why, then thoroughly investigates the botanist's methods and conclusions, start to finish.

5. See [2008] or [2011b], chapter 1.

6. In yet another example, in [2011b], pp. 104–111, I consider a top-down characterization of the Second Philosopher: instead of starting with perceptual beliefs and working upward, imagine her starting with

it's the best I can do: I imagine trying to call a dog's attention to something by pointing at it, only to have him study the tip of my finger. The device of the Second Philosopher is intended to motivate a series of reactions to and investigations of traditional philosophical questions; those reactions and investigations themselves are the intended focus. Some might find them surprisingly amenable; some sympathetic to second-philosophical method may think I've misapplied it and propose a better way; many are likely to find second-philosophical theories just not the right sort of thing at all, in which case there's an implicit challenge to explain exactly why not. For this purpose, the character of the Second Philosopher has to be suggestive enough to get the project going, but any further attention to the detail of her characterization is a distraction.⁷

I think this is what Barry Stroud has in mind when he speaks of 'the transparency of naturalism':

Maddy calls the kind of investigation she has in mind 'Second Philosophy'. I find this congenial . . . since that name means nothing at all until Maddy specifies what she means by it. To understand her, we have to look right through that label to see what she means to draw our attention to, just as I think we have to do with the term 'naturalism' itself. (Stroud [2009b], p. 253)

As for the Second Philosopher's method, apart from beginning without the common assumption that philosophical questions require a priori treatment, it comes to this:

we should just try to find the best ways we can to understand the things that puzzle us, and proceed carefully. If this is naturalism, it sounds like a good idea. But it is not a definable method . . . we are left with nothing much more determinate than the idea of careful, informed investigation,

the full variety of common sense, natural science, mathematics, and more, and gradually whittling away the parts that aren't well supported. I argue that this would make no difference for her ultimate stand on the ontology and epistemology of mathematics.

7. For that matter, asking for a precise definition of the Second Philosopher can be a roundabout way of asking for a demarcation criterion: 'science' would then be 'that which this inquirer generates.'

or perhaps just responsible inquiry. That is not nothing to be left with. It is all we've got. But there is no need to decide in advance how inquiry into whatever interests us must, or must not, proceed. (Ibid., p. 255)

Stroud has put the point eloquently: to understand what Second Philosophy is, who the Second Philosopher is, look to which questions she takes up and how she goes about investigating them. This is the sense in which all the essays in this volume are exercises in Second Philosophy.⁸

The collection consists of eleven essays, roughly five new—##1, 2, 5, 6, and 11—and six previously published—##3, 4, 7–10. (I say ‘roughly’ because an extract from Essay #1 has been published, and Essay #4 is somewhat reworked from its appearance in [2017a].) Let me comment briefly on each of these in turn.

The two new essays in Part I aim to illuminate the roots and methods of Second Philosophy. The title essay, ‘A plea for natural philosophy’, is the long version from which I intended to extract my APA presidential address. Unfortunately, the April 2020 meeting of the Pacific Division was abruptly canceled in the early, chaotic days of the pandemic, and in the end all that came of it was a shortened text included in the *Proceedings* ([2020]). Before all that, though, when starting out, I imagined that the occasion would grant me some license, an excuse to let my hair down. That permission in hand, I sketched a broad, perhaps somewhat irresponsible, history of the changing relations between what we tend to think of as ‘science’ and what we tend to think of as ‘philosophy’—from the early moderns (especially, Descartes, Boyle, Locke, Newton, Berkeley, and Reid) to the present. These 17th- and 18th-century thinkers seamlessly combined what we nowadays regard as two separate disciplines, an approach that was largely lost in the wake of Kant’s Copernican Revolution and his a priori critical inquiry. Still, I argue, the early modern notions and concerns live on in what we now think of as interdisciplinary work—in the foundations of mathematics, physics, biology, vision theory, consciousness studies, and more. The term ‘Second Philosophy’ occurs only in a footnote, but my intention is to cast early modern ‘Natural Philosophy’ as exemplary and ‘Second Philosophy’ as its rightful heir.

8. Also, the bulk of [2007], [2011b], and [2017a].

The second essay, ‘On the question of realism,’ picks up the fundamental issue of scientific ontology from earlier efforts ([2001], [2007], IV.1 and IV.5). It begins by updating the Second Philosopher’s reactions to Bas van Fraassen’s constructive empiricism in light of some of his more recent writings. Once again, I suggest that his debate with the ‘realist’ hinges on transcendental considerations that are methodologically removed from and irrelevant to her ordinary scientific efforts; from her perspective, the whole discussion appears mysterious and unmotivated. In sharp contrast, Kyle Stanford’s epistemological instrumentalism mounts a historicist critique internal to her empirical inquiry. I do my best to follow the contours of his argumentation as it’s developed over the years and eventually conclude that if the typical labels can be set aside, our apparent disagreement may come to a mere difference of emphasis. Still, this would mean that the ‘realists’ with whom Stanford contends, like van Fraassen’s opponents, do not include the Second Philosopher who simply believes in atoms on the basis of the famous Brownian Motion experiments of Jean Perrin in the early 20th century and subsequent evidence from cloud chambers, electron microscopes, and so on. This strikes me as an odd result! If her belief in atoms doesn’t qualify her as a ‘realist,’ perhaps this philosophical debate has slipped its moorings.

In Part I of [2007], the Second Philosopher is confronted with radical skepticism about our knowledge of an external world, including its appearance in the thought of historical figures like Descartes and Hume. The three essays in Part II continue in this line. Essay #3, first published in 2011, appears here with only slight adjustments under a new title, ‘Hume and Reid’. Observers of the early modern period often identify Hume as the first ‘naturalist’, determined to bring empirical methods to the study of human nature. His supposed atheism, for example, sets him in contrast to the believers—Descartes, Locke, the Bishop Berkeley, and the good pastor Reid—and establishes him as fully rejecting appeals to the supernatural.⁹ I argue that despite this, it isn’t Hume at all, but rather Reid who first articulated a version of naturalism and stands as direct ancestor to the Second Philosopher.

9. In the paper quoted above (Stroud [2009b], pp. 241–242), Stroud sees Hume in this way. Though we disagree about Hume, we both take the contemporary mainstream focus on a priori methods to be ‘largely the legacy of Kant’ (p. 244, see Essay #1).

Neither his faith nor his dualism prevents his purely empirical inquiry into the human mind. This complements Reid's leading role in Essay #1.

The next two essays shift the historical focus from early modern to early analytic philosophy, in particular, to Moore and Wittgenstein. 'Moore's hands' adapts the discussion of 'Proof of an external world' (Moore [1939]) embedded in [2017a] to a stand-alone piece on Moore. This sets the stage for the new essay, 'Wittgenstein on hinges'. Wittgenstein's final writings on skepticism are collected in *On Certainty* (Wittgenstein [1949/51]), a difficult book to interpret for various reasons, beginning with the fact that it isn't a book at all, but a collection of notes assembled by editors. We know from scholarly work on his published writings that notes like these typically undergo a long process of revising and reordering before Wittgenstein is even semi-satisfied with them, and those in *On Certainty* are all just first drafts. To reduce this difficulty as much as possible, I work with newly revealed biographical information about the period of Wittgenstein's last writings (see Rogers [2010], [2011]), then attempt to extrapolate from interpretations of his published works, the *Tractatus* and the *Philosophical Investigations*. The resulting interpretation stands in opposition to contemporary views of the book as the ur-text of so-called 'hinge epistemology' and instead hews closer to Moore (as understood in 'Moore's hands') than Wittgenstein's rhetoric would suggest.

In the end, though the goal of these two essays is to illuminate the thinking of Moore and Wittgenstein, there is an unspoken theme as well: though Second Philosophy isn't mentioned,¹⁰ the position these two thinkers are portrayed as favoring is akin to the one further developed in [2017a]. Together, the essays of this group suggest that early versions of second-philosophical thinking can be traced in these three historical greats.

Part III, on Logic and language, begins with 'A note on truth and reference'. My original plan for this piece was to extract the characteristically second-philosophical account of truth and reference from Part II of [2007], where it was so thoroughly embedded in discussions of various other themes as to be somewhat lost in the weeds.¹¹ I imagined that this would be straightforward,

10. Apart from a footnote or two in 'Moore's hands'.

11. E.g., the backstory in Tarski, Quine, Field, and Leeds; fine points about the configuration of the debate itself (e.g., the role of physicalism); contrasts with minimalist thinkers like Crispin Wright; and so on.

but getting down to it, I realized my views on these matters had shifted. The result was an entirely new essay that develops a story less distinctly deflationist and more Austinian in spirit—at least as I understand Austin in his ‘Truth’ (Austin [1950]) and ‘A plea for excuses’ (Austin [1956a]).

The other two essays in this group concern the philosophy of logic. ‘The philosophy of logic’ is my retiring presidential address to the Association for Symbolic Logic. The ASL was founded in the 1930s by figures not easily classified as either philosophers or mathematicians, but in the decades since, mathematical logic, philosophical logic, and the philosophy of mathematics and logic have developed into distinct disciplines. Though all three remain within the purview of the Association, mathematical logic has so outpaced the others in sheer numbers that the membership today consists largely of mathematicians. For various historical and sociological reasons, the philosophy of mathematics has continued to be represented, nonetheless, but the fundamental issues in the philosophy of logic—what is logical truth and how do we know it?—have been less in evidence. The goal of my address was to describe the range of positions on offer in the philosophy of logic to non-philosophers already aware of the three great schools in philosophy of mathematics in the early 20th century: Formalism, Intuitionism, Platonism. At the very end, I use this taxonomy to locate and sketch the second-philosophical account of logic in Part III of [2007].

That extended account of the second-philosophical view in [2017] derives it in stages from a reading of Kant on the status of logic. The strategy behind this approach was to ease the reader into a frame of mind where the structuring of the world into objects with properties isn’t so obvious as to be nearly invisible, as it is, for example, when we hardly notice that a ‘possible world’ is taken to resemble a model of first-order model theory. By the time Kant’s position is ‘naturalized’ into the second-philosophical one, the possibility of this worldly structuring is modernized (for example, to a post-Fregean objects-standing-in-relations) and de-idealized to an objective fact about our world—but the advantage to having started with Kant is that this objective fact is more easily seen to be contingent. Still, despite these advantages, the Kantian approach led some readers to expect or discern some idealistic or transcendental residue. When Penelope Rush invited me to contribute an overview of the Second Philosopher’s position to her

volume on *The Metaphysics of Logic*, it seemed the perfect opportunity to demonstrate that it could be described and defended without any reference to Kant. This also opened the way to engage with those who question the scientific reality of ordinary, medium-sized objects—think of Eddington’s two tables, Sellars’s pink ice cube, or Ladyman and Ross’s reductionism. ‘A second philosophy of logic’ is this pocket-sized summary.

The final group, Part IV, consists of three essays in the philosophy of mathematics, beginning with two in the philosophy of arithmetic. ‘Psychology and the a priori sciences’ picks up the developmental psychology central to the Second Philosopher’s epistemology of logic, elaborating some points on representation and the concept of object, then extends the picture to examine psychology’s role in a second-philosophical account of elementary arithmetic claims ($2 + 2 = 4$) and of full Peano arithmetic. While the account of logic purports to show that rudimentary logic is correct simply due to the contingent structure of our macro-world, this robust truth extends into arithmetic only as far as its elementary claims; the study of the ‘...’, of a full omega sequence, of Peano arithmetic, is linked not to the world but to a human conceptual structure related to the language-learning device.¹² The account of logic teaches us both that logic is contingent and that, given our cognitive apparatus, it will seem necessary to us; the account of arithmetic should leave us somewhat less confident that our notion of an omega sequence is coherent, unique, and determinate.

‘Do numbers exist?’ is an outlier here, written for the second edition of the *Norton Introduction to Philosophy*, a textbook addressed to beginners. As a young math major, I often heard tell of Hilbert’s remark that one should be able to explain one’s mathematics to ‘the man on the street’,¹³ so I welcomed the challenge to present the second-philosophical view of arithmetic to such

12. And, by the way, higher set theory, with its extrinsic considerations in pursuit of mathematical depth, is another matter entirely (see [2014b], [2011b]). We shouldn’t expect a single philosophical treatment to cover all areas of mathematics.

13. The quotation comes from Hilbert’s famous address to the International Congress of Mathematicians in 1900, in which he attributes it to ‘an old French mathematician’: ‘A mathematical theory is not to be considered complete until you have made it so clear that you can explain it to the first man whom you meet on the street’. See Barrow-Green and Siegmund-Schultze [2016] for an investigation of Hilbert’s sources.

an audience. With an assist from Gideon Rosen, one of the editors, I hope to have shown that this can be done!

The final essay, ‘Enhanced if-thenism’, is new, an attempt to revive a view familiar from Philosophy of Mathematics 101 and, it might be noted, from conversations with mathematicians eager to avoid philosophical quicksand. The idea that all metaphysical and epistemological questions about mathematics could be reduced to those about the logical connection between the ‘if’ and the ‘then’ will probably never lose its superficial appeal. My goal in this essay is to demonstrate that the view is more attractive and resilient than traditional objections have made it seem, that it can meet these challenges and more given just one enhancement: attention must be paid to what guides and constrains the choice of what’s put in the ‘if’, because this isn’t arbitrary. In the case of set theory, this enhancement comes to an account of the rationality of axiom choice, long my leading hobbyhorse, but the discussion in this essay is more general, extending to concept-formation across the range of mathematics.

I hope this gives some sense of what’s in the individual essays and how they fit together. Some trace historical antecedents of Second Philosophy; some follow the dictates of transparency and get on with doing it. This often looks for all the world like history, epistemology, philosophy of science, philosophy of language, philosophy of logic, philosophy of mathematics, and maybe even metaphysics, but perhaps collecting these essays side by side helps expose the fact that they all ultimately spring from one particular point of view.

PART I

METHOD

A Plea for Natural Philosophy

‘A plea’ in my title is obviously a nod to the great J. L. Austin and his plea for excuses (Austin [1956a])—a transparent ploy to align myself with my betters.¹ But apart from this trace of vanity, I hope the comparison might also help orient the project of this talk. So let me begin there.

Part of Austin’s genius was the way he employed and juggled a number of different philosophical methods—including the common sense² embodied in his Plain Man and a brand of therapeutic philosophy quite different from Wittgenstein’s—but he’s most famously associated with, perhaps the leading figure of, the now much-maligned school of Ordinary Language Philosophy.³ Unlike some less discerning advocates of that approach, Austin was keen to emphasize that an examination of ‘what we should say when’ isn’t appropriate in all cases:

Such a method . . . is plainly preferable to investigate a field where ordinary language is rich and subtle, as it is in the pressing practical matter of Excuses, but certainly not in the matter of, say, Time. (Austin [1956a], p. 182)

In cases like Time, ordinary language ‘has not been fed from the resources of the microscope and its successors’ and ‘our interests are more extensive or intellectual than the ordinary’ (ibid., p. 185).⁴ The point of his plea then is to

1. Just how much better is illustrated by how much more cleverly Austin crafted his own such nod: *Sense and Sensibilia* (Austin [1962]).

2. Descendent of G. E. Moore, and Thomas Reid before him.

3. All three of Austin’s methods are discussed in [2017a].

4. Indeed, even in cases where attention to ordinary language is appropriate, ‘modern scientists have been able, it seems to me, to reveal its inadequacy at numerous points, if only because they have had access to

recommend a particular topic—excuses—if one wishes to apply a particular method—ordinary language analysis.

My plan here is to reverse the polarity of this line of thought: rather than starting with a method and recommending an appropriate topic, instead starting with a topic and recommending an appropriate method. To speak in vague and general terms, we might say that the topic in question is the world and our place in it. Admittedly, this rough characterization isn't much help; it's not immediately clear what it leaves out. For example, I take 'our place in it' to include the fact that humans engage in mathematics, so that asking what kind of discourse math is, what governs its practice, to what, if anything, it's answerable, would all count as part of the overall project.⁵ In what follows, I focus on particular subtopics that clearly fall within the scope of 'the world and our place in it'; toward the end, I indicate a few cherished items that may fall outside it. But first, I'd like to take up the other term in the relation, the proposed method, that is, 'natural philosophy'.

1. THE METHOD

The story begins in 17th-century Europe. The historical literature on what's known as the Scientific Revolution is deep, subtle, and contentious, including an important debate over the extent to which there even was such a thing.⁶ However that may be, the 'natural philosophy' I have in mind arose in the early modern period in the wake of this purported change. The preceding Scholasticism, descended from Aristotle, comprised theology, morality, and alchemy, in addition to something then termed 'natural philosophy'. The

more comprehensive data and have studied them with more catholic and dispassionate interest than the ordinary man . . . has had occasion to do' (Austin [1956a], p. 203).

5. There would be analogous questions for any human undertaking—including ethics, theology, and astrology—but answering them takes a different course for mathematics, because of its special role in our investigations of the world (see, e.g., [2011b]).

6. See, e.g., Shapin [1996], p. 1: 'There is no such thing as the Scientific Revolution, and this is a book about it'; and Dear [2009], p. 2: 'For all that it was exaggerated and self-congratulatory, the [18th-century] idea that there was a fundamental difference between medieval learning and the new learning brought about by the recent 'revolution' contains an important insight'. For an overview of the controversy, see Henry [2008], chapter 1.

basis for this last study was generalities of common knowledge⁷—that the sun rises, that heavy bodies fall, that we perceive the features of ordinary objects—and its goal was to understand them. Scholastics are sometimes ridiculed as positing ‘dormitive virtue’ to explain opium’s tendency to put people to sleep, but their accounts typically have more moving parts. So, for example, I see that an object has a certain sensible quality because that quality in the object alters the medium between the object and my eye, producing a ‘species’ in the medium that shares the form but not the matter of that quality; the species is then propagated through the medium, eventually reaching and being realized in my eye.⁸ This Scholastic natural philosophy is a contemplative discipline, proceeding from commonplaces (we see the properties of objects) by pure thought (since the object is distant from the eye, there must be some intervening medium,⁹ and so on).

The early moderns dramatically altered both aspects of this Scholastic account. First, the empirical input was no longer commonplace generalities, but individual observations—not ‘what happens in nature’ but ‘what actually *happened* in nature.’¹⁰ Great care was taken to collect these observations as broadly, methodically, and responsibly as possible, a process that eventually came to include deliberate experimentation. This initial phase is sometimes called ‘natural history’, to distinguish it from the second undertaking—theorizing about the causes of these phenomena—sometimes termed ‘natural philosophy’ proper.¹¹

Of course the Scholastics had so theorized themselves, but in a second departure, the new school had more demanding standards. Rather than positing one species for each sensory quality—flirting with the specter of

7. See, e.g., Dear [2009], p. 5, Shapin [1996], pp. 81–82.

8. Cohen [1982] explores the sense in which the eye takes on the species. Pasnau [2017], chapter 4, points out that even when the eye or the mind ‘takes on’ the form of the perceived quality, the species is not itself perceived and can’t be introspected even if we try (it’s posited as a theoretical entity). The early modern’s ‘idea’ plays a quite different (perhaps less plausible?) role as the object of perception.

9. Pasnau [2017], pp. 71–72, especially note 1, traces the influence of this ‘no action at a distance’ line of thought in both the Scholastics and the early moderns.

10. See Shapin [1996], pp. 85–90 (the quotations come from pp. 89–90), and Dear [2009], pp. 6–7. Shapin draws attention to ‘the ‘cabinets of curiosities’ then fashionable in gentlemanly circles throughout Europe. These cabinets eloquently testified to nature’s particularity and startling variety. Stuffed with rarities and oddities, such cabinets were *accessible* proof that there were indeed more things in heaven and earth than were dreamed of in traditional philosophies’ (p. 90).

11. See Shapin [1996], p. 85.

dormitive virtue—the explanatory basis was restricted: everything was to be explained in terms of tiny particles with a small range of features, at the beginning, of course, those of Corpuscular Mechanism (size, shape, motion, number . . .). As Bacon had counseled, this was no longer a project for the study; the new inquirer was to observe broadly, to experiment, to generalize, predict, and confirm.¹² In this way, the story goes, the old ‘natural philosophy’ was replaced by the new ‘natural philosophy’, and the new ‘natural philosophy’, the story continues, is what we now call ‘science’. So, for example, we shouldn’t be surprised, don’t you know, that what we regard as one of history’s most brilliant ‘scientific’ works, Newton’s *Principia*, bears the full title *Mathematical Principles of Natural Philosophy*.

So far so good, but then one can hardly help wondering: where in this picture is what we now call ‘philosophy’?¹³ The mystery deepens when we consider the canonical lists of ‘natural philosophers’ of this new sort: Galileo, Boyle, and Newton, of course, but Descartes was as central as Boyle to the development of Corpuscular Mechanism. For that matter, Descartes actively dissected eyes to observe the retinal image,¹⁴ traced nerves for his work on the physiology of distance perception,¹⁵ and propounded laws of motion.¹⁶ Even his quintessential ‘philosophical’ text (in our terms), the *Meditations*,¹⁷

12. See, e.g., Gaukroger [2002]. This schematic telling suppresses various disagreements on fundamental points between individual early moderns, over certainty vs. fallibilism (see, e.g., Shapin [1996], pp. 101–103), over the role of experimentation (see, e.g., *ibid.*, pp. 82–84), over the nature of explanation (see, e.g., Gaukroger [2014], p. 20, on ‘horizontal’ vs. ‘vertical’ explanations (more on this distinction below)).

13. Lüthy [2000], pp. 175–177, describes how the early moderns themselves debated terminological issues in the vicinity of ‘natural philosophy’.

14. See *Optics*, Descartes [1632], pp. 91–93: ‘the objects we look at do imprint very perfect images on the back of our eyes . . . you will be . . . certain of this if, taking the eye of a newly deceased man, or, for want of that, of an ox or some other large animal, you carefully cut through to the back the three membranes which enclose it . . . Then, having covered it over with some white body thin enough to let the daylight pass through it, as for example with a piece of paper or with an eggshell . . . place this eye . . . in such a manner so that it has its front . . . turned toward some location where there are various objects . . . illuminated by the sun . . . if you look at the white body . . . you will see, not perhaps without admiration and pleasure, a picture which will represent in natural perspective all the objects . . . outside it’.

15. In *Treatise of Man*, Descartes [1633].

16. In *Principles of Philosophy*, Descartes [1644].

17. Descartes [1641a].

was aimed at opening the way for his physics, as Descartes explained in his famous letter to Mersenne.¹⁸ So Descartes is typically included as well.

But now what about Locke? Praising ‘Master-Builders’ like Boyle, Huygens, and ‘the incomparable Mr. Newton,’ he famously describes himself as an ‘Under-Labourer . . . clearing the Ground a little, and removing some of the Rubbish’ in service to these greats.¹⁹ Presumably what needs clearing away is Scholasticism (in concert with Descartes) and some Cartesianism (innate ideas, for example), but despite his modesty, Locke’s thinking goes considerably beyond this. His early work in medicine and physiology, his collaborations with Boyle, his study of seeds and principles of generation all fall under the broad umbrella of natural philosophy, often in its natural historical branch.²⁰ However he thinks of himself, the fact remains that he both helped formulate Corpuscular Mechanism and—with admirable intellectual integrity—brought attention to its points of tension: its apparent inability to explain object cohesion, contact forces, and ultimately Newton’s gravitational force acting at a distance.²¹ He engaged directly with Newton on the nature of ‘body’—Newton’s striking exploration of the notion²² apparently came to him in conversation with Locke²³—and evidence suggests that the third of Newton’s famous ‘Rules for the study of natural philosophy’²⁴ was influenced by Locke’s argument for the inseparability of primary qualities.²⁵ Lisa Downing, astute analyst of the period, places Locke squarely in ‘the historical transition from Boylean mechanism to Newtonianism.’²⁶

18. Descartes [1641b], p. 173: ‘I may tell you, between ourselves, that these six Meditations contain all the foundations of my physics . . . I hope that readers will gradually get used to my principles, and recognize their truth, before they notice that they destroy the principles of Aristotle.’

19. Locke [1689], ‘Epistle to the Reader’, pp. 9–10. Jacovides [2017], pp. 33–34, notes a remarkable parallel between Locke’s ‘under-labourer’ remark and an earlier passage from Boyle, who describes himself as ‘a drudge’ and ‘underbuilder’ for his natural philosophical betters.

20. Jacovides [2017] gives a comprehensive account of Locke’s ‘scientific’ training and activities, and their interconnection with his ‘philosophy’. See also Anstey [2011].

21. Locke’s criticisms of Corpuscular Mechanism are documented in the influential Wilson [1979]. (Cf. footnote 26.)

22. In the much-discussed ‘De gravitatione et aequipondio fluidorum’ (Newton [1962], pp. 121–156).

23. See Downing [2014], p. 105.

24. Newton [1687], p. 795. Rule 3 appears in the second edition of 1713. (See footnote 67.)

25. Locke [1689], II.8.9. See Jacovides [2017], pp. 89–91, and footnote 67 below.

26. Downing [2014], p. 99. In the paper cited in footnote 21, Wilson takes Locke to suppose that God simply imbues objects with extra (‘superadded’) powers that don’t spring from their corpuscular makeup, which conflicts with his assumption that all qualities of the object derive from that fundamental

Easing into the 18th century, we find Berkeley, too, directly engaged with Newton, both in his cogent critique of the calculus²⁷ and in his innovative instrumentalist account of forces.²⁸ Most influential of all was his re-orientation of the theory of vision from optics to psychology in *New Theory of Vision*.²⁹ Just a few decades later, Reid makes his Newtonian method explicit:

The man who first discovered that cold freezes water, and that heat turns it into a vapour, proceeded on the same general principles, and in the same method, by which Newton discovered the law of gravitation and the properties of light. His *regulae philosophandi* are maxims of common sense, and are practised every day in common life; and he who philosophizes by other rules, either concerning the material system, or concerning the mind, mistakes his aim. (Reid [1764], I.1, p. 12)

True to his word, Reid's work on external world skepticism comes embedded in a close empirical study of the physiology and psychology of vision and of the senses more broadly.³⁰

Perhaps a set of principles could be developed to disentangle what we call 'philosophy' or 'philosophers' in all this from what we call 'science' or 'scientists', but fair judgment suggests that to do so would be untrue to the spirit of the time. Margaret Wilson, who has done as much as anyone to illuminate the thought of this period, describes 'the seventeenth-century framework' as one 'in which what we now regard as two distinct modes

structure. Downing [1998] disagrees. Her Locke does believe that all qualities of the object derive from some fundamental features, but he takes his examples to show that these aren't exhausted by its corpuscular makeup: God doesn't tack on extras with no basis in the fundamental features; he's given objects a more complex array of such features than Corpuscular Mechanism allows. Since Locke also believes that the corpuscular account is the only one fully intelligible to us—because of its close connection with sense experience—this is a deeply discouraging realization. I return to this theme below.

27. In *The Analyst*, Berkeley [1734].

28. In *De Motu*, Berkeley [1721]. See Downing [2005].

29. Berkeley [1709].

30. Of course, Hume [1739] also set out to apply 'the experimental method of reasoning'—with a clever analogy positioning himself as 'the Newton of the science of man' (Stroud [1977], p. 5)—but the naturalistic credentials of his approach are undermined when he insists that 'the science of man is the only solid foundation for the other sciences' (ibid., introduction, paragraph 7), that it promises to change and improve them but not vice versa. See Essay #3 for a comparison of Hume and Reid on method.

of intellectual activity were often seamlessly combined' (Wilson [1992], p. 481).³¹ This seamless combination is what I have in mind with the term Natural Philosophy: an empirical study, beginning with careful observation of phenomena ('natural history'), moving on to deliberate experimentation, theory formation, and testing ('natural philosophy' in the stricter sense), always assessing and re-assessing its methods as it goes—a mode of inquiry that makes no principled distinction between questions we tend to think of as 'scientific' and questions we tend to think of as 'philosophical'.³² This is the method I'm out to explore.

2. THE TOPIC

With this sense of the Natural Philosophical method, we're returned to the question of the relevant topic. Just now, Reid identified the suitable topics for his Natural Philosophical method as 'the material system' and 'the mind'—not far from my own rough description, 'the world and our place in it'. It's worth noting that this characterization needn't involve any form of physicalism or materialism or reductionism. Reid, for example, was a confirmed dualist, but he was happy to engage in empirical study of the mental based on introspection³³ and the reports of others, either in ordinary life or as subjects in experimental conditions.³⁴ Notice also that this approach to

31. Others are less delicate: Lüthy [2000], p. 174, describes attempts to separate 'philosophy' from 'science' in the early modern period as 'a laughable farce' and quotes another historian, George Molland, as declaring 'for this period, to distinguish rigidly between philosophy and science would be grossly anachronistic'. Interpreters of Locke and Berkeley tend to agree. Speaking of Locke and Boyle, Winkler [1989], p. 255, characterizes some interpretations of the former as suggesting 'that scientists and philosophers operated within clearly defined spheres of authority . . . But there was no such division of labour'. Speaking primarily of Berkeley, Atherton [1991], p. 67, writes: 'Berkeley was concerned with the debates that were prevalent and important to the physics of the time . . . both Locke and Berkeley ought to be seen as taking . . . a more active role in the development of scientific theory than they are generally given credit for . . . they saw themselves as advancing considerations which would have the result of reshaping prevailing views of ongoing scientific endeavor'.

32. Readers of [2007] (or [2011b] or [2014b]) or [2017a] will recognize the Second Philosopher or the Plain Inquirer, respectively, viewed from yet another angle.

33. Though he insisted that introspection is as fallible as any of our other faculties (see Reid [1785], I.6, pp. 59–64).

34. See Essay #3, especially footnote 29. Hatfield [1995], p. 188, locates an empirical approach to mental substance as a common feature of 18th-century thought about the mind: 'if one believes that immaterial entities exist and that some of them inhabit human bodies, it makes good sense to seek to

psycho-physical laws represents a methodological shift: for the Corpuscular Mechanist, explanation is, as you might say, vertical—the macro-qualities of an object are explained in terms of its microstructure—but these are horizontal explanations—some macro-qualities are explained in terms of others.³⁵ As Natural Philosophical method developed, in this and other ways, over the course of the early modern period, the work of Galileo, Boyle, and Newton clearly demonstrated its effectiveness for study of ‘the material’, and Locke, Berkeley, and Reid did the same for ‘the mind’, leading the way to empirical psychology.

This bird’s-eye perspective certainly suggests that the Natural Philosophical method is well-suited to the topic, but I think we can trace the connection more effectively by narrowing to a familiar subtopic and engaging in a bit of compare-and-contrast with ‘philosophy’ as we think of it today. Among the many questions that can be and are posed about the world and our place in it, one of the most characteristically ‘philosophical’ (as we now use the term) concerns the relationship between the two: our knowledge of that world, that is, epistemology. As a way into this subject, I hope I may be forgiven a touch of autobiography . . .

As an eager young math major, I was baffled by the question of mathematical knowledge and gradually stumbled unawares into philosophy. While gaining my feet, I naively assumed that the disciplinary specialization of epistemology would have something helpful to offer, but in fact, the main role it seemed to play was as a potential spoiler—for example, when Benacerraf deployed the then-popular causal theory of knowledge to pose a stark challenge to many forms of mathematical realism.³⁶ At the time, I turned for help

determine the powers and capacities of such substances empirically, by studying the manifestation of the mind in the behavior of others and in one’s own experience of mental phenomena.’

35. I borrow the vivid terminology of vertical and horizontal explanations from Gaukroger [2014]. As Gaukroger and others have noted (see also Downing [2002], pp. 348–349), Boyle’s actual explanations were horizontal, as were Locke’s, given our ignorance of corpuscular structure (Jacovides [2017], pp. 80–81), but for both, these were second-best. Berkeley wasn’t opposed to corpuscular theory (despite what his immaterialism might suggest; see Winkler [1989], pp. 238–275, Atherton [1991], Downing [2005]), but recognized no principled distinction between vertical and horizontal explanations; both are regularities or ‘physical causes’ (though the only true causal agents are minds) (see Winkler [1989], pp. 260–261). Despite his openness to corpuscular explanation, Berkeley was more closely aligned with Newton (Atherton [1991], p. 63, Downing [2005], p. 235). Gaukroger [2014] traces the path of horizontal explanation up through Newton.

36. Benacerraf [1973].

to empirical psychology instead,³⁷ but years later, when I strayed into metaphilosophy and adopted external world skepticism as a diagnostic instrument, it was this quintessentially epistemological challenge that led me back to its sources in Descartes and the early moderns. What struck me then—and this is the point of this autobiographical digression—was how fresh, how immediate these writings seemed! Unencumbered by centuries of philosophical theorizing, Descartes, Locke, Berkeley, and Reid addressed the problem of knowledge head-on, with breathtaking directness. Though what they were doing was undoubtedly epistemology, I couldn't imagine that any of these thinkers would much care about whether the agent in some intricately described situation should count as knowing—their concerns were more raw and immediate than that.³⁸

With this puzzling apparent mismatch in the back of my mind, I was delighted to come across Robert Pasnau's fascinating history of epistemology, *After Certainty*.³⁹ Noting the contemporary impression that giving an account of knowledge has always been the main job of epistemology, he writes:

To be sure, there was Plato in the *Meno* and the *Theaetetus*, and the casual follower of philosophy's history might be forgiven for supposing that his example carried forward more or less continuously until the present day. [In fact,] from Aristotle to the Middle Ages and well beyond, philosophers took an interest in carefully circumscribing one or another particular kind of cognitive grasp of reality—perception, imagination, assent, deduction, and so on—but showed little interest in defining the broad category of knowledge. That English contains this very general word of positive cognitive appraisal did not strike philosophers, even those who worked in English, as calling for any special definitional inquiry. (Pasnau [2017], p. 2)⁴⁰

37. See [1990].

38. At the time, analysis of the concept of knowledge was the leading goal of mainstream epistemology. More recently ([2017a], pp. 205–206), I noted this mismatch between early modern and more contemporary epistemology in much the same terms and was rightly criticized for implying that contemporary epistemologists remain largely engaged in this same project, as in the heyday of Gettierology (e.g., in Williamson [2017]). Of course this isn't true—see footnote 41.

39. Pasnau [2017].

40. In an extended note, Pasnau [2017], p. 141, is more pointed: "The casual assumption that the Platonic definitional quest informs the whole history of philosophy is so rampant that it might seem

Here Pasnau puts his finger on what had so impressed me in the early moderns: they weren't out to analyze a general concept of knowledge or to study a general state of knowing;⁴¹ their attention was on isolating and evaluating particular forms of 'cognitive grasp'. Their leading question wasn't 'what counts as (what is) knowledge?', but 'how do creatures like us manage to get reliable information about the world?'^{42,43} This is just the question I once asked about mathematics—how do we manage to get reliable information

unfair to single out individual scholars . . . The irresistibility of such 'ever since the dawn of philosophy' claims attests to our vulnerability to a cognitive shortcut we might call the heuristic of historical continuity: our propensity to take data points scattered across an historical timeline and to fill in the gaps on the assumption that things have always been so. As vision fills in for the blind spot where the optic nerve meets the retina, so [does] our historical imagination fill in the inevitable gaps in our knowledge with an assumption of homogeneous continuity. In the present case, then, even if all we know is that Plato wanted a definition of knowledge and that philosophers today similarly seek such a thing, those two data points alone are enough to foster the conclusion that this quest must have continued uninterrupted over millennia about which we are wholly uninformed.

41. I allude here to more recent developments in mainstream epistemology (recall footnote 38). In her [1994], Zagzebski argued that 'knowledge' can't be analyzed as the post-Gettier project supposed (i.e., by adding a condition to the justified-true-belief account); in his [2000], Williamson proposed that 'knowledge' be taken as primitive, that 'justification', 'belief', etc., be analyzed in terms of 'knowledge' rather than the other way round; and finally, in the wake of his [2007], Williamson turned his back on conceptual analysis entirely and simply theorized about knowledge itself ('To characterize knowledge as the most general factive state is not to engage in conceptual analysis . . . It is simply to theorize about knowledge' (Williamson [2009], p. 290)). The project of conceptual analysis has by no means disappeared (see Ichikawa and Steup [2018] for a recent survey), but 'knowledge-first epistemology' is now a highly influential alternative. While this represents a sea change in contemporary analytic epistemology, it hardly registers from the perspective of our 17th- and 18th-century Natural Philosophers: the topic is still 'knowledge', a level of generality that the early moderns didn't engage. (See also footnote 98.)

42. Here again Hume is an outlier, actually troubled by skeptical worries, unlike Galileo, Descartes, Boyle, Locke, Berkeley, or Reid. In contemporary terms, what I'm describing may sound like Quine's epistemology naturalized (Quine [1969]), but I would argue (have argued, see [2007], I.6) that Quine—for all his many virtues!—goes seriously astray in a number of ways. See footnote 73 for one example.

43. As it happens, this isn't Pasnau's assessment of the early modern epistemological project. He takes it to center on the notion of an epistemic ideal, with the leading question, not 'what counts as knowledge?' but 'what should we believe?' (Pasnau [2017], p. 8). My concern is that contemporary epistemologists could just as easily embrace this leading question, holding that the reason the 'what counts as knowledge?' is so important is that its answer tells us what we should believe. So I'm not convinced that Pasnau has correctly identified the early modern project and how it differs from the contemporary one. (Our disagreement here may spring from his identification of the Galileo-Newton mathematical method as the beginning of 'science' proper, so that Locke (and Berkeley and Reid, and maybe even Boyle) were doing something else (*ibid.*, pp. 14–19). Important as the Galileo-Newton method was (see [2008] or [2011b], chapter 1), it wasn't the only method in use even in areas that now strike us as 'science', so I stand by the conclusion above that 'scientific' and 'philosophical' threads during this period can't be separated.)

about mathematics?—the sort of question the contemporary epistemological literature hadn't helped me address!

When our early modern Natural Philosophers set out to address their version of the central question, they explore cognitive routes from basic perception—distance information by sight, the significance of color vision, the relations of vision and touch⁴⁴—to more theoretical matters—the nature of explanation, the significance of force laws, general rules for Natural Philosophizing. Given the nature of their question, this piecemeal approach makes perfect sense: the way we gain reliable information about medium-sized objects by vision is unlikely to have much in common with the way we gain reliable information about molecular structure; to insist on a general answer would be to set aside the case-specific details that actually do the epistemic work. It's easy to see why the question of 'knowledge' in general just didn't arise.

From our contemporary point of view, we might try to separate the 'philosophical' from the 'scientific' strands in these inquiries, but these thinkers did not. Given that we now regard this as conflating questions of two disparate types, we should expect just what we find: contemporary epistemologists, as 'philosophers', are asking different questions. The puzzle then becomes: how did we get from there to here?⁴⁵

44. Reid [1764] considered our senses of smell, taste, and hearing as well.

45. Pasnau also answers this question differently than I'm about to do. For him, the key is Buridan's 14th-century category of 'moral certainty': 'There is still another, weaker evidentness which suffices for acting well morally . . . if someone, having seen and investigated all the attendant circumstances that one can investigate with diligence, judges in accord with the demands of such circumstances, then that judgment will be evident with an evidentness sufficient for acting well morally—even if that judgment were false on account of invincible ignorance regarding some circumstance' (quoted by Pasnau [2017], p. 34). By the 17th century, Christianity had come to be classified as morally certain, and this turned the demarcation of moral certainty into a crucially pressing question: 'Epistemology now begins to consider where to draw the line between what is and is not sufficiently certain to warrant firm belief' (ibid., p. 43). Here again (see footnote 43), it seems to me that the seeker of moral certainty, like the contemporary theorist of knowledge, could take herself to be pursuing what Pasnau identifies as the early modern epistemological project: the question of moral certainty is important because these are the things we should believe. In other words, there doesn't seem to me to be enough distance between Pasnau's version of the early modern project and the contemporary project for the question of how we got from one to the other to arise.

3. TWO JUST-SO STORIES

A likely thought is that this shift in focus in the theory of knowledge has something to do with the forces that gradually separated the two now-disparate practices over the intervening centuries. Perhaps there is an established historical account of this transition, but if so, I blush to admit that it's unknown to me. There is, however, a story that philosophers sometimes tell themselves. Here it is, coming from Russell:

As soon as definite knowledge concerning any subject becomes possible, this subject ceases to be called philosophy, and becomes a separate science. The whole study of the heavens, which now belongs to astronomy, was once included in philosophy; Newton's great work was called 'the mathematical principles of natural philosophy'. Similarly, the study of the human mind, which was part of philosophy, has now been separated from philosophy and has become the science of psychology. Thus, to a great extent, the uncertainty of philosophy is more apparent than real: those questions which are already capable of definite answers are placed in the sciences, while those only to which, at present, no definite answer can be given, remain to form the residue that is called philosophy. (Russell [1912], p. 155)

Austin puts the sentiment in more colorful terms:

In the history of human inquiry, philosophy has the place of the initial central sun, seminal and tumultuous: from time to time it throws off some portion of itself to take station as a science, a planet, cool and well-regulated, progressing steadily towards a distant final state. This happened long ago at the birth of mathematics, and again at the birth of physics: only in the last century [Austin is writing in 1956] have we witnessed the same process once again, slow and at the time almost imperceptible, in the birth of the science of mathematical logic, through the joint labors of philosophers and mathematicians. Is it not possible that the next century may see the birth, through the joint efforts of philosophers, grammarians, and numerous other students of language, of a true and comprehensive

science of language? Then we shall have rid ourselves of one more part of philosophy (there will still be plenty left) in the only way we ever can get rid of philosophy, by kicking it upstairs. (Austin [1956b], p. 232)

Austin is imagining the birth of linguistics, and indeed Chomsky's groundbreaking *Syntactic Structures*⁴⁶ appeared the year after this passage. For Russell and Austin, in the beginning there is philosophy, and the various sciences spin off from that source. Of course, in our terminology, what 'science' spins off from isn't 'philosophy' (in the contemporary sense), but Natural Philosophy in the sense we're pursuing here. I assume Russell and Austin would take this as a friendly amendment.

Appealing as it may be, I'm not sure this picture matches our case. Between them, Russell and Austin offer a range of examples: two that we've touched on already—celestial mechanics and empirical psychology—and two more recent—mathematical logic and linguistics. In each of these cases, I think it's fair to say that some 'residue' has remained in philosophy: methodological questions about Newton's inferences, puzzles about sensations and the brain, unclarities about the nature of logical truth and linguistic meaning. As Austin says, 'there will still be plenty left'; indeed, the subsequent progress of the spun-off sciences tends to create new 'residues', like the puzzles of quantum mechanics.⁴⁷ The trouble with applying this picture to our present case is that the questions of contemporary epistemology, questions about the scope and nature of 'knowledge', don't look like this kind of residue. To take just one comparison, the question 'which are the logical particles?' was central in the 'pre-scientific' 'joint labors of philosophers and mathematicians' and remains as 'natural philosophical' residue when mathematical logic spins off into a 'science'. In contrast, contemporary questions about 'knowledge' weren't ones our early modern Natural Philosophers even felt the need to ask.

For cases like ours, I like to tell a different sort of just-so story. It begins with Hume's skepticism in the early 18th century. It's well-known that Hume

46. Chomsky [1957].

47. In reality, those planets aren't as 'cool and well-regulated' as Austin implies.

occasioned Kant's awakening from his 'dogmatic slumbers'.⁴⁸ What's perhaps less well-known is that Hume had much the same effect on Reid, as expressed in this letter:

I have learned more from your writings . . . than from all others put together. Your system appears to me not only coherent in all its parts, but likewise justly deduced from principles commonly received among Philosophers: *Principles, which I never thought of calling into Question, untill the conclusions you draw from them in the treatise on Humane Nature made me suspect them.* (Reid to Hume, in Reid [1764], p. 264, italics mine)

Though Kant and Reid were primarily motivated by different aspects of Hume's skepticism—inductive and external world, respectively—it's easy to imagine both of them flummoxed by section I.4.2 of the *Treatise*,⁴⁹ where Hume undertakes to locate the source of the idea of an external object.⁵⁰ That source, he argues, can't be either the senses or reason; in the end, he concludes that it's a 'mistake' of the imagination,⁵¹ and even this much only by a tortured argument of highly questionable cogency.⁵² Kant and Reid could be forgiven for concluding instead that the concept of an object—clearly present in the cognition (for Kant) or the perception (for Reid)—must arise from some source that Hume overlooked. Their contrasting answers to this puzzle define their respective philosophies.

48. Kant [1783], 4:260: 'I freely admit that the remembrance of *David Hume* was the very thing that many years ago first interrupted my dogmatic slumber and gave a completely different direction to my researches in the field of speculative philosophy'.

49. Hume [1739].

50. Continued and distinct body, in Hume's terms.

51. See Hume [1739], I.4.2.35.

52. See, e.g., this excerpt from Stroud's extended analysis (Stroud [1977], pp. 103–104): 'Imagining or conceiving of X requires that one already have the idea of X; we cannot think of something of which we have no idea. In particular, therefore, we could not 'imagine . . . a change in the time without any *variation* or *interruption* of the object' unless we already had the idea of the invariableness or uninterruptedness of an object. But that is just the idea of identity for Hume. It follows that we could not perform that act of mind that is said to produce the idea of identity in us unless we already had the idea of identity to begin with. Hume seems to 'explain' our acquisition of the idea only on the assumption that we already have it, so he does not explain it at all. But without an explanation of the source of the idea of identity his explanation of the origin of the idea of continued and distinct existence cannot get off the ground. He would not have explained how we get into the conflict which the idea of continued existence is intended to resolve'. As if that weren't enough, Stroud continues, 'But even if he had, his account would still run into difficulties'.

Reid recognizes that,

If Nature had given us nothing more than impressions^[53] made upon the body, and sensations in our minds corresponding to them, we should in that case have been merely sentient, not percipient beings. We should never have been able to form the conception of any external object, far less a belief of its existence. (Reid [1764], VI.21, p. 176)

Lucky for us, Nature has formed our ‘constitution’ so that ‘our perceptions . . . correspond’ to ‘impressions on the organs, nerves, and brain’ and ‘vary in kind, and in degree, as they vary’ (Reid [1785], II.2, p. 76); by these means ‘the mind passes immediately from the sensation to . . . conception and belief’ in external objects (Reid [1764], VI.21, p. 177). On careful review, he laments that then-current accounts of the physiological processes from organ to nerve to brain, from Descartes to Hartley, are largely ‘mere conjecture’, but he clearly believes this is the kind of study that’s needed—it just has to be carried out in line with strictly Newtonian methods.⁵⁴ In the end, he’s confident that our perceptual systems arose as ‘necessary in order for our supplying the wants of nature, and avoiding the dangers to which we are constantly exposed’, and that they are ‘admirably fitted . . . to this purpose’ (Reid [1785], II.5, p. 101). Of course, the good pastor Reid attributes this happy circumstance to the ‘parental care and goodness’ of his Maker (Reid [1764], VI.20, p. 170), but it would be entirely consistent with his Natural Philosophical approach to regard Darwin as a friendly and welcome amendment. In the same spirit, his response to Hume’s external world skepticism rests on an empirical case against the Argument from Illusion and the theory of ideas, and a challenge to the underlying assumption that perception is less reliable than our other faculties (introspection, reason, memory).⁵⁵

53. For Reid, ‘impressions’ are physiological: ‘In perception, the object produces some change in the organ [of sense] . . . the organ produces some change upon the nerve . . . and . . . the nerve produces some change in the brain . . . we give the name of an *impression* to those changes’ (Reid [1785], II.2, pp. 75–76).

54. See Reid [1785], II.3.

55. See Essay #3, [2017a], pp. 92–109.

In stark contrast, Kant takes a very different route: he invents an entirely new kind of inquiry, transcendental critique. This transcendental inquiry is defined in explicit contrast to the Natural Philosopher's empirical inquiry:⁵⁶ for example, transcendental psychology—'explanation of the way in which concepts relate to their objects a priori'—is sharply distinguished from empirical psychology—'how a concept is acquired from experience and reflection on it.' Just as we might expect, he locates the beginnings of empirical psychology in Locke:

Such a tracing of the first endeavors of our powers of cognition to ascend from individual perceptions to general concepts is without doubt of great utility, and the famous Locke^[57] is to be thanked for having first opened the way for this. (Kant [1781/7], A85–86/B117–119)

Exemplary as this Natural Philosophical inquiry is, Kant sets it aside in favor of his new transcendental inquiry, presumably to be conducted with new, a priori methods.⁵⁸ The concept of an external object is then located transcendently, in the pure categories, in the structure of the discursive intellect, and not, as Reid would have it, in ordinary human psychology. The same transcendental move informs his response to Hume's inductive skepticism.

So by the end of the 18th century, there were two very different reactions to Hume on offer, one naturalistic, one transcendental. Kant being the philosophical genius that he was, bestriding the subject to this day, it's no surprise that his approach dominated. Various forms of neo- and post-Kantian idealism ensued, eventually provoking the anti-idealist reaction of Moore and Russell, the rise of conceptual analysis and analytic philosophy more

56. This is what makes Kant the first First Philosopher (in the terminology of [2007]). I was slow to realize that if First Philosophy is understood on this Kantian model—as a different sort of enterprise—then Descartes doesn't qualify. (I made this mistake in [2007], I.1, though an inkling of the truth emerged here and there (e.g., p. 308). The light began to dawn in [2011a], pp. 124–125.) Though Descartes uses a priori methods, he does so as part of his investigation of the world, as a Natural Philosopher in the sense intended here. To put the point another way: Kant isn't out to correct Locke, he's engaged in a different activity; Descartes is out to change physics.

57. Kemp Smith's translation offers the charming but apparently inaccurate phrase 'the celebrated Locke.'

58. See footnote 103.

broadly. At that point, the quest for an analysis of ‘knowledge’ naturally became a central undertaking of epistemology, which now often gives way to direct theorizing about ‘knowledge’ itself.⁵⁹ On this picture, then, what we call ‘philosophy’ separated itself from Natural Philosophy when Kant posited an explicitly separate inquiry with a priori methods.

This story reverses the Russell-Austin line: they have the sciences spinning off from Natural Philosophy; my story has philosophy breaking free from the same source. Though both are irresponsible just-so stories, mine seems a better fit for the question at hand: how did epistemology get from there to here? But once again, I think we can learn more by narrowing our focus still further, this time by looking more closely at the fine grain of the early modern epistemological project, and in particular, at the role of the crucial distinction between primary and secondary qualities. This time, instead of asking how the current epistemological project took over from the early modern one, my thought is to ask what became of the leading questions of the early modern project as ‘epistemology’ came to focus on the nature of knowledge.

4. PRIMARY/SECONDARY THEN

A distinction between qualities like size, shape, motion, and number,⁶⁰ on the one hand, and those like color, taste, odor, and sound, on the other, first appeared in familiar form in Galileo:

Upon conceiving a material or a corporeal substance, I immediately feel the need to conceive simultaneously that it is bounded and has this or that shape; that it is in this place or that at any given time; that it moves or stays still; that it does or does not touch another body; and that it is one, few, or many. I cannot separate it from these conditions by any stretch of my imagination. But that it must be either white or red, bitter or sweet, noisy or

59. Recall footnote 41.

60. Different thinkers give slightly different lists, but these differences won’t matter.

silent, of sweet or foul odor, my mind feels no compulsion to understand as necessary accompaniments . . . tastes, odors, colors, and so forth . . . have their habituation only in the sensorium . . . If the living creature were removed, all these qualities would be removed and annihilated. (Galileo [1623], p. 309)

To the dismay of those keen to separate the ‘philosophers’ from the ‘scientists’, here Galileo offers something like a conceptual analysis of what it is to be a substantial body. In the decades to come, as a distinction of this general sort became embedded in the rising Corpuscular Mechanism of Descartes and Boyle, Descartes’s defense followed a conceptual route reminiscent of Galileo:

[T]he nature of matter, or of body considered in general, consists not in its being something which is . . . coloured, or affects the senses in any way, but simply in its being something which is extended in length, breadth and depth. (Descartes [1644], p. 224 (AT VIIIA, p. 42)

Here in Galileo and Descartes, we see precursors to Locke and Newton in the tradition of analyzing what it is to be a ‘body’.

Though perhaps not entirely immune to a priori considerations himself,⁶¹ Boyle most often strikes a different tone, taking empirical evidence as the deciding factor: Corpuscular Mechanism as a whole is ‘to be either Confirmed, or Disproved by, the Historical Truths’, where ‘Historical Truths’ are the fruits of natural history, in other words, observations.⁶² So, for example,

since [Corpuscular Mechanism] maintains that sensory qualities arise from the relations between bodies and our sense organs . . . it predicts a sort of relativity that is in fact observed (e.g., water feels cold to one hand, and warm to another, depending on the state of the hand). (Downing [2002], p. 346)

61. See Downing [2002], p. 346. I draw heavily on pp. 345–347 of Downing’s paper for the rest of this paragraph.

62. For the quotation, see Downing [2011], p. 129.

Appropriately circumspect, Boyle takes this empirical case to be a work in progress, so far inconclusive, but he also appeals to the explanatory power of Corpuscular Mechanism, especially as compared to Aristotelean Scholasticism:

he expressed optimism that corpuscularianism would eventually accumulate explanatory success sufficient for it to merit the assent of reasonable people. (Downing [2002], p. 347)

Boyle also claims other theoretical virtues, including clarity and parsimony.

With these precursors in the background, the canonical treatment of the distinction, its sources and its defense, indeed the very terminology of ‘primary’ and ‘secondary’, comes in Locke’s *Essay*, especially the famous section II.8. Locke dabbles with both a Galileo-Descartes-style conceptual derivation (see below) and a Boylean appeal to empirical predictions and explanatory power,⁶³ but setting aside how the distinction is supposed to be defended, I’d like to focus here on the question of its content: what is it that differentiates a primary quality from a secondary quality? This turns out to be a surprisingly difficult question to answer:

Careful readers of the *Essay* have noticed that Locke makes a number of puzzling claims about the nature of this distinction . . . On occasion, the claims he makes appear to be flatly inconsistent . . . It is not surprising, then, that scholars are deeply divided over how best to interpret Locke’s remarks . . . it is tempting to see this lack of unanimity as revealing ‘those unfortunate inconsistencies for which Locke’s work is so famous’ (Curley [1972], p. 440). It may be that scholars have fastened onto different strands in Locke’s thought, strands that he simply did not recognize as being mutually inconsistent. (Rickless [1997], pp. 297–298)⁶⁴

63. E.g., Locke [1689], II.8.21: with the distinction, ‘we may be able to give an Account, how the same Water, at the same time, may produce the *Idea* of Cold by one hand, and of Heat by the other’.

64. Though he sets out this temptation, Rickless [1997], p. 299, does not himself succumb: ‘The purpose of this paper is to argue that II.8 of Locke’s *Essay* is a unified, self-consistent whole, and that the appearance of inconsistency is due largely to misreadings and misunderstandings’. For two more samples of a vast literature, see Jacovides [2007] and Downing [2009].

Fortunately, the problem of either knitting Locke's apparently conflicting claims together or convicting him of inconsistency isn't one we need to address here. For our purposes, it will be enough to isolate a few of Locke's most salient threads.⁶⁵

The obvious place to start is with the distinction already on the table: having primary qualities is part of what it is to be a body, while having secondary qualities is not. In Locke's version, primary qualities are 'such as . . . the Mind finds inseparable from every particle of Matter' (Locke [1689], II.8.9). Whatever Galileo or Descartes might have had in mind with this sort of conceptual criterion, we're now concerned with Locke, for whom all cognition begins with the senses:

Men have in their Minds several *Ideas*, such as are those expressed by the words, *Whiteness, Hardness, Sweetness, Thinking, Motion, Man, Elephant, Army, Drunkenness*, and others . . . The Mind [is] white Paper . . . without any *Ideas*; how comes it to be furnished? . . . To this I answer, in one word, from *Experience*. (Locke [1689], II.1.1–2)

Among these ideas must be that of 'body'. In our experience, we isolate those properties 'such as Sense constantly finds in every particle of Matter' (ibid., II.8.9)—like size, shape, motion or rest—then exclude any property that the mind can separate—like color, fragrance, taste.⁶⁶ The result is what Locke calls the nominal essence of 'body', an essence constructed by us.

But Locke also recognizes real essences: 'the very being of any thing, whereby it is, what it is . . . the . . . Constitution of Things, whereon their . . . qualities depend' (Locke [1689], III.3.15). According to Corpuscular Mechanism, the real essence is the corpuscular micro-structure, the features and arrangement of its tiny parts; these are responsible for all the properties of a body. Locke draws a crucial link between the nominal essence, drawn from ordinary experience, and Corpuscularianism:

65. As will be obvious, this discussion draws heavily on Downing's reading, especially her [1998]. I depart from her, with considerable trepidation, in footnote 72.

66. Why doesn't color qualify as inseparable? Downing [1998], p. 403, footnote 46, suggests: 'I suspect that Locke agrees with Descartes that experience presents us with uncolored bodies'. (It would presumably be enough if we could conceive of a colorless body.) Of course, Berkeley would disagree.

Take a grain of Wheat, divide it into two parts, each part still has *Solidity*, *Extension*, *Figure*, and *Mobility*; divide it again, and it retains still the same qualities; and so divide it on, till the parts become insensible, they must retain still each of them all those qualities. . . . These I call the . . . *primary Qualities* of Body, which I think we may observe to produce simple *Ideas* in us, viz., of Solidity, Extension, Figure, Motion, or Rest, and Number. (Locke [1689], II.8.9)⁶⁷

It may seem as if Locke is arguing that the qualities isolated by Corpuscular Mechanism—size, shape, motion or rest, etc.—comprise the real essence of body, but this seems unlikely given that (as we’ve seen) he was an early critic of Corpuscularianism and that he hedges his support in passages like this:

I have here instanced the corpuscularian Hypothesis, as that which is thought to go farthest in the intelligible Explication of the Qualities of Bodies; and I fear the Weakness of the humane Understanding is scarce able to substitute another . . . whichever Hypothesis be clearest and truest . . . it is not my business to determine. (Locke [1689], IV.3.16)

Downing ([1998]) makes a compelling case that Locke means just what he says: in agreement with Boyle, he holds that Corpuscular Mechanism is uniquely intelligible because of the direct connection it draws between the fundamental micro-qualities⁶⁸ and the qualities of ordinary experience—the link described in the ‘grain of wheat’ passage. Taking the real essence to be the nominal essence brings it within our epistemic reach. Corpuscular

67. This extension of what’s true of all observed objects to all objects is what may have influenced Newton’s Rule 3 (Newton [1687], p. 795): ‘Those qualities of bodies that cannot be [increased or diminished] and that belong to all bodies on which experiments can be made should be taken as qualities of all bodies universally’. Recall footnote 24.

68. Given that a quality is a ‘Power to produce [an] *Idea* in our mind’ (Locke [1689], II.8.8) and that an individual corpuscle is ‘less than to make it self singly be perceived by our Senses’ (ibid., II.8.9), Downing, [1998], pp. 390–391, points out that, strictly speaking, the size or shape of a corpuscle can’t be a quality. Still, the features of an individual corpuscle do contribute to the aggregate corpuscular structure that produces an idea. In any case, Locke clearly intends that corpuscles and perceivable objects have, e.g., shape in the very same sense, so perhaps his usage here can be thought of as the sort of thing mathematicians happily shrug off as ‘an abuse of notation.’

Mechanism may not be true, for all that, but if it isn't, the true alternative must necessarily be, to some degree, unintelligible to us.⁶⁹

In sum, then, the upshot of this first version of the distinction depends on whether or not we adopt the Corpuscular hypothesis. If we don't, primary qualities are special in that they are part of the nominal essence of 'body', the collection of qualities present in every body of ordinary experience and inseparable by the Mind. If we do, primary qualities are special in the deeper sense of being present in the real essence, in what it is to be a body, in that from which an object's other qualities flow.

This brings us to the doorstep of a second version of the distinction, in terms of explanatory role. All of an object's qualities are to be explained in terms of ('flow from') the object's real essence. Primary qualities, then, are those included in the real essence—shape, size, and so on (again assuming Corpuscular Mechanism)—they are explanatorily basic, and secondary qualities are explained in their terms. As an epistemologist, Locke is particularly interested in explaining the object's ability to produce ideas in us:

The power to produce any *Idea* in our mind, I call *Quality* of the Subject wherein that power is. Thus a Snow-ball having the power to produce in us the *Ideas* of *White*, *Cold*, and *Round*, the Powers to produce those ideas in us, as they are in the Snow-ball, I call *Qualities*. (Locke [1689], II.8.8)

Why does the snowball look spherical and white? Because of the arrangement and qualities of the snowball's corpuscular parts. It looks spherical because of the way its corpuscles are arranged;⁷⁰ it looks white because of the qualities of the corpuscles at its surface.

But now there's a puzzle. What distinguishes the snowball's whiteness from its sphericalness? After all, both result directly from its corpuscular micro-structure. A first thought is that corpuscles, like the snowball, can be spherical, while they can't be white. But this seems too loose a connection: even if some of the corpuscles making up the snowball happen to be

69. Recall footnote 26.

70. 'The way the corpuscles are arranged' is intended to refer to their overall geometric pattern (which is spherical), not a detailed accounting of where each and every corpuscle is located. Here I side

spherical, this isn't what gives the snowball the power to produce an idea of sphericity; that power of the snowball results, not just from the qualities of individual corpuscles, but from their arrangement in aggregate. The power of the snowball to produce an idea of whiteness also depends on the aggregate, but there is a key difference: the aggregate arrangement that produces the idea of sphericity is itself spherical, while an aggregate arrangement of corpuscular sizes, shapes, motions, etc., at the surface just amounts to a larger, more complicated size, shape, and motion.

This is where a third version of the distinction comes into play: primary qualities are special because their ideas resemble the powers that produce them while the ideas of secondary qualities do not.^{71,72} Of course the snowball example makes this look simpler than it is: the circular arrangement of a penny's corpuscles doesn't produce an idea of roundness when the penny is seen from an angle. It does, however, produce an idea of roundness-at-an-angle, which is enough to distinguish it from the qualities of the corpuscles at its surface that produce an idea of copper-color.

Putting all this together, and assuming Corpuscular Mechanism, we have the beginnings of a Natural Philosophical answer to the early modern epistemological question: how do we manage to get reliable information about the world? The answer is that the world consists of corpuscles with primary qualities; these produce ideas in our minds; the ideas of primary qualities inform us of the primary qualities of the objects around us; the ideas of secondary qualities allow us to distinguish between objects, though

with Putnam [1973], pp. 295–296, in thinking that the former is what explains my idea, not the latter. (His well-known example involves explaining why a square peg doesn't fit in a round hole.)

71. Connecting this to another Lockean locution, though both primary and secondary qualities are powers, secondary qualities are 'mere powers' because they don't resemble. Notice, for the record, that even if color is identified with the surface micro-structure, that micro-structure wouldn't resemble the idea of color—so color ontology is irrelevant to the point.
72. Downing [1998], p. 388, footnote 15, seems prepared to accept a weaker notion of resemblance implicit in the 'first thought' in the previous paragraph: 'my idea of a cone would be resembling, even if no ultimate particles are cone-shaped, because the idea gives me an accurate conception of the sort of spatial properties which are had by those particles'. Reluctant as I am to depart from so keen an observer of Locke, it seems to me contrary to his spirit for our ideas to be so uninformative: the idea of sphericity produced by the snowball would only tell me that there are shaped things out there. (The idea of whiteness produced by the snowball only allows me to distinguish it from other objects that don't produce an idea of whiteness. See Locke [1689], II.32.14.) If 'resemblance' operates instead at the level of aggregates of corpuscles, that idea tells me there's something spherical out there.

without giving us direct information about them.⁷³ Formulating this answer involves an interplay between all three versions of the primary/secondary distinction: primary qualities are part of the real essence of ‘body’, secondary qualities are separable from body; primary qualities are explanatorily basic, secondary qualities are explained in terms of primary qualities; the ideas of primary qualities resemble the powers in the body that produce them, the ideas of secondary qualities resemble nothing in the body that produces them. Between them, the three interrelated distinctions touch on conceptual, metaphysical, and epistemic themes.

5. PRIMARY/SECONDARY NOW

Given this portrait of the early modern primary/secondary distinction and its role—central to their Natural Philosophical epistemology—let’s now ask how primary and secondary qualities figure in today’s marketplace of ideas. What topics in this vicinity capture the imagination, command the attention of the contemporary philosophical mainstream? I hardly count myself as finely attuned to such matters, but I think it’s fair to say that sessions at disciplinary meetings, books and articles published, PhDs awarded, all suggest widespread concern with the nature of secondary qualities, particularly color, and more particularly, color ontology.⁷⁴ A first pass at the range of positions would include various forms of subjectivism—color is in the head—objectivism or physicalism—color is in the object—and relationalism or dispositionalism—color is a relation between the object and the head, color is a disposition in the object to produce certain experiences in the head—but philosophical ingenuity has moved the debate far beyond these crude beginnings: for example, the opening chapter of Cohen’s *The Red and the Real*⁷⁵ develops a taxonomy of (at least) seven general categories, each with two additional toggles, for a total (at least) 28 possibilities!

73. If ‘epistemology naturalized’ is supposed to take place wholly within psychology (Quine [1969]), then this account doesn’t qualify. Answering the early modern question requires an account of the world—Corpuscular Mechanism—as well as an account of the mind.

74. Cf., e.g., Cohen [2009], p. vii: ‘A brief glance at recent philosophical journals, book catalogs, conference schedules, or graduate seminar offerings reveals that color has recently returned to its place at the center of philosophical inquiry’.

75. Cohen [2009].

Of course the participants in this lively debate are well aware that the primary/secondary distinction traces back to the early modern period.⁷⁶ Unfortunately, those looking to the early moderns for inspiration are frustrated to find them fluctuating between quite different positions—without apparent embarrassment! So, for example, Zed Adams, in his thought-provoking history of color concepts, lays out an array of possible positions and a string of quotations from Descartes, then remarks: ‘If we ask ourselves which of these positions Descartes holds, the answer is that he seems to hold all of them—which is obviously problematic’ (Adams [2016], p. 91).⁷⁷ Similarly, Jacovides ([2017], p. 194) describes Locke’s ‘careless attitude’ toward color ontology, and Downing ([2002], p. 344) remarks that ‘Boyle does not seem overly concerned with this question.’ Wilson sums up this way:

Descartes, Boyle, Locke and other writers of the . . . period could vacillate rather unselfconsciously among the views of, say, colors as dispositions or power to cause sensations, as the mechanistic structures in objects that accounted for the ‘powers’, or as the sensations themselves. Present-day philosophers may find it necessary to defend *one or the other* of these (or still different) positions, while conscientiously seeking to demonstrate the untenability of its rivals. (Wilson [1992], p. 477)

She traces the discrepancy here to what I think many would regard as its source: the contemporary embrace of ‘more exacting standards of rigor than were characteristic of seventeenth-century treatments’ (ibid.).

But there’s something odd in all this. If we look back at the early modern epistemological story sketched a moment ago, a story with a starring role for the distinction between primary and secondary qualities, we can’t help but notice it was easily told without taking any stand on the ontology of color

76. E.g., Cohen [2009], p. 3, footnote 3, remarks that ‘color physicalism has sometimes gone by alternative . . . labels, including . . . (adverting to the modern distinction among primary and secondary qualities of matter) “primary quality theory of color”’.

77. Adams [2016], pp. 84–92, gives a thorough account of commentators’ efforts to distill a coherent view: ‘Many have noted that Descartes appears to vacillate between different positions on the nature of color. The almost universal response . . . has been to attempt to explain it away’ (p. 91). See also Nolan [2011], pp. 81–82.

or other secondary qualities. It involves primary qualities in the object producing ideas in us, some of which resemble features of the object and some of which don't. Nothing in this requires us to determine whether color *is* the qualities and arrangement of the corpuscles, the disposition to cause the idea, or the idea itself; all that matters is that the idea of red doesn't resemble the corresponding quality and arrangement of the corpuscles. So maybe the reason Descartes, Locke, Boyle, and others don't take a firm stand on this matter isn't because they're sloppy or unrigorous, but because they don't care—or better, maybe it isn't even a question that presented itself to them, in something like the way the nature of a general notion of 'knowledge' apparently didn't.

In fact, there are hints in the writings of some of our early modern Natural Philosophers that the question of color ontology isn't just without interest, but that it may even be ill-posed. As he works his way up from smell and taste to hearing and touch, Reid writes:

If it is asked, Whether the smell be in the rose, or in the Mind that feels it? The answer is obvious: That there are two different things signified by the smell of the rose; one of which is in the mind . . . the other is truly and properly in the rose . . . All the names we have for smells, tastes, sounds, and for the various degrees of heat and cold, have a like ambiguity . . . They signify both a sensation, and a quality perceived by means of that sensation. (Reid [1785], II.16, p. 195)⁷⁸

This ambiguity in the terms leads to a conflict:

The vulgar say, That fire is hot, and snow cold, and sugar sweet; and that to deny this is a gross absurdity, and contradicts the testimony of our senses. The Philosopher says, That heat, and cold, and sweetness, are nothing but sensations in our minds; and it is absurd to conceive, that these sensations are in the fire, or in the snow, or in the sugar. (Ibid., II.17, p. 205)

78. See also Reid [1764], II.8–9 and V.1.

If there's a real metaphysical fact of the matter to be settled here, either the Vulgar or the Philosopher must be wrong, but Reid believes 'this contradiction between the vulgar and the Philosopher is more apparent than real' (ibid.); it's what we would call merely verbal,

[O]wing to an abuse of language on the part of the Philosopher, and to indistinct notions on the part of the vulgar. (ibid.)

The Philosopher is abusing language because on the ordinary, ambiguous sense, the fire is hot; the Vulgar notion is indistinct, because his usage is ambiguous. Once each understands the sense in which their respective claims are meant, the disagreement vanishes.⁷⁹

And what about color? Here Reid surprises us:

Colour differs from other secondary qualities in this, that whereas the name of the quality [in the object] is sometimes given to the sensation which indicates it, and is occasioned by it, we never, so far as I can judge, give the name of *colour* to the sensation, but to the quality only. (Reid [1764], VI.4, p. 87)

He speculates that this may be because

the appearances of the same colour are so various and changeable, according to the different modifications of the light, of the medium, and of the eye, that language could not afford names for them. (Ibid.)

Perhaps, he's suggesting, it's more efficient to use the word for the constant underlying quality in the object. But however that may be, Reid is clearly making an empirical claim about how color vocabulary is in fact used, just as he was with the other secondary qualities, not a metaphysical claim about what color really is.

79. Reid [1785], II.17, p. 206: 'This speech of the Philosopher . . . is meant by him in one sense; it is taken by the vulgar in another sense. In the sense in which they take it, it is indeed absurd, and so they take it to be. In the sense in which he means it, it is true; and the vulgar, as soon as they are made to understand that sense, will acknowledge it to be true.'

In the context of Newton's more sophisticated account of color vision—which includes light rays as potentially 'colored' alongside the object and the experience—Howard Stein writes:

it seems to me a strange view of philosophy that holds that there is a single canonically 'right' use of the word 'color'. (Stein [2004], p. 167, footnote 12)

Reid, at least, rejects this 'strange view'.⁸⁰ The causal story of the object, the light, the stimulations of the visual system, and the phenomenal experience can be described without any decision on which portion really is the color. In a similar key, Stein remarks:

The questions that arise about colors—about, in ordinary parlance, 'the colors of things' (of the sky; of the sun; of the nocturnal and the diurnal moon . . .)—can be discussed, in a way I have tried to exemplify in my earlier remarks, without *any* doctrine about the ontology of color; and such discussion—I hope I have convinced you by those examples—can genuinely satisfy the curiosity expressed by such questions—again, without any need to take a stand, *one way or the other*, on the issue of whether colors are in some fundamental sense 'really in things'. (Ibid., pp. 155–156)

Given that our questions about how color vision works and why the sky looks blue can all be answered without a stand on color ontology, the Natural Philosopher might well conclude that which element of the causal chain we regard as 'color' is a linguistic matter, a matter of labeling, that the ontological question is asking for a fact of the matter where there is none.⁸¹

80. Immediately before the remark just quoted, Stein [2004], p. 167, footnote 12, notes that 'Newton distinguishes *various senses* of "color" "'—'what colors are in the object, in the rays, and in the sensorium'. These distinctions are—I should say *obviously*—clarifying'. It might appear that Stein is attributing a Reid-like position to Newton, but in fact, in the very passage Stein is quoting, Newton insists that 'the Rays to speak properly are not coloured', and that it's the sensations only that have 'the Forms of Colours' (Newton [1704], pp. 124–125). So Newton certainly appears to hold the 'strange view of philosophy'.

81. Stein [2004], p. 156, doesn't draw this strong conclusion: 'I call this a 'moderate' rejoinder because it does not claim to refute a theory of the ontology of color. [I take him to mean that it doesn't claim to show that the ontological question about color is ill-formed.] It does claim to refute the view

At the very least, then, these early modern Natural Philosophers weren't interested in the ontological question about color; at worse, they thought it ill-posed. This suggests that in terms of our just-so stories, the emergence of color ontology as a central theme, like that of the nature of 'knowledge', wasn't a Russell-Austin-style case of a Natural Philosophical matter left behind as 'science' spins off, but a Kant-started-it-style case of newly independent 'philosophy' posing a question of its own, to be settled by its own methods. This returns us to a more focused version of the question posed a while back: if these mainstream contemporary concerns are new, not the same as the early modern concerns, then what has happened to those earlier concerns? In particular, as color ontology has taken center stage, what's become of the focal points of the early modern discussions of primary and secondary qualities?

6. WHAT'S BECOME OF THE EARLY MODERNS' CONCERNS?

Recall, once again, that the early modern epistemological story involved three versions of the primary/secondary distinction: primary qualities are part of the concept of 'body', secondary qualities are not; primary qualities are explanatorily basic, secondary qualities are not; our ideas of primary qualities resemble those qualities, our ideas of secondary qualities do not. When the distinction reaches its full development in Locke, the role of the conceptual distinction is to anchor the case for the unique intelligibility of Corpuscular Mechanism: the very properties that comprise our nominal essence of 'body'—size, shape, motion, etc.—also comprise its real essence, that is, the properties from which all the others flow, the explanatorily basic properties. The key components of the story, then, are two: intelligible explanatorily basic qualities, and ideas, some of which resemble and some of which don't. Let's take these up in turn.

that one such theory is *clearly prerequisite* to the understanding either of ordinary discourse or of our best current knowledge of color'.

As we've seen, Locke's hope that the uniquely intelligible primary qualities of Corpuscular Mechanism would explain all the properties of objects was under pressure even in the *Essay* from the cohesion of bodies and the workings of contact forces. As understanding has progressed, Locke's worries have been sustained; both these phenomena turn out to be beyond the reach of Mechanistic thinking, both turn out to be electromagnetic phenomena. Here the explanatorily basic properties are the values of the electro-magnetic field—qualities undreamed of in early modern thinking and entirely absent from the nominal essence generated from Locke's perceivable grain of wheat. Here, as elsewhere, the requirement of 'intelligibility' in Locke's sense necessarily falls away.⁸²

Today, we have the wild menagerie of particles detailed in the Standard Model, with properties like spin and color charge.⁸³ Some would take these particles as fundamental, and spin, color, and so on, as explanatorily basic; others seek to explain particles in terms of more fundamental quantum fields, whose values would then be explanatorily basic; still others would reduce both these to the mysterious strings of string theory. So this aspect of the early modern role for primary qualities—marking the search for physical fundamentals—is very much alive today.

A natural label for this contemporary study would be 'foundations of physics'—for that matter, much of what our early moderns were up to could also be described in that way. Nowadays, when 'science' and 'philosophy' are regarded as distinct enterprises, it's worth noting that some

82. See Wilson [1992], pp. 477–478, for reflections on this development. Stein [2004] traces the overthrow of Locke's intelligibility requirement to Newton's optics, in particular to his isolation of a specific number, the 'fit of easy reflection' (Newton [1704], p. 281), for each type of ray: Newton 'calls these "*original and connate properties*" of the rays; and the role of these properties in the . . . understanding of light and vision that Newton develops is entirely analogous to the role that *should* be played, according to Locke, by *primary* qualities' (Stein [2004], p. 142). This may well be right, but I worry that some additional argumentation would be needed to show that the Corpuscular Mechanist couldn't reduce rays to aggregates of corpuscles and fits to features those aggregates enjoy due to the qualities and arrangement of those corpuscles (e.g., if the fit is a motion, the aggregate might have that fit due to the combined motions of its component corpuscles); if so, fits wouldn't be explanatorily basic, after all. In any case, for our purposes, it isn't important exactly when the intelligibility of explanatorily basic properties, their connection to sensory qualities, was first rejected, just that it was.

83. Of course, color charge has nothing to do with color. See Feynman [1985], p. 136: 'The quarks have an additional type of polarization that is not related to geometry. The idiot physicists, unable to come up with any wonderful Greek words anymore, call this type of polarization by the unfortunate name of 'color', which has nothing to do with color in the normal sense.'

contributors to this discussion are housed in philosophy departments and some in physics departments.⁸⁴ Those in philosophy departments would most often be regarded as philosophers of physics, but of a particular breed. Some philosophers of physics, some philosophers of science more generally, take their questions from mainstream philosophy—say, the problem of universals⁸⁵—and look to physics to answer them; this is philosophy informed by physics. In contrast, those participating in what I’m here calling ‘foundations of physics’ take their questions from physics, lending a hand with questions that arise—like those above about the explanatory basicness of particles, field values, or strings—within the ‘science’. Philosophers of physics engaged in this project are sometimes challenged to explain why their work isn’t just physics, isn’t just ‘science’; their rightful membership in the distinctive inquiry of ‘philosophy’ is sometimes questioned. Likewise, those theoretical physicists similarly engaged might be accused of straying from their own disciplinary norms. This is unsurprising, given that both groups are addressing questions of a piece with those of our early moderns, questions formulated in an arena where ‘philosophy’ and ‘science’ aren’t separable. This is the Natural Philosophical residue left over after most of ‘physics’ spins off, Russell-Austin-style, into a distinct, well-regulated discipline.

So, if the worldly side of the early modern epistemic story, the side that counts primary qualities as explanatorily basic, lives on in the foundations of physics, what’s become of the cognitive side, the relation between the world as so described and our cognition? The key notion here is resemblance: primary qualities are those whose ideas resemble them. As it happens, the most salient case—perception of size, shape, and distance by sight—has been central to theorizing about vision from the start. By the early modern period, Kepler’s discovery of the retinal image had contributed to the rise of the theory of ideas,⁸⁶ and with it the tendency to phrase the problem in its terms: how do we get accurate information about three-dimensional shape, size, and distance from a two-dimensional projection? Descartes, Berkeley, and Reid all discussed these matters in detail, often including experimental

84. See, e.g., the contributors to Saunders et al. [2010].

85. This example is drawn from chapter 3 of Maudlin [2007], a book devoted to addressing metaphysical questions by appeal to physics (as its title suggests).

86. See, e.g., Jacovides [2017], chapter 7. See also [2017a], pp. 124–136.

work, and (as noted) Berkeley especially had a profound influence on the discipline that eventually spun off, Russell-Austin-style, into contemporary vision science. Meanwhile, Newton was founding the study of color vision, showing how to predict the color experience that would result from a given combination of rays.⁸⁷

In the centuries since, visual perception of the world's geometric features has turned out to be a complex and subtle business, involving intensive processing (e.g., stereopsis) and visual cues (e.g., the bluish cast of distant mountains).⁸⁸ Though space itself is locally Euclidean, our visual experience of it is neither perfect resemblance nor a two-dimensional projection: a long alleyway looks narrower in the distance, but it doesn't look like two converging lines on a plane, as a projection would; it's some kind of foreshortened three-dimensional space, whose structure is a lively topic of contemporary research.⁸⁹ The mechanics and physiology of color vision are comparatively well-understood.⁹⁰

So in this case the updated version of the early modern primary/secondary distinction would be the question: are worldly size, shape, distance, etc., registered differently from the workings of color vision? Phenomenally, the answer would seem to be 'yes'—it's natural to think that our experiences of size, shape, and distance present themselves as geometric, while our color experiences bear no such resemblance to anything external—but physiologically both are simply encodings of certain features of the world. The question, then, is whether there's a principled difference between the way the system for spatial vision and the system for color vision register information.⁹¹ This

87. See Stein [2004], §6.

88. See, e.g., Palmer [1999], chapters 5 and 7.

89. See, e.g., Hatfield [2003a], Wagner [2006].

90. Palmer [1999], p. 95: 'Certain aspects of color perception are among the best-understood topics in vision science, perhaps in all of cognitive science.'

91. This question is even trickier than it seems because it requires a prior characterization of what information the visual system is out to encode. For example, if we take color vision to be encoding the reflectance distribution along an object's surface (its SRD), then the phenomenon of metamers (two SRDs that look the same) is a failure, but if we take the goal to be encoding, e.g., the location of ripe fruit against a leafy background, then metamers that don't effect this function (e.g., metamers that aren't present in the environment) are irrelevant (see Hatfield [1992]). A contemporary mainstream philosopher might be tempted to answer the characterization question simply—the color vision system is out to encode color—then stage the debate just mentioned (between SRDs or 'psychobiological properties' (Hatfield [2003b])) as a disagreement over color ontology. From the perspective being developed here,

is part of a larger inquiry into the mechanics of vision, what might be called ‘the foundations of vision science’. Here once again we find a joint project of researchers from what are officially different forms of inquiry, this time, from departments of philosophy and of psychology or cognitive science.⁹² In other words, here, too, we find a residue left behind when a science spins off into independence—this time vision science—and practitioners who straddle disciplines, or maybe better, practitioners who pursue a single enterprise that’s artificially segmented by our contemporary insistence that ‘philosophy’ and ‘science’ are distinct undertakings. Let’s just call them Natural Philosophers and leave them be!

7. THE PLEA

This brings us full circle, at last. We began with Austin, who recommended a certain topic—excuses—as well-suited to a given method—ordinary language philosophizing. Reversing the order, I set out to recommend a certain method—Natural Philosophizing—as well-suited to a given topic—the world and our place in it. To describe the method, I returned to the early moderns—especially Descartes, Boyle, Locke, Newton, Berkeley, and Reid—and called attention to their seamless integration of what we now call ‘philosophy’ and what we now call ‘science’. Tracing history forward, the concerns of these Natural Philosophers survive in what we now consider as interdisciplinary inquiries, in the foundations of physics and of vision science. Once we recognize the phenomenon, it can be found in many other areas: in consciousness research,⁹³ in the foundations of mathematics,⁹⁴ in the theory of evolution,⁹⁵ and more. The fact is that pretty much any ‘science’

the underlying issue is a difficult evolutionary question: what purposes was the visual system designed to serve (in the terms of Marr [1982], a task analysis)? Masking this as a metaphysical question is at least a distraction, and at worst, the replacement of a real question with an ill-posed question that invites the application of inappropriate methods.

92. See, e.g., the contributors to Hatfield and Allred [2012].

93. See, e.g., Metzinger [2000].

94. See, e.g., Koellner and Woodin’s ‘Exploring the frontiers of incompleteness’ project at Harvard, 2011–2012: <http://logic.harvard.edu/colloquium.php>.

95. See, e.g., Orzack and Sober [2001].

that's spun off from the Natural Philosophical core has left behind some Russell-Austin-style foundational residue.

So, where does this leave the questions that spun off instead as Kant-started-it, independent 'philosophical' questions? I've suggested that inquiry into the nature of 'knowledge' is such a question. When the topic 'the world and our place in it' was introduced, I postponed consideration of what it leaves out; the nature of knowledge now presents itself as a candidate. It seems we could have a complete account of how, for example, the visual system registers information without touching the question of whether the agent 'knows' or 'sees that' in the presence of papier-mâché barns; presumably even a complete account of all the ways we gain reliable information wouldn't tell us whether 'knowing' is a factive purely mental state.⁹⁶ These contemporary epistemological questions appear to float free of what's going on in the world. To put the point another way: does the world contain a concept 'knowledge' to be analyzed⁹⁷ or a worldly item 'knowledge'^{98,99} to be

96. Cassam [2009], p. 27, describes a slight variant of Williamson's project that 'relies on armchair reflection rather than empirical science'.

97. Austin mounted a case against concepts/meanings/universals that runs parallel to Quine's (see [2017a], pp. 60–66). Psychologists speak of 'concepts', but insofar as this marks a serious theoretical posit, I doubt it can do the philosophical work required.

98. This includes both knowledge-first views stemming from Williamson (see footnote 41) and 'naturalistic' views like Kornblith's (see the following footnote). As an example from the first group, Cassam ([2009], p. 27) describes a Williamson-like view and remarks that it 'does not treat knowledge as a natural kind or as something to be studied by analyzing the concept of knowledge into more basic concepts. Instead [it] explains what it is to know A by identifying different means of coming to know it . . . all one needs is an open-ended list of means of knowing that A.' Cf. Williamson [2000], p. 34: 'knowing that A is seeing or remembering or . . . that A if the list is understood as open-ended' (quoted by Cassam [2009], p. 28). It's difficult to see how any of this can be done 'from the armchair' (see footnote 96); e.g., when Einstein and Perrin discovered means of coming to know that there are atoms, this was an empirical breakthrough. With no compunction about leaving the armchair, the Natural Philosopher might be forgiven for thinking that if 'know' is just the label for an open-ended list of ways of gaining information about the world, then the epistemologist's job is to study and assess the items on that list, from the senses to scientific confirmation.

99. Kornblith [2002], p. 29, takes the opposite approach, holding that 'knowledge constitutes a legitimate scientific category. In a word, it is a natural kind'. In particular, he argues that 'reliably generated true belief' is a collection of 'homeostatically clustered properties' and identifies this natural kind with 'knowledge'. One engaged in the contemporary epistemological debate might object to the last step: even if there is such a natural kind (see Roth [2003] for some doubts), on what grounds is it identified as 'knowledge' when 'there is a host of additional candidates, each corresponding to a different theory that was floated in response to the Gettier problem' (Goldman [2015])? Our early modern Natural Philosophers and their descendants would question the wisdom of lumping, e.g., reliably generated true visual beliefs with reliably generated true beliefs about molecular structure; as noted earlier, expecting an account at this level of generality blinds us to real work on how we get reliable information by vision,

studied or just a word ‘knowledge’ with an impressive array of subtle and effective ordinary language uses, a suitable topic for Austin’s ordinary language philosophizing?¹⁰⁰ I suggest that what we’ve seen supports the third option. If this is right, then the contemporary philosopher’s questions about knowledge, intended in some other sense, outstrip the target topic here: the world and our place in it.¹⁰¹

This conclusion runs parallel to our previous musings on the question of color ontology. It seems we could have a complete account of color vision—from the micro-physics of the object to the behavior of light to the workings of the visual system—and still not have an answer to the central question of color ontology: what is color? Updating Reid, the contemporary psychologist Davida Teller writes:

Our goal is to unite three interestingly diverse kinds of entities: *Visual stimuli* (e.g., physical objects and their properties), *neural states* (the states of ensembles of neurons at many processing stages within the visual system), and *conscious perceptual states* (our visual perceptions of particular physical stimuli) . . . As far as I can see, color realism is the view that of the vision scientist’s three entities—surface spectral reflectance, neural signals, perceived color—one *is* color, and the other two are not. But if you ask a color scientist which of the three entities *is* color, she will answer that the question is ill-posed. We need all three concepts, and we need a conceptual framework and a terminology that makes it easy to separate the three, so that we can talk about the mappings among them . . . We care much more about our fundamental distinctions than we do about who

which is quite different from real work on how we get reliable information in physical chemistry. It’s as if Kornblith wants to use the Natural Philosophical method but is foiled by accepting the contemporary formulation of the leading epistemological question: ‘It is worth pointing out that this conception of philosophy [i.e., his ‘naturalism’] sits well with the practice of philosophy throughout most of its history. The great philosophers of the past were not narrowly trained specialists. . . quite a number of them were able scientists in their own right’ (Kornblith [2002], p. 176). (See also [2017a], pp. 218–219.)

100. This is what I take Austin to be doing in his well-known [1946], in implicit response to the skeptic. See [2017a], pp. 65–66.

101. I wonder if this is what Cassam [2009], p. 27, means when he says that knowledge, the target of his study, isn’t a ‘natural kind’.

owns the word *color* . . . The argument . . . seems to me to collapse to an uninteresting terminological dispute. (Teller [2003], pp. 48–49)¹⁰²

As Stein put it, why insist on a single ‘right’ use of the word? If something does determine the right use, then it, too, seems to fall outside the scope of our topic.

In general, then, Kant-started-it independent ‘philosophical’ questions float free of the early modern inquiry and its Natural Philosophical descendants: everything that needs explaining—how we gain reliable information about the world, the functioning of color vision, the ‘colors’ of objects—can be explained without answering them. This isn’t to say that they’re necessarily pseudo-questions; presumably they’re intended as part of a different sort of inquiry with its own methods, just as Kant claimed.¹⁰³ But for those of us set on investigating the world and our place in it—if that’s your topic—I stand by my recommendation: take counsel from the early moderns, stick with Reid’s modest response to Hume, and adopt the seamless methods of the Natural Philosopher.¹⁰⁴

102. Along the way, Teller also mentions usages like ‘red light’ and ‘red cones’. See also the psychologist Richard Warren [2003], p. 51: ‘emphasis on reserving color terms for the inherent ‘physical’ color of objects . . . can make understanding color appearance more difficult . . . Perhaps it is not necessary to choose between these views [preferring a] laissez-faire approach: Use the same term to describe both the stimulus and the sensory/perceptual response, and allow the context to make it clear which aspect is being considered’. The title used by another psychologist, Paul Whittle [2003], makes his point ‘Why is this game still being played?’: ‘Why are we (particularly philosophers) still caught up in the problem of objectivity vs. subjectivity? . . . We seem unable to let the problem rest. Scientists are as divided in their opinions as philosophers, although in my experience their work is remarkably independent of their metaphysics, and when necessary they set up dual definitions’ (p. 203). In the end, he makes a pitch for what I’m calling Natural Philosophy: ‘It might be thought that these grumbles reflect a scientist’s impatience with philosophy per se. Quite the opposite . . . The science of perception needs philosophy just now, and philosophy has much to offer . . . It has been my dawning awareness of these resources that has made me impatient with philosophy that seems to be playing old tunes with only minor variations’ (p. 204).

103. Of course, the Natural Philosopher will want to know what this other sort of inquiry is, what questions it sets out to answer, what methods it uses in pursuit of those answers, and what reason we have to think those methods are reliable. E.g., see [2007], I.4, in reaction to Kant, I.5, in reaction to Carnap, or Essay #2, in reaction to van Fraassen.

104. Thanks to Lisa Downing and Jeffrey Schatz for their helpful comments on earlier drafts. Neither should be assumed to second my more audacious conclusions.

On the Question of Realism

In a number of recent, and some perhaps now not so recent, writings, I've been describing a straightforward but highly idealized inquirer called the Second Philosopher.¹ She begins with simple perceptual beliefs, progresses to generalizations, then experimentation, then theory formation and testing, always circling back to examine and correct her beliefs and methods as she goes. Along the way, I've confronted her with a series of traditionally 'philosophical' questions to explore how they look from her perspective: the challenge of radical skepticism, the nature of truth and reference, the ground of logical truth, and the ontology, epistemology, and methodology of mathematics, from arithmetic to higher set theory. None of this amounts to a direct argument for approaching philosophy as the Second Philosopher does, but my hope has been that at least some readers will feel the draw.

Though I never set out to address the status of the theoretical posits of natural science, the issue came up indirectly in a discussion of Quinean holism, the linchpin of his indispensability argument for the existence of mathematical entities.² I'd felt free to help myself to indispensability considerations—then widely accepted—in my early realistic philosophy of mathematics,³ but nagging conscience soon led me to doubt that the role of mathematics in 'our best scientific theory' could bear this kind of ontological weight. In fact, the assumptions of science seem a remarkably varied lot, from explicit idealizations (like continuity in fluid dynamics), to requirements that may or may not be accurate (like the continuity of space-time), to items currently

1. See [2001], [2007], [2011b], [2014b], [2017a].

2. See [1997], II.6.

3. See [1990].

posited but whose relation to the world is still largely mysterious (like the wave function of quantum mechanics).

This line of thought brought me to the case of atoms, before and after Perrin's famous experiments at the turn of the 20th century.⁴ I argued that they were part of 'our best scientific theory' before Perrin, but weren't fully confirmed until after, and thus, that holistic indispensability alone is not enough to support ontological conclusions. (The same would presumably go for other general conformational rubrics like inference to the best explanation.) And so, incidentally, the presence of mathematical entities in 'our best scientific theory' wouldn't be enough to establish their existence.

In this way, pursuit of a question in the philosophy of mathematics landed me—and the Second Philosopher with me—at least in the vicinity of the realism/instrumentalism debate in the philosophy of science: I'd claimed, on her behalf, that we have rationally compelling evidence, now, for the existence of atoms—better evidence even than Perrin, given cloud chambers, electron microscopes, and so on—which would apparently make me a realist.⁵ I say 'apparently' because, paradoxically, claiming rational belief in atoms doesn't seem to be enough to qualify one as a 'realist' as the term is currently used, or at least as it's used by two influential instrumentalists: Bas van Fraassen and Kyle Stanford. My goal here is to trace how this paradox plays out in the views of these two thinkers. The point is comparatively straightforward for van Fraassen, so I start with him, then use his constructive empiricism as a foil for the more complex case of Stanford's epistemological instrumentalism. In the end, I hope to show that the distance between Stanford's position and the Second Philosopher's is considerably shorter than labels and rhetoric would suggest.

4. See [1997], pp. 135–143, [2007], pp. 404–407.

5. Van Fraassen [2015], p. 7, footnote 5, remarks that 'The debate between empiricists and scientific realists is not over the reality of unobservable entities, but over the *telos* of scientific activity', but he admits that 'What is true is that the debated questions would be moot given an epistemology that would entail that the existence of unobservable entities . . . could be established on the basis of the evidence'. Unpacking this precisely will have to wait, but for now, the Second Philosopher and I are claiming that the existence of atoms, and thus the existence of unobservable entities, has been established (at least within the bounds of scientific fallibility) by experimental evidence.

1. VAN FRAASSEN

In van Fraassen's terms, both Second Philosophy (what the Second Philosopher does) and Empiricism are stances:

By a stance I mean a position that consists in attitudes, commitments, and a characteristic approach to philosophical problems, possibly including or presupposing some beliefs as well, but not defined by a thesis or doctrine. (p. 65)⁶

Maddy's Naturalism [is] characteriz[ed] quite explicitly as a *philosophical stance*. (p. 66)

Even the quick sketch of the Second Philosopher given above should demonstrate that this attribution is entirely accurate.⁷ This raises the question whether rational disagreement between two stances is possible; here van Fraassen's example is the analytic metaphysician:

Can we not say: the metaphysician disagrees, has a different and incompatible stance, but disagreement does not make one unscientific [or] irrational. (van Fraassen [2002], p. 48)

Arguing that the answer is no, van Fraassen switches to the example of the crank:

[A] disregard for evidence, a refusal to submit one's ideas to natural selection by relevant experiment or to engage in vigorous testing when nature itself does not put one to the test (these are some examples, none of them factual beliefs!) can certainly take one beyond the scientific pale. So the paradigm of science does not suggest that disagreement in the sphere of attitudes, commitments, values, and goals is invulnerable to empiricist critique. On the contrary, it suggests that it is. (van Fraassen [2002], p. 48)

6. Unaccompanied page numbers in the text refer to van Fraassen [2015].

7. I chose to introduce Second Philosophy by describing the behavior of the Second Philosopher because I took it to be what van Fraassen would call a stance (see, e.g., [2007], p. 1).

So perhaps there is room for rational adjudication of any disagreement between the Empiricist and the Second Philosopher.

Since science, in some sense, is to act as judge, and since both parties tout their fundamental respect for that practice,⁸ here the Empiricist is faced with a more delicate case than those of the analytic metaphysician and the crank. Under the circumstances, it's obviously important to get the fine points of the Second Philosophical stance right. I don't think van Fraassen quite does this, so I'll try to set the record straight here and there. Some of these corrections, especially at the beginning, may seem nitpicky, but my hope is to escalate in small steps and display along the way what I take to be the particular virtues of the Second Philosopher's approach. Whether this amounts in the end to a comparative advantage for her stance over the Empiricist's is obviously a separate question.

But first a note on terminology. Van Fraassen discusses the Second Philosopher under an alias: the Naturalistic Native, or Native for short. His aim is to distinguish the inquirer I describe from me, the person doing the describing—that person he calls the Naturalistic philosopher.⁹ This distinction seems less important to me than it does to him because I take his Naturalistic philosopher—me—to be using the same methods, to be taking the same stance, as his Naturalistic Native—my inquirer.¹⁰ The sources of our disagreement on this point should be clarified below (see (ii)).

8. Cf. van Fraassen [2015], p. 64: 'Naturalism and Empiricism place the admiration of empirical science, as a paradigm of rational inquiry, at center stage'.

9. See van Fraassen [2015], p. 67.

10. In an APA book session on [2007], Barry Stroud raised the question whether I, the author of the book, was a Second Philosopher. When I'd asked myself this question in the past, I'd felt confident that the answer was yes, because I wasn't using any methods that aren't available to the Second Philosopher, but it turned out that Stroud had something else in mind: were the questions I was addressing ones the Second Philosopher would naturally be led to ask? Here again, I think the answer is yes. I take this as obvious for cases like the nature of truth and reference, the ground of logical truth, the proper methods of mathematics and their rational defense, and even for some of the issues surrounding radical skepticism, but a few others, like the analysis of the enduring attractions of the dream argument and the argument from illusion in [2017a] or of contemporary worries about the prospects for metaphysics naturalized in [2007], V.5, arise only when the Second Philosopher inquires into the nature of the human practice of philosophizing. But she will get to that eventually!

(i) *Scope*

Here I have in mind both questions about where the Native starts and about where she ends up. Van Fraassen takes her to begin with ‘the *currently accepted scientific view*’ as ‘the rock to build on’, as her ‘personal foundation . . . as the received knowledge, or securely founded opinion, about what is the case.’¹¹ He sets this in contrast with the Empiricist, who takes the same body of beliefs ‘as our natural first target for analysis and reflection, for critique and interpretation, as precisely what you need to subject to critical reflection.’ I have used the phrase ‘native to the contemporary scientific worldview’—a nod to Quine’s ‘naturalistic philosopher’ who ‘begins his reasoning within the inherited world theory as a going concern’ (Quine [1975], p. 72)—but the Second Philosopher is first introduced along the lines given above:¹² she begins with ordinary perception (not some pre-certified conclusions to be taken for granted) and gradually extends her methods, always circling back to reassess and correct both those methods and her beliefs.¹³ Notably, in none of this does she employ a distinction between scientific beliefs and the rest, which she would need in order to identify the starting point van Fraassen assigns to her. (I assume we’d all count this as a virtue, given that there’s no way of corralling the ever-evolving ‘methods of science’ into a fixed criterion.)¹⁴ This means, for example, that she meets Descartes’s Method of Doubt on its own terms: she doesn’t denounce the search for a priori certainties as ‘unscientific’—if there’s a way to get them, as Descartes promised, she’d like very much to know about it!—it’s just that as she follows his line of thought, she doesn’t see that his various claims are justified.¹⁵

11. All the quotations from van Fraassen in this paragraph come from his [2015], p. 76.

12. Though van Fraassen [2015] makes heavy use of [2001], the Second Philosopher doesn’t make her appearance until [2007].

13. See [2007], p. 14, [2011b], II.1, [2014b], p. 2. Though I consider the possibility of characterizing the Second Philosopher top-down instead of bottom-up (in [2011b], IV.5), even in that case, her starting point isn’t ‘the currently accepted scientific worldview’ but the entirety of our beliefs, reasonable and not. (Van Fraassen [2015], p. 75, may have this in mind when he says ‘if we ask where we are just now, we are confronted by a plethora of answers from our neighbors and peers, as well as by uncertainties and ambiguities in what we find ourselves tending to answer.’) The Second Philosopher’s job is then to figure out how to whittle this sprawling mass down to its justified portions.

14. See [2007], p. 1.

15. See [2007], I.1, [2017a], pp.7–18.

Though he doesn't attribute this explicitly to the Native, van Fraassen also includes a strain of ontological parsimony in his general characterization of Naturalism:

[T]he ontological view that everything is physical, material, or within the domain of the natural sciences . . . is typically understood to classify, as needing to be 'reduced' or explained away, such items as consciousness, reference, goodness, beauty, purpose, functions, etc. (p. 63)

Another corollary to the Native's starting point is that she begins with no such restrictions in place. As to where she ends up, there's also no pre-existing restriction to what's typically regarded as the disciplinary hard sciences. She observes particular details that fall beneath the notice of physics, botany, or metallurgy—the pattern of bruising in the grass where the cannonball falls or the remnants that cling to the ball itself as it rolls to a stop—and her passion for generalizing and understanding would reach into disciplines like history, sociology, anthropology, and linguistics. Interdisciplinary thinkers behaving second-philosophically are currently engaged in vision science and consciousness studies, and there's no predicting what revamping of our ontological expectations might be needed as these inquiries develop. For that matter, my own second-philosophical work in the foundations of higher set theory makes room for mathematical abstracta.¹⁶ I don't know what might emerge when the Second Philosopher turns her attention to the human practice of moral judgment, but it obviously falls within her purview, and the discovery of objective moral facts, for example, can't be ruled out in advance.

(ii) *The force of 'native'*

At one point, van Fraassen describes the 'Naturalistic Native' as someone who 'lives [in the scientific worldview] as unselfconsciously as a fish in water or a bird in air' (p. 92). Modulo the caveat above about 'the scientific worldview', this sounds about right for the second-philosophical Native,¹⁷ but elsewhere

16. See the Thin Realism of [2011b].

17. Unless, of course, the 'unselfconsciously' here prevents the Native from subjecting her beliefs and methods to critical scrutiny. Presumably, the 'scientific worldview' must include this sort of self-consciousness.

we find descriptions that seem slightly off: van Fraassen's Native embodies 'the common sense of the educated layperson in our culture today' (p. 66); she's 'the "model reader" of the texts that the sciences offer the public' (p. 67, see also p. 68). Here the Native is pictured as standing apart, being guided by what she regards as 'scientific'; we hear of 'the relevant scientific community to which [the Native] defers' (p. 84). An odd consequence of what might seem a trivial deviation is that the Native is now compared to a scriptural fundamentalist; the utterances of scientists are to be taken as gospel, at face value, with no room for further investigation.¹⁸ The distance between this figure and the second-philosophical inquirer I've described should be obvious: the Second Philosopher doesn't defer, she simply speaks as a scientist, citing evidence and subjecting her beliefs to scrutiny. When asked why she believes in atoms, the second-philosophical Native doesn't say 'because the scientists say so'; she says 'because of this evidence', citing results from Brownian motion, cloud chambers, electron microscopes, and so on. Let's reserve the term 'Native' in what follows for the occupant of this stance.

So what distinguishes van Fraassen's other figure, the 'Naturalistic Philosopher'? She's the one who describes the Native, but, of course, this isn't enough to set her apart—a person can describe herself. Her defining characteristic, as far as I can tell, is that she rejects all forms of transcendentalism,¹⁹ that is, views that posit a second level of inquiry standing over and above, separate from, that of the Native. (Kant's critical philosophy is the prime exemplar.) I can see why van Fraassen might think this rejection can't be the work of the Native—she has no tools to address positions that explicitly set her entire framework aside—but this is only a problem if the 'rejection' of these views is understood to involve a direct counterargument. In fact, all that's intended is what van Fraassen calls the Native's 'disturbingly uncomprehending stare' (p. 82), though there's more to it than that. The template for the Native's interaction with the transcendental philosopher

18. See van Fraassen [2015], pp. 81–82. For example, van Fraassen suggests that this leads the Native to misunderstand Dumas's remark that 'never in chemistry must we go beyond the realm of experiment' as expressing a 'positivist philosophical prejudice'. I don't see that I myself attributed such a prejudice to Dumas (see [1997], pp. 136–137), but the current point is that the Native is perfectly capable of assessing the historical context of a past scientist's remark and understanding it accordingly.

19. See van Fraassen [2015], pp. 66–67.

goes roughly like this: the transcendentalist explains that his inquiry is different from, disjoint from, the Native's empirical study; the Native asks what this special inquiry is designed to find out and what methods it uses to do so; when the answers to these questions are unsatisfying and/or unconvincing, the Native sees no reason to sign on to the transcendentalist project. That's what the 'uncomprehending stare' of the 'rejection of transcendentalism' comes to—and it falls well within the Native's capabilities. If van Fraassen's Naturalistic Philosopher is to attempt something stronger, a direct refutation of transcendentalism, then she goes beyond the powers of my Second Philosopher.

One reason, then, that van Fraassen feels the need to distinguish the Naturalistic Philosopher from the Native is an overly strong understanding of what the Naturalistic Philosopher does, but the Native is also being understood as too weak: the 'Native is no philosopher, and cannot be confronted with philosophical questions' (p. 68). On the contrary, as noted above, I've argued at length elsewhere that the Native will eventually face, and even answer, many traditionally philosophical questions. So, in the end, I remain unconvinced of the need to distinguish the Native from the Naturalistic Philosopher, unconvinced of any 'studied ambiguity' or unstable 'vacillation' between the two (pp. 68, 84). Still, there does remain one stark difference between me, the author, and the Second Philosopher I describe: she's highly idealized; she knows so much more and is so much smarter than I am!

(iii) *Interpretation*

In van Fraassen's eyes, the central shortcoming of the Native is that she is incapable of 'interpretation'. He freely admits that she doesn't take 'a completely uncritical attitude' (p. 84) toward her current beliefs and methods, that she assesses and corrects them as she goes, but interpretation is somehow more than, or different from, just that. What she's incapable of doing, according to van Fraassen, is 'bracketing' something that she currently accepts. This sounds at first like the claim that the Native can't call any of her well-confirmed beliefs into question, which is obviously too strong,²⁰ but van

20. This would seem to be the upshot of the lifejacket example (van Fraassen [2015], p. 85).

Fraassen's intentions come clear when he describes the new challenges that arise when we bracket:

questions as to how to understand . . . scientific activity, of the criteria of success apparently applied in intra-community assessment of the work in that area, and therefore of its aim. (p. 85)

Unsurprisingly, the two salient options for that aim are truth and empirical adequacy. This is where the realism/instrumentalism debate is joined.

At this point, van Fraassen is careful to distinguish the Native from other exemplars of what he terms 'naturalism'. He bemoans her inability to see that the Empiricist—in claiming the aim of science is empirical adequacy, not truth—isn't raising a scientific question about, say, the evidence for the existence of atoms based on Brownian motion. In fact, this way of (mis) understanding Empiricism is just the Native's attempt to distill an intelligible question out of a transcendental debate.²¹ As we'll see in more detail in connection with Stanford, there were those in Perrin's day who doubted we could ever know unobservables and who regarded atomic theory as a useful fiction. When those doubts are removed by Perrin's experimental successes, the Native takes it that a realism/instrumentalism debate integral and essential to scientific activity has been resolved—science itself would be hamstrung if this sort of doubt and resolution were relegated to some extra-scientific realm of evaluation. So, the Native understandably imagines that when van Fraassen denies we know there are atoms, he's raising a similar doubt, but that he's doing so now, when there's abundant rationally compelling evidence for the existence of molecular, atomic, and even sub-atomic structures. Understandably, she sees him as challenging the epistemic force of that evidence and sets out to defend it.

Van Fraassen contrasts this slow-witted reaction on the Native's part with that of the 'naturalist' Stephen Leeds, who addresses the Empiricist directly, on his own terms. As understood by van Fraassen, Leeds argues that since

21. As van Fraassen [2015], p. 87, recognizes: 'the only question . . . that makes sense to the Naturalist Native is the factual question whether unobservable entities such as atoms or molecules exist. That is a question she takes to have been addressed, and answered, by scientists themselves.'

inference to the best explanation is a principle of scientific inquiry, it should be applicable, too, in the interpretive mode to establish the correctness of Realism as opposed to Empiricism (p. 79). Obviously, this sort of thing falls nowhere near the Native's repertoire; she doesn't even regard inference to the best explanation as conclusive in the first place!²²

Still, some features of other naturalisms do seem to bleed over into van Fraassen's treatment of the Native. For example, he emphasizes Leeds's insistence that there's no need to justify or validate scientific inferences (p. 79), then goes on to claim that despite the Native's reliance on observation, 'she sees no need to investigate the relation between *observation* and the final answer as to what is so or not' (p. 82). I'm not sure what to make of 'final' here, but the Native certainly does engage in serious investigation of how the senses register information about the world, of when they tend to be reliable and when not—for the case of vision, this is one of the foci of contemporary vision science. More dramatically, van Fraassen takes Arthur Fine's Natural Ontological Attitude to entail that it makes no sense to consider alternative understandings of quantum mechanics, then remarks that this is '[u]ndoubtedly the Naturalist Native's reaction as well!' (p. 83). But it isn't. Offering an account of how quantum mechanics relates to the world is only a more elaborate cousin of the earlier question about whether atomic theory describes actual physical entities. This is the sort of criticizing and assessing that the Native does all the time, the sort of thing that she offered up as a possible understanding of van Fraassen's 'interpretation' and that van Fraassen rejected.

And this is the point, really. 'Interpretation', for van Fraassen, isn't part of the Native's ordinary empirical inquiry. It's a reflection on that inquiry, not from within, but from a point outside.

22. Section II.6 of [1997] makes a case that inference to the best explanation, along with Quinean holism, is not a principle of scientific inquiry. Perhaps it's also worth noting that it's possible to think the aim of science is truth without thinking our best theories aim to be 'true throughout' (van Fraassen [2015], p. 86).

(iv) Transcendentalism

This brings us to van Fraassen's transcendentalism, a label he appears to embrace without protest in this passage:

Maddy goes on to display her Naturalism as a recurring revolt against transcendentalism, in a broad sense—against 'two-level' philosophical views, citing Kant, Carnap, and me as the main examples. (p. 66)

What makes a view 'two-level' for me is its methodological independence from ordinary science. Van Fraassen isn't offering a direct critique of Perrin's experiments²³ or of the scientific community's subsequent embrace of atomic theory. His objection, rather, is to Perrin's prose and that of those around him, which would include Einstein, Poincaré, Ostwald, and the Nobel Prize committee:

Such pronouncements are important for the historian, to indicate the terms in which such episodes were discussed, but we must always keep in mind that these words do not come in the context of a philosophy seminar, where our distinctions are made, or the conceptual problems are disentangled in the way we do. (van Fraassen [2009], pp. 22–23, footnote 20)

Viewing the matter in this special philosophical way, one comes to appreciate that what was established wasn't 'the reality of molecules' (*ibid.*, p. 22) but the bankruptcy of

23. A possible exception is his suggestion that understanding Perrin's results 'as providing evidence for the reality of molecules . . . is a rather strange reading' partly because his 'research was entirely in the framework of the classical kinetic theory in which atoms and molecules were mainly represented as hard but elastic spheres of definite diameter, position, and velocity' (van Fraassen [2015], pp. 22–23, see also van Fraassen [2009], p. 8). If the scientific use of idealizations were enough to make the Empiricist's case, the debate would be over before it started! What Perrin and company claim isn't that every detail of atomic theory is correct—they're well aware that it involves these false assumptions—but that the model it describes resembles the world in important ways, most importantly, in the particulate nature of matter. (See previous footnote.)

the idea that it might be good for physics to opt for a different way of modeling nature, one that rivaled atomic theories of matter. (Ibid., p. 23)

But this rejection of the prose makes no difference in practice: the Native, inclined to speak, and dare we say believe, with Perrin and company, is left to carry on as before.²⁴

This returns us, at last, to the question whether there's room for rational adjudication of the case between the Native and the Empiricist. Van Fraassen admits that 'it is still possible [to] read [Perrin's] results as providing evidence for the reality of molecules' (van Fraassen [2009], p. 22), which might suggest that this is a dispute between stances with no definitive resolution, but I suspect that in fact this possible reading of Perrin's experiments isn't being given by the Native, but by those who share the seminar room with van Fraassen—Realists like Leeds, for example, who directly oppose the Empiricist there. This comes out more clearly in his earlier extended analysis of Perrin's work: there he debates Salmon, Glymour, and Achinstein, but remarks that 'Maddy takes for granted . . . Perrin's reasoning . . . and does not offer a competing account to these' (van Fraassen [2009], p. 6, footnote 4). ('Taking for granted' here recalls the earlier 'deferring'; we've already noted that the second-philosophical Native doesn't defer.) Of course, the Native doesn't simply offer Perrin's evidence uncritically any more than Perrin himself did—it's rightly submitted to all appropriate scientific scrutiny—but it is true that whatever assessing gets done is simply part of scientific activity. Though van Fraassen likens the Native to the Kantian 'transcendental realist',

24. Imagine this debate between Perrin, standing in for the second-philosophical Native, and van Fraassen. Perrin claims to have strong evidence for the existence of molecules. Van Fraassen disagrees: what Perrin has done is establish the 'empirical grounding' of atomic theory. Perrin responds that, yes, before we had evidence only from large collections of molecules; I figured out how to get direct empirical access to the individual behavior of molecules, and along the way verified the most unlikely aspect of that behavior, namely, the random walk. Again van Fraassen disagrees: what you and your coworkers did was to develop both theory and measurement procedures so as to enhance the empirical adequacy of atomic theory. Perrin wonders why van Fraassen is so keen on this strange way of describing the situation—and here van Fraassen's reasons will have to do, not with the specifics of Perrin's work, but with his Empiricist stance and his seminar-room debate with the Realist. From that point of view, van Fraassen's instrumentalist would say that Perrin is just wrong about what he's accomplished; the realist would say that he's right, but that he doesn't fully understand the reasons why. However, they would both heartily agree that his science is fine, even exemplary, quite unaffected by his errors in the seminar room.

I think the label applies more properly to those Realists who undertake to rebut the Empiricist on his own terms—these are the thinkers van Fraassen takes himself to have wrestled to something like a draw (though ‘in retrospect’, theirs is ‘rather a strange reading’ (ibid, p. 22)). Despite the Native’s claim to have good evidence for the existence of atoms, the term ‘realist’ as it’s used in this debate doesn’t appear to apply; Realists, properly so-called, reside in the seminar room, engaging directly with Empiricists about matters that transcend ordinary science. The Native and van Fraassen’s Empiricist aren’t in direct conflict after all; no adjudication is required.

Notice, by the way, that from the Native’s point of view, this Realist is almost as mystifying as van Fraassen’s instrumentalist. Faced with evidence from Perrin and his successors, both seem to think that it isn’t enough, that it requires supplementation (the sort of thing Leeds’s use of inference to the best explanation in the seminar room is intended to provide); they just disagree on whether or not that supplementation can be provided. The Native naturally wonders what’s wrong with her evidence as it stands. This isn’t to say that she doesn’t subject that evidence to critical scrutiny—we’ve seen that she does, just as Perrin and his contemporaries did, just as scientists do all the time—but she doesn’t insist, as the seminar-room Realists do, on some peculiarly philosophical line of assessment. She attends, rather, to the specifics of her particular case: were the experimental protocols sufficiently stringent? Were the calculations correct? Were all relevant variables accounted for? In the end, both sides of this transcendental debate escape her.

One last note. I’ve claimed that despite her inability to cotton on to the discussion in the seminar room, the Native—or Second Philosopher—still manages to address many traditional questions about skepticism, logic, language and the world, arithmetic and higher mathematics. *Second Philosophy* and subsequent writings aim to demonstrate just how much she can do. It’s hard to know what to call these inquiries if not ‘philosophy’—maybe ‘Natural Philosophy’ in something like the sense of the early moderns?²⁵ In any case, if there remains any doubt that I, the author of these works, function as a (sadly limited) Second Philosopher, I can report that my own reaction to transcendentalism in general, and van Fraassen’s transcendentalism

25. See Essay #1.

in particular, is the same as hers: I can't see what the point of the project is supposed to be or what methods it might properly employ.

2. STANFORD'S EPISTEMIC INSTRUMENTALISM

In sharp contrast to van Fraassen, Stanford is no transcendentalist. His views have evolved over the years—I hope to trace the broad outlines of that development—but his naturalism has been constant, beginning in his influential book, *Exceeding Our Grasp*:

I am not reaching beyond or outside of science itself for evidence of some supposedly higher or purer kind with which to sit in global judgment on the scientific enterprise as a whole. With those philosophical naturalists who emphasize the essential continuity of philosophical and scientific efforts to acquire knowledge, I hold that there is only good and bad evidence, not higher and lower evidence or scientific evidence and some other kind. Indeed, I expect my argument to be congenial to . . . naturalists of this sort. (Stanford [2006], p. 37)

Later he proposes, more specifically, what he calls 'integrative naturalism':

For . . . integrative naturalists, understanding what our best scientific theories are telling us about the world and understanding how we go about entheorizing that world in the first place are not distinct challenges: both are part of the overarching and more fundamental challenge of trying to *simultaneously* understand both the world and our own place within it . . . That project has plenty of moving parts, from developing new statistical methods and quantitative measures of confirmation to building better microscopes, and from proposing and testing novel empirical hypotheses to interpreting the historical record of scientific inquiry, but these are all parts of a single interconnected inquiry. (Stanford [2016], p. 93)

So Stanford begins from a stance very like that of the Second Philosopher.²⁶

From this perspective, Stanford mounts an original historicist critique of scientific realism and replaces it with his own ‘epistemic instrumentalism’. Crucially, his characterization of the line between those parts of our theory that qualify as literal and those merely instrumental makes no appeal to a notion of observation. Typically, the instrumental parts will concern domains where

it is difficult to acquire information . . . because the entities inhabiting them are too small or too large or too amorphous for us to readily perceive; because the causal interactions between those entities are too fast or too slow or too rare or take place on too grand a scale for us to engage with in ordinary ways; these entities and interactions occur in times and places either far removed from our own or otherwise inconveniently located (e.g., at the dawn of life on Earth, in remote regions of the universe, at the center of the Sun), and so on. (Stanford [2006], p. 3)

But membership in such a domain, by itself, isn’t enough;²⁷ for example, though dinosaurs admittedly ‘roamed the earth long ago,’

fossilization is not a hypothetical postulated mechanism but a process we can study in action and we simply have no specific reason to doubt that its products in the remote past were any different from its present ones. (Ibid., p. 33)

The determining feature, then, is the kind of evidence we have for our account of a given theoretical posit—the topic of (i), below. For now, it’s enough to note that the Second Philosopher’s theory of atoms, based on the evidence

26. As he acknowledges (see Stanford [2016], p. 93, footnote 1).

27. See, e.g., Stanford [2006], p. 35: ‘we neither invariably reason eliminatively about unobservables, nor invariably reason otherwise about observables, and it is the application of eliminative inference outside its domain of reliable operation, not observability as such, which represents a legitimate source of concern.’

of Perrin and his successors, concerns a domain ‘too small . . . for us to readily perceive’ and clearly qualifies as a potential target of Stanford’s critique.²⁸

I think it will help clarify the contours of Stanford’s historicist critique and his epistemic instrumentalism if we consider a historically located version of the Second Philosopher, an embellished version of Perrin, newly in possession of the evidence described in his *Atoms* (Perrin [1913]). At the turn of the 20th century, atomic theory was fundamental to chemistry, and kinetic theory had flowered in physics.²⁹ Still, there were doubts about the reality of atoms, as attested by this remark in a leading textbook by Wilhelm Ostwald, one of the founders of physical chemistry and an eventual Nobel Prize winner:

[T]he atomic hypothesis has proved to be an exceedingly useful aid to instruction and investigation . . . One must not, however, be led astray by this agreement between picture and reality and combine the two. (Ostwald [1904], p. 151)³⁰

The copious evidence for atomic theory at that point was all abductive: the theory as a whole was explanatorily and predictably powerful in dramatic ways, but the tiny molecules in endless random motion that it posited were entirely theoretical, entirely inaccessible.

At the time, Perrin agreed that ‘the skeptical position . . . was . . . legitimate and no doubt useful’ (Perrin [1913], p. 216), and so did Einstein, who set out to correct the situation:

My major aim in [developing ‘the statistical mechanics and the molecular-kinetic theory of thermodynamics’] was to find facts which would guarantee as much as possible the existence of atoms. (Einstein [1949], p. 47)

28. See Stanford [2006], p. 210, where ‘rocks are made up of atoms with a specific internal composition’ is offered as an example of a belief that separates the realist from his instrumentalist. I worry a little that ‘with a specific internal composition’ might commit the realist to every detail of contemporary atomic theory (something Perrin* avoids in the description below), but let’s set this aside for now.

29. I discuss this case in more detail in [1997], pp. 135–143, and [2007], pp. 404–407.

30. Quoted by Richard Miller [1987], p. 473.

What Perrin accomplished was to design and carry out experiments so precise that this restless motion was revealed for the first time in the movements of his tiny manufactured particles—all in perfect harmony with Einstein's detailed predictions. Ostwald and other skeptics were convinced.

Thus the historical Perrin, the basis for my imaginary, fully second-philosophical Perrin*. This figure is a fallibilist to begin with—he makes no claim to certainty about anything—and he realizes, of course, that theoretical science—the sort of thing the instrumentalist calls into question—is more epistemically risky than, for example, ordinary perceptual belief. In the particular case of atomic theory, he worried at one time that despite its immense success, it was in some sense unverifiable, the domain it describes ultimately inaccessible. But he now believes that he has a new and decisive kind of evidence—a means of access, of direct verification—based on his experiments and other considerations available at the time. This doesn't mean that he believes every part of the theory he embraces—it naturally involves idealizations and simplifications, plus a few somewhat dicey assumptions³¹—but he does at least tentatively believe that he's closed the case for the particulate structure of matter in the small and its accompanying randomness.

Unlike van Fraassen's Empiricist stance, Stanford's arguments come from within integrative naturalism, so they ought to engage directly with Perrin*, they ought to present considerations he'd be duty-bound to address. My plan, then, is to confront this figure with Stanford's historicist challenge.

(i) Stanford's historicist challenge

Any gloomy appeal to the historical record naturally begins with the classic pessimistic induction: many successful scientific theories of the past have turned out to be false; why should we expect our own apparently well-confirmed theories to be any different? Despite his sympathy for this line of thought, Stanford believes 'there is at least some justice'³² to the reply that

31. See van Fraassen [2009], p. 8: "Perrin worked throughout with the 'billiard ball' version of the kinetic theory. In his models, molecules are perfectly hard, perfectly elastic spheres, taken as relevant approximation." He points out that Rutherford's work on atomic structure was known at the time.

32. See Stanford [2006], 2.3. The quotation is from p. 44.

our best current theories have important virtues that distinguish them from the failed theories of the past, so he offers a critique that rests on the failures of theorists rather than the failures of theories. His thought is that even if our current theories are better in significant ways, it's unlikely that our current theorists—mere humans, after all—are significantly better than those of the past.

To isolate the systematic failure of scientists in the problematic domains, Stanford suggests that the evidence we have for many of our theories has a distinctive character:

[T]he reasons we can offer for believing [their claims] would seem . . . to be limited to the fact that each of the fundamental hypotheses in question offers the most powerful and convincing systematic account we have for explaining, predicting and intervening with respect to a wide range of empirical phenomena . . . and we can neither offer nor even imagine any alternative hypothesis whose performance in these respects would be equally impressive. (Stanford [2006], p. 34)

This is precisely the kind of holistic, abductive evidence that Perrin and his peers considered insufficient in the late 19th century. In Stanford's telling, this type of inference to the best explanation, as it's also called, consists of an eliminative induction to the last theory standing.³³ The trouble with this—and here we have Stanford's fundamental point—is that the range of alternatives we've eliminated includes only those we've been clever enough to think up, and the historical record suggests that scientists of the past very often didn't think up all the relevant alternatives; indeed they missed alternatives that ended up being accepted later on in the progress of science.³⁴ Given it's unlikely that humans are more imaginative now than they were in the past, Stanford presents Perrin* with the kind of ground-level challenge

33. I use the terms 'abductive inference', 'inference to the best explanation', and 'eliminative induction' interchangeably in what follows.

34. For the record, Stanford doesn't impugn all abductive inferences in science, only those for which we lack a clear overview of the space of possibilities (see Stanford [2006], p. 32). This happens most often in the problematic domains, and since our focus will be on one of these, the case of atomic theory, I won't continue to insert this caveat.

that didn't turn up in van Fraassen's constructive empiricism: in problematic domains, which include the micro-structure of matter, human theorists are prone to misapply eliminative induction—what makes you think you aren't doing the same? In other words, what makes you think your new evidence is of a qualitatively different kind?

One line of defense against Stanford's argument would be to argue that his examples aren't representative, that they're all three drawn from the early history of genetics, before that science had found its footing,³⁵ but I imagine Perrin* turning instead to his own house, asking whether his inference is badly flawed in the way Stanford has identified. Answering this question properly would require careful examination of the reasoning of Darwin, Galton, and Weismann in Stanford's examples and even-handed comparison with the detailed structure of Perrin*'s evidence and argumentation. Without attempting anything like this here, I imagine Perrin* might at least begin his response by pointing to a few disanalogies.

The biologists in question were contemplating an ill-understood phenomenon, heredity, and searching for a reasonable theory that would explain it; this is the kind of situation that invites an eliminative induction or an inference to the best explanation and thereby risks the kind of error Stanford identifies. In contrast, Perrin* has a well-developed theory in hand, a theory that has been effectively explaining and predicting chemical and physical phenomena since Dalton's work a century prior. Perrin* isn't pulling a new theory out of the air to explain Brownian motion; he's using the case of Brownian motion to put a highly successful existing theory to the most stringent test available, a test Einstein had developed for precisely that purpose.

Some realists have taken a different line of defense, arguing instead that the scientists in Stanford's examples weren't so badly wrong, after all, that they were right about certain central features of their theories and the rest was irrelevant. One way to execute this strategy would be to argue for a general criterion that allows us to differentiate the confirmed parts from the rest. Stanford points out that such efforts often depend on considerable

35. E.g., Godfrey-Smith [2008], p. 142, thinks the historicist critique should rest on 'mature' theories, which classification wouldn't include the genetics of Stanford's period. Magnus and Callender [2004] raise doubts about the comparison class from which the historicist's examples are supposed to be drawn.

hindsight: we identify these confirmed parts by looking back at a theory and selecting those that coincide with our more mature understanding of the situation. But this is unhelpful, Stanford argues, because what the realist needs is a way of telling which parts of our current theory are confirmed, and that requires a criterion that can be applied without the benefit of this kind of Monday morning quarterbacking.

Now we've seen that Perrin* doesn't take all details of atomic theory to be confirmed by his experimental results—there are all those idealizations and simplifications and possibly dodgy hypotheses—but he does take them to confirm the particulate nature of matter in the small and its accompanying randomness. He offers no general criterion for identifying the relevant morals of any other scientific theory—if I were to put a few more words in his mouth, I'd say he was skeptical that there is such a thing (see (iv), below)—but for now, what's of interest is that there is a participant in Stanford's realism/instrumentalism debate, Stathis Psillos, who represents something that sounds compatible with Perrin*'s stance. Stanford summarizes Psillos's 'selective realism' like this:

[His] intriguing alternative approach promises to evade these persistent problems by avoiding the need for any *explicit* criterion of selective confirmation at all . . . he seeks to finesse this problem by arguing that working scientists *themselves* routinely judge different parts, features, or aspects of extant theories to be differentially confirmed by the empirical evidence and that the historical record shows these judgments to be generally reliable. If so, we can . . . safely rely on *scientists' own* judgments in identifying the selectively confirmed, trustworthy aspects of existing theories. (Stanford [2006], pp. 173–174)

Stanford responds to this with a second historical induction whose goal is to establish that scientists' own judgments in such matters are not in fact 'generally reliable'.

Viewed from Perrin*'s perspective, this presents a fresh challenge: not 'given Stanford's historical examples, am I relying on the same sort of flawed eliminative induction?' but 'given more of Stanford's historical examples, am I making the same sort of mistake about which parts of atomic theory my

evidence confirms?’ Once again, there’s room for a general argument about whether Stanford’s examples are representative, but as before, let’s suppose Perrin* looks to his own house: can he distinguish his case from those Stanford cites? As before, I won’t attempt to answer this question in any detail, but a quick review of the examples—19th-century vitalists, Maxwell’s defense of the ether, Lavoisier’s insistence on caloric fluid—finds repeated arguments, not only that such-and-such exists or happens, but that an alternative is inconceivable. So, for example, how could something be transmitted from here to there without some ‘medium or substance in which [it] exists after it leaves one body and before it reaches the other’?³⁶ The quasi-conceptual flavor of these discussions contrasts sharply with Perrin*’s circumstances: he’s open not only to the conceivability but to the real possibility that matter isn’t particulate; that’s why he (along with Einstein) is so keen to find a test! So I think there is room for Perrin* to reply to this historicist challenge as well.

The important point in all this is that Stanford’s epistemic instrumentalist has issued a direct challenge to Perrin*—show that your evidence is truly different in kind from the previously available purely abductive evidence—and this challenge is one that Perrin* would find both understandable and appropriate. In this way, Stanford, unlike van Fraassen, opens a direct dispute over the epistemic force of Perrin*’s experimental results. I take this ground-level dispute to be the crux of Perrin*’s confrontation with the epistemic instrumentalist. My goal here isn’t to mount a case in favor of Perrin*’s side of the argument, though I believe there is one to be made and I often gesture in that direction. Rather—to anticipate—it seems to me that, contrary to what one might expect, this debate, this crux, often isn’t Stanford’s primary focus, and much of what follows is an effort to trace how and why this is so. For now, let me just note that this isn’t the issue Stanford forefronts at this point in *Exceeding Our Grasp*.

Instead, Stanford focuses on the claims of his opponents in the professional literature, in particular, on the realist who appeals to some refinement of the Success Argument:

36. This comes from Maxwell on the ether, quoted by Stanford [2006], p. 152.

[T]he only satisfactory explanation for the success of our scientific theories is that they are (at least approximately) true . . . any other view of the matter leaves it a complete and utter miracle why our best scientific theories are so successful. (Stanford [2006], p. 6)

Such a realist believes in atoms because they are (the right kind of) posits of (the right kind of) successful scientific theory. Obviously this is not Perrin*—it was his own specific experimental findings, not a claim about scientific theories in general, that resolved the issue for him.

Thus, Stanford's historicist challenge to Perrin* and to the scientific realist. Let's now consider what he takes to follow from his critique.

(ii) *The moral*

To orient ourselves on this question, let's first consider what moral Perrin* would draw if he were convinced that his conclusions rested on one of Stanford's bad eliminative inductions, on an open-ended inference to the best explanation, that his evidence for atomic theory consisted exclusively of its explanatory and predictive powers, the interventions it makes possible, and our inability to think of anything else that could do as well. How should he react? From his perspective, all he's accomplished is to add more evidence of the same kind as he had before; accepting Stanford's critique would mean that he and Einstein had failed to achieve their goal of establishing the reality of atoms. Presumably, at that point, he would try again. Nothing in Stanford's position rules out the possibility that he might eventually find the right kind of evidence, evidence as effective as fossils are for our theory of dinosaurs.³⁷

Surprisingly, Stanford doesn't see the situation this way:

This does not mean the moral suggested here is that we must somehow constrain and regulate the inferences we draw in the course of our scientific theorizing by some perfectly abstract and general commitment to the

37. What would Perrin* do if all these efforts failed? What if it appeared that atomic theory was wildly successful, but that there were no atoms? I submit that Perrin* would then seek some other explanation, some other, more subtle relation between the theory and the world, to explain the theory's success. See the M/W principle in (iv), below.

likely existence of completely unspecified but serious unconceived theoretical alternatives. What could this amount to but . . . a sure recipe for inferential (and therefore conceptual) paralysis? (Stanford [2006], p. 135)

He would deny that Perrin* was ‘somehow irresponsible or careless’ or ‘made reckless inferences’ (ibid., p. 134), or that his instrumentalism is a ‘narrow and defensive creed,’ as Popper once claimed (ibid., p. 209):

The epistemic instrumentalist will insist no less than the realist that we continue to challenge our best scientific theories by uncovering and testing further empirical implications they have, that we use them to unearth new phenomena and new ways to predict and intervene in the course of events around us, that these theories serve as the appropriate starting point in trying to determine how they themselves can be refined, improved, and developed . . . She will not . . . pursue the further implications of our theories less doggedly, or invest those implications with less significance, than the realist . . . In short, the instrumentalist is in a position to take the claims of our best scientific theories about nature every bit as *seriously* as the realist does. (Ibid., p. 210)

The recommendation seems to be that Perrin* behave just as he would have done if his inference to the reality of atoms had been sound.

Having ruled out Perrin*'s reaction—back to the drawing board!—Stanford considers two other candidate responses to the fundamental error he's uncovered. One is ‘a vapid agreement to tack the phrase “but of course there may be something I haven't thought of yet” piously and toothlessly onto every conclusion we draw’ (Stanford [2006], p. 135). Oddly enough, this sounds like what Perrin* already does, given his general fallibilism and his awareness of the heightened epistemic risk involved in theoretical science; he says this about his conclusion that atoms are real, despite being unconvinced by Stanford's arguments that his evidence is weaker than he thought.

Whether this ‘pious’ addendum is ‘toothless’ is a question we'll get to in a moment, but as the rhetoric makes clear, it isn't Stanford's choice. He prefers,

instead, that we adjust our attitude toward our theories, that we recognize that scientists are ‘not entitled to believe the conclusions of their eliminative inferences’, that we restrict our belief to consequences of the theory that ‘can be understood independently of the theory . . . toward which we are adopting an instrumentalist stance’ (Stanford [2006], p. 197). This instrumentalist stance is nothing exotic, he tells us; it’s the same as the contemporary realist’s attitude toward Newton’s theory of gravitation when ‘she allows that we can perfectly well make use of [it] to send rockets to the moon without believing . . . [that] gravity is . . . a force exerted by massive bodies on one another, [that] there is . . . absolute space or time, and so on’ (ibid., p. 204). The contemporary realist is also an instrumentalist about quantum mechanics, except that in this case we have no more inclusive theory that explains its success. The instrumentalist and the realist both take instrumentalist stances toward some theories and realist stances toward others; they just draw the line in different places.³⁸

(Notice, by the way, that here Stanford places the realism/instrumentalism debate within science, as van Fraassen does not. In particular, presumably Stanford would take Perrin* and Ostwald to have been instrumentalists about atomic theory before the decisive experiments and realists after; they simply redrew the line in light of new evidence. Here the ground-level debate comes back into view: was Perrin’s evidence sufficient to make this move rationally compelling?)

At this point, it’s hard not to worry that Stanford’s withholding of belief has no more teeth than Perrin*’s fallibilist scruples, that he’s simply being asked to add a pious ‘but I don’t believe it’ instead of ‘but I could be wrong’. And Stanford does worry.³⁹ There are hints of a response in *Exceeding Our Grasp*,⁴⁰ but the fully developed version doesn’t appear until the aptly named ‘Catastrophism, uniformitarianism, and a scientific realism debate that makes a difference’ (Stanford [2015a]).

Stanford borrows the terms ‘Catastrophism’ and ‘Uniformitarianism’ from geology, but in his usage, Uniformitarians ‘hold that the future of the

38. See Stanford [2006], pp. 204–205.

39. Recall that there’s no such worry for Perrin*: as we’ve seen, the success of Stanford’s argument would have enough teeth to send him back to the drawing board.

40. See Stanford [2006], pp. 209, 211. Also [2014], p. 124.

scientific enterprise will be characterized by revolutions and transformations every bit as profound and consequential as those we find throughout its history', while Catastrophists 'believe that truly profound and fundamental revisions to our scientific understanding of the world are now largely confined to the past' (Stanford [2018], pp. 214–215). His thought is that the instrumentalist is a Uniformitarian, while the realist 'thinks that contemporary theories have things sorted out at least roughly right and that our remaining errors are simply errors of detail' (Stanford [2015a], p. 874) and is thus apparently a Catastrophist. Catastrophists, unlike Uniformitarians, he takes to be theoretically conservative by nature—that is, disinclined to pursue 'transformative' or 'revolutionary' science. And here, Stanford concludes, we have a difference that makes a difference.

Obviously, much depends on how we understand these revolutionary changes. Stanford doesn't attempt a general characterization or posit any Kuhnian incommensurability. Instead, he takes the admirable path of explicating the notion by what he calls 'historical ostension' (Stanford [2018]); that is, he gives us examples of past theories that were superseded by such changes—Newtonian mechanics, Dalton's atomic chemistry, Lavoisier's caloric theory of heat, Maxwell's account of electromagnetic ether—all of which he takes to be 'radically false', 'fundamentally and profoundly mistaken', not 'even approximately true'.⁴¹ Stanford assumes the reader shares his dismay at the defects of these theories—that gravity turned out not to be a force, that atoms turned out to have substructure, that there is no subtle fluid behind heat phenomena or mechanical medium for the transmission of electromagnetic waves—so that we would count no one a realist who thought her current best theory is akin to these. But Perrin* would be delighted if his theory turned out to give us as much fresh insight into the workings of the world as Newton's or Dalton's or Lavoisier's or Maxwell's!⁴² This reaction is now commonly described as the claim that these theories weren't really so terrible after all, that they were in fact 'approximately true'.

41. Descriptions like these appear throughout, but these three are from Stanford [2014], p. 117, Stanford [2006], p. 157, and Stanford [2014], p. 117, respectively.

42. (iii), below, expands on this reaction.

Here the debate threatens to devolve into one over the viability of an appeal to ‘approximate truth’, but Stanford argues that this is a distraction. The real disagreement remains that between Catastrophists and Uniformitarians, because any realist who professed also to be a Uniformitarian, who embraced the idea that her theory is comparable to the cited theories of the past, would be indistinguishable from an instrumentalist inspired by the historical critique (Stanford [2015a], pp. 876–877). Perrin* does believe that scientific progress will produce changes like those in the cases Stanford cites; this realization informs his conviction that theoretical science is especially epistemically risky. It would seem to follow that, despite his confidence in the particulate structure of matter, Perrin*’s ‘realism’ is indistinguishable from ‘instrumentalism.’ Fortunately, this puzzle can be set aside for now because Stanford comes to concede that a realist can consistently embrace Uniformitarianism. So this effort to locate ‘a difference that makes a difference’ fails (though the charge of conservatism eventually reappears in the new context (see (iv))).

For what it’s worth, it seems to me that there may well be real methodological differences between Perrin* and the epistemic instrumentalist:⁴³ first, if Perrin* has a theory incompatible with another well-confirmed theory, he counts this as evidence against his theory until a way can be found to reconcile them, and second, when Perrin* arrives at a new theory by a strong inference to the best explanation, he will want to know why this theory is so successful. (In the case of atomic theory, the second of these took the form of a concerted effort to find a stringent test for particulate structure and random behavior.) Stanford considers the first in passing, coming to what seems to me an ambivalent stance,⁴⁴ and the second reappears in (iv), below

43. Recall that Perrin*, as we’re understanding him, doesn’t accept Stanford’s critique of his evidence. If he did, then (as noted above) there’d be a clear methodological difference: Perrin* would go back to the drawing board, while the epistemic instrumentalist would simply withhold belief.

44. See Stanford [2006], pp. 209–210: ‘the epistemic instrumentalist will insist . . . perhaps even that we work to unify our various scientific theories with each other and with whatever else we believe to produce a single coherent, consistent, and systematic account of the natural world as a whole’. But then, in a footnote to this passage (*ibid.*, p. 213), he admits that ‘the rationale for this . . . may not be entirely obvious, as instrumentalism seems to offer little epistemic motivation for insisting that our theories not contradict one another’. He concludes that ‘the instrumentalist’s motivation for consistency is pragmatic rather than epistemic’ and ‘of course the instrumentalist would seem better positioned than the realist to make sense of and live with the fact that our best theories seem at present to be neither fully mutually consistent nor maximally unified’.

(as the M/W principle). But before leaving this stage of the discussion, I'd like to make two observations.

The first concerns the route we've taken to arrive at this point. In his effort to give methodological teeth to 'but I don't believe it', Stanford has shifted his focus away from the distinctive critique of abductive evidence that separates his instrumentalism from a generic pessimistic induction and toward a more general feature that the two versions share: a distrust of current theories based on a negative assessment of past theories. The contrast between Uniformitarianism and Catastrophism rests on the identification of sample historical cases and two attitudes toward them—there will be more like these and there won't—where the former is taken (for now) to characterize instrumentalism in general and to preclude realism. This switch seems to me unfortunate, because it leaves behind the ground-level debate between Perrin* and the epistemic instrumentalist that we've been tracing. The question there is whether Perrin*'s specific evidence has the force he claims, not whether a selection of superseded theories from the past should give us pause about what we're tempted to believe now.

The second observation begins more locally, in Stanford's case against the selective realist. Here the challenge for the realist is to identify those parts of her current theory that she takes herself to be in a position to stand by, so to speak—in Perrin*'s case, the particulate nature of matter and its attendant randomness. My question is what this 'standing by' comes to. I've been understanding it to characterize those elements of our current theory that we take to be confirmed, that we take ourselves, now, to be rationally justified in believing. Sometimes, it's possible to read Stanford's text in this way, but more often he includes a predictive element: to stand by a claim is to predict that it will survive in future theories.⁴⁵ Before taking up the final, or at least the latest, turn in Stanford's overall line of argument, I'd like to pause to lay out my concerns about this way of putting the point. This will raise a general question about the historicist method, but this remains tangential to the crucial ground-level debate just recalled.

45. That is, sometimes what matters is 'whether we should *trust* or *believe* the accounts of otherwise inaccessible natural domains given by our best scientific theories'; other times what's involved is 'distinguishing the parts of present theories that will be preserved or retained in their successors'. (Both quotations come from Stanford [2006], p. 183.)

(iii) On historicism

Historically motivated instrumentalism obviously depends on a negative assessment of now-superseded theories of the past. When it comes to characterizing this attitude, we've seen that Stanford presents us with examples—'historical ostension'—and counts on us to share his negative assessment. Often enough, the assessment is that they are simply false, indeed 'radically false' (Stanford [2014], p. 117), but his thinking on truth and falsity takes a subtle turn in his recent, delightfully titled "'Atoms exist" is probably true, and other facts that should not comfort scientific realists' (Stanford [2015b]).

Stanford is considering a flat-footed objection:

Historicist critics of scientific realism are often met with a wry (even indulgent) smile and some version of the understandably incredulous inquiry, 'Surely you don't seriously doubt that atoms exist?' (Stanford [2015b], p. 5)⁴⁶

In response, he argues that contemporary confidence that 'Atoms exist' is true doesn't guarantee that earlier atomic theories weren't fundamentally and profoundly mistaken—thus able to contribute to the historicist critique—despite the implications of his cheeky interlocutors.⁴⁷ Along the way, he adopts a view about reference, and hence about truth, that seems to me problematic for a proponent of the historicist approach. This is what I'd like to explore.

Stanford opens this line of thought with a quotation from Howard Stein at his delightfully irascible finest:

46. I blush to admit that I recognize myself among these annoying interlocutors, though I don't suppose that I'm alone in this!

47. For what it's worth, I'm not sure these interlocutors are out to claim that our current acceptance of 'Atoms exist' undermines Stanford's disparagement of earlier atomic theories. Rather, their point seems to me to resemble a Moorean anti-skeptical gambit: you've given an argument, I don't care what, that implies we don't know that there are atoms; which is more plausible, your argument or the claim that we know there are atoms? Or perhaps more bluntly, an attempted reductio: your argument implies that we don't know there are atoms; can you really live with that consequence?

For my part, I throw up my hands at this: why should we say that the old term ‘ether’ failed to ‘refer’?—and that the old term ‘atom’ did ‘refer’? . . . our own physics teaches us that there is *nothing* that has *all* the properties posited by nineteenth-century physicists for the ether *or* for atoms; but that, on the other hand, in *both* instances, rather important parts of the nineteenth-century theories are correct. (Stein [1989], pp. 56–57)

Stanford admits that it’s hard to imagine any principled reason for this difference, going so far as to note that some decisions about terminological continuity seem to be inspired by public relations as much as anything else. Next comes Mark Wilson, making a persuasive case that we imagine our concepts equipped to make definitive classifications in novel cases, when often enough a new usage in fact arises from irrelevant historical contingencies.⁴⁸

Stanford’s diagnosis of this situation calls on a proposal of my own for a disquotational theory of reference.⁴⁹ The feature of this account that matters for his purposes is one of two diagnostics I use to distinguish disquotational from correspondence theories: when Pierre says ‘la neige est blanche’, is the truth or falsity of his utterance completely determined by facts about Pierre, his language, and his worldly surroundings? For the correspondence theorist, the answer should be yes; the truth or falsity of Pierre’s utterance depends only on the disposition of the worldly referents of his words. For the English-speaking disquotationalist, on the other hand, the truth predicate is native to English, so the French sentence has to be translated before ‘true’ can be applied to it either way—which means that his having said something true involves not only facts about Pierre, French, and his surroundings, but also about English and the context and vicissitudes of translation. In cases like Pierre’s, the proper translation is obvious, but our retrospective readings of past scientific theories are a more subtle matter.

Consider the much-discussed example of Joseph Priestley and his ‘dephlogisticated air’. Suppose he remarks that ‘dephlogisticated air is good for breathing’. To assess the truth of this utterance, the disquotationalist must first translate it into our current idiom—‘interpret’ is a better term (though

48. See Wilson’s indelible example of the Druids from his [1982].

49. See [2007], Part II, or more recently, Essay #6.

not in van Fraassen's sense!), since Priestley wrote in English—and here the choice of translation may depend on the context of translation: for beginning chemistry students, to impress them with Priestley's accomplishments, one might choose 'oxygen is good for breathing'; for historians of science out to track the development of ideas in early chemistry, the homophonic 'dephlogisticated air is good for breathing' would be more appropriate. Once the role of the translator's interests is recognized, any manner of other factors intrude; imagine the complex historical and sociological story we'd need to tell to account for the difference Stein notes between contemporary translations involving the 19th-century terms 'ether' and 'atom'!⁵⁰

Stanford adopts this point of view:

This sensitivity of our judgments of referential continuity to terminological decisions . . . helps to illustrate something that I think has been widely overlooked in approaches usually taken to questions about reference and meaning by philosophers: those approaches have tended to obscure the fact that judgments about referential continuity (and therefore about the truth of past existential commitments) always involve *interpretive* decisions concerning past speakers and linguistic communities. (Stanford [2015b], p. 406)

He takes various correspondence theories to task for supposing

that the empirical facts . . . at the time that Priestley used the term 'phlogiston' are sufficient to determine the facts about which objects or properties in the world were those to which [his term] did and did not refer. But this way of seeing the situation simply ignores the fact that we ourselves are making decisions about how to *interpret* . . . Priestley. (Stanford [2015b], p. 407)

The moral, for Stanford, is that our current confidence that 'Atoms exist' reflects a host of factors irrelevant to the underlying physical facts of the case,

50. This might be called 'the indeterminacy of translation' if that term weren't already taken for a more ambitious claim. See Essay #6 for more.

and for that reason, it doesn't conflict with his assessment that, for example, Dalton's atomic theory was radically false.

But surely this moral extends beyond existential claims. Was Maxwell right to think 'the ether must be substantial'? If we're out to document the historical fall of mechanism, then we'd say 'no, of course not'; if we're out to emphasize the fact that the electromagnetic field has mass-energy content, we might say 'yes, though of course Maxwell was wrong to think that being substantial should be understood mechanistically'.⁵¹ So, was Perrin* right to think that 'matter is particulate'? From the point of view of fairly recent quantum field theory, you might say 'it turned out there aren't any particles, just fields', or you might say 'it turned out that particles are small concentrations in the field'.⁵² These vagaries would explain why the debate over the 'approximate truth' of past theories seems so stubbornly unresolvable: there's no fact about the theory's subject matter alone that decides the point; our answer also depends on our context of interpretation. The historical case for instrumentalism is mounted by pointing to past theories we're all supposed to agree were radically false, but those judgments depend on how we interpret the theories, and these interpretations are influenced by a range of factors unrelated to the worldly facts of the case.

If this is right, the historicist critique of scientific realism is in serious trouble: it draws consequences for our current theories based on retrospective assessments of past theories, but those retrospective assessments have been shown to be sensitive to irrelevant contextual factors. More locally, challenging selective realists to identify those portions of their theory that will survive in future theories is similarly misguided. For example, it would be pointless to ask Perrin* to predict that future theories will agree that matter is particulate.

51. Stanford considers this example in his [2006], p. 161, footnote 13. There he rejects the 'yes' answer on the grounds that Maxwell's 'substantial' must be understood as mechanistic. It's not clear whether this opinion would shift in light of Stanford [2015b].

52. See Steven Weinberg [1987], pp. 78–79 (quoted by Malament [1996], p. 1): 'A quantum field theory is a theory in which the fundamental ingredients are fields rather than particles; the particles are little bundles of energy in the field.' (This comes close to saying both things in the course of one sentence.) The precise manner of recovering particulate structure in contemporary field theories is a subtle matter, a subject of continued discussion and debate.

The turning point in this line of thought is the recognition of the role of translation/interpretation in a disquotational theory of reference. If some form of correspondence theory could be defended instead—a theory according to which the referent of Pierre’s ‘neige’ and the truth of his ‘la neige est blanche’ are determined by facts exclusively about Pierre, his language, and his physical environment—then there would be something for our retrospective assessments to get right or wrong, but without that, it’s not clear that Stanford can afford to help himself to the idea that our assessments of past theories involve translation/interpretation in anything like the way Stein, Wilson, and I propose.⁵³ But Perrin* can so help himself, continuing to hold the theories of the past in high regard⁵⁴ and stating his commitments in terms of current confirmation, not prediction.

Perhaps surprisingly, I think this conclusion is less troubling for Stanford’s core position than it appears, because I doubt that historical critique is essential to his epistemic instrumentalism. There was a hint of this in the observation that Perrin, Einstein, and no doubt others were initially wary of atoms despite the overwhelming abductive support for atomic theory. Stanford has argued that they were right to be wary at that point, because the historical record shows that human scientists are bad at cataloguing the entire range of candidate explanatory theories. But it’s unlikely that any of these practicing scientists were moved by anything like Stanford’s painstaking historicist critique. It was enough for them to observe that any inference to the best explanation is really an inference to the best explanation we’ve thought of so far, and that as long as the only evidence for atomic theory was exclusively its explanatory and predictive power, there was always the open possibility that something else was responsible for the phenomena in question, that some other relationship between theory and world—more devious than truth—was responsible for its success. In this way, the initial ground-level

53. Stanford [2000] has championed a causal theory in the past, so this might be a natural resting place for him.

54. Of course, Stanford could forgo ‘truth’ and ‘reference’ altogether and simply challenge the realist to explain what reason she has to think her current theory is better than the superseded theories of the past. But the very considerations that tempt her to interpret those theories generously must also remove the sting from the suggestion that her current theories are ‘no better’.

challenge to Perrin* remains intact: does your new evidence extend beyond the abductive in a way comparable to fossil evidence for dinosaurs?

As it happens, the historicist contrast between Catastrophism and Uniformitarianism recedes in Stanford's more recent work, as we'll now see.

(iv) *The middle path*

So far, we've reviewed Stanford's distinctive historicist argument,⁵⁵ then traced the quest for a 'difference that makes a difference' through the advent of the Catastrophism/Uniformitarianism contrast, and finally to the charge that the realist must be theoretically conservative in objectionable ways. Now, in his latest paper, 'Realism, instrumentalism, particularism: a middle path forward in the scientific realism debate' (Stanford [2021]), we find Stanford identifying a new 'middle path' that's recently arisen between 'historically sophisticated realists and instrumentalists' (Stanford [2021], p. 1). This middle path has three tenets.

Perhaps surprisingly, the first of the three shared tenets of this middle path is Uniformitarianism, which has come to be embraced by some selective realists:

So-called 'selective' scientific realists have argued, for example, that although we should indeed anticipate further radical and fundamental changes in our theoretical conception of the natural world, we can nonetheless identify particular elements, aspects, or features of our best scientific theories that we can justifiably expect to find preserved throughout the course of such future changes. (Stanford [2021], p. 3)⁵⁶

As we've seen, Perrin* has been a Uniformitarian all along, so he passes this test for the middle path.

The second shared tenet is a denial that those 'radical and fundamental changes of our theoretical conception' involve any kind of Kuhnian

55. Though I've just argued (in (iii), above) that the historical element here isn't strictly necessary—and that that's a good thing.

56. In light of (iii), above, note that the selective realist could instead claim to identify those parts of current theory that they take to be confirmed, leaving out the problematic predictive element.

incommensurability. The realist easily agrees to this, so the concession here, if any, comes from the instrumentalist side. In fact, incommensurability has never been part of Stanford's epistemic instrumentalism (see Stanford [2006], p. 22).

Third and finally, those on the middle path share this simple conviction:

[W]hen scientific theories are able to achieve robust empirical and practical success, it is surely at least reasonable to think that the reasons for that success will consist in *some* systematic relationship or connection between how the theory represents (some part of) the world as being and how things actually stand there. (Stanford [2021], p. 7)

Stanford generously calls this the Maddy/Wilson principle, but it's not entirely clear that he understands it in the same way that Wilson or I would. The only guidance he gives is to note that the historical record is rich with cases of subsequent theories explaining the success of their predecessors. Perhaps he intends a methodological injunction to seek such explanations.

If I may speak for Wilson, I believe we would both happily agree to this, but expect something more. Two of the overriding morals of *Wandering Significance* (Wilson [2006]) are that we shouldn't be too fastidious or dismissive about theories whose success we can't currently explain (e.g., quantum mechanics) and that we should seek such an explanation nonetheless and reasonably hope for one in the fullness of time (as we do for quantum mechanics). This describes Perrin*'s initial attitude: he has an extremely effective theory; he certainly doesn't intend to give it up (neither does Ostwald, for that matter), but he wants to understand why it works. One possible answer—the first answer to try—is that matter is in fact particulate, so Perrin* sets out to test this explanation as best he can. If he hadn't succeeded, and if other attempts had also failed, presumably he would have looked for alternative explanations.⁵⁷ This is what scientists routinely do: it's why Michelson and Morley attempted to measure the ether wind; it's why Thorne, Weiss, and Barish sought out gravitational waves; it's why more than half a dozen underground laboratories are now looking to detect dark matter.

57. Recall footnote 37.

When Stanford agrees that ‘there must *be* some reason why a cognitive instrument that works well does so’ (Stanford [2021], p. 7), he doesn’t explicitly say that this third shared tenet includes a methodological injunction to ask what that ‘systematic relationship or connection’ is, to actively seek an explanation of why a given theory works so well. Still, since he’s engaged in a search for ‘a difference that makes a difference’, it seems safe to assume that he would have noted a disagreement on this point if there were one.⁵⁸ In (i), above, I suggested that their respective attitudes toward this injunction might mark a methodological difference between Perrin* and Stanford’s epistemic instrumentalist, but this potential way of distinguishing them seems to have vanished on the middle path.⁵⁹

So, what difference does Stanford find? Though the middle-path realist is no longer a Catastrophist, Stanford repeats the charge that she will be objectionably conservative about the portions of her theory that she regards as confirmed, this time on the grounds that the middle-path epistemic instrumentalist

believes that even more instrumentally powerful alternatives radically distinct from contemporary scientific theories are actually out there still waiting to be discovered. (Stanford [2021], pp. 11–12)

But as a Uniformitarian, doesn’t the middle-path realist believe this, too? Assuming she shares Perrin*’s steadfast fallibilism, why shouldn’t she, too, see the wisdom in pursuing or funding ‘transformative’ scientific possibilities? For that matter, Stanford once told us that the epistemic instrumentalist takes our current theories ‘as the appropriate starting point in trying to determine how they themselves can be refined, improved, and developed’, that ‘she will not . . . pursue the further implications of our theories less doggedly, or invest those implications with less significance, than the realist’ (Stanford [2006], pp. 209–210). So it’s not clear that there’s a significant difference in levels of conservativeness here, without which Stanford hasn’t yet

58. In comments on an earlier draft of this essay, Stanford confirmed that he agrees with Wilson and me on this point.

59. Whether Stanford’s instrumentalist has reason to seek to unify his various theories, as Perrin* does, is a question that remains open.

given convincing evidence that his instrumentalist's 'but I don't believe it' has methodological teeth (though Perrin*'s 'but I may be wrong' apparently does—helping to guard him from a complacent conservatism).

However this question is resolved, 'Middle path' also contains a new proposal for a methodological separation between the middle-path realist and the middle-path instrumentalist, but first, there's a tantalizing return to what I've been calling the ground-level dispute. Recall how, in *Exceeding Our Grasp*, Stanford emphasized that his epistemic instrumentalist wasn't issuing a blanket injunction against theoretical claims about inaccessible domains; his objection was to such claims that are justified by abduction only, claims for which there's no evidence comparable to fossil evidence for dinosaurs. Similarly, the realist needn't commit herself to every posit or claim in her current best theory, so the difference between them was just over where to draw the line. This point returns in 'Middle path':

Realists and instrumentalists alike can recognize exceptions to their general or generic expectations in particular cases based on evidence or other considerations specific to the case in question. (Stanford [2021], p. 15)

Elsewhere, Stanford explains in some detail how 'evidence . . . specific to the case' works to convince the instrumentalist of the epistemic force of fossils.⁶⁰

But this is precisely the sort of thing Perrin* sees himself as doing: he presents what he takes to be 'considerations specific to the case' that go beyond the previous, exclusively abductive evidence; like Stanford's instrumentalist, he takes this new and distinctive evidence to support rational belief in the particulate character of matter and its attendant randomness. Here the original, ground-level debate rises back into view, and furthermore, surprisingly, it now appears that Perrin* qualifies as a middle-path instrumentalist!

But if Perrin* is the instrumentalist, where is the realist in this story? We're told that, roughly speaking,

realists are generously presuming our successful theories (or privileged parts thereof) innocent unless and until proven guilty, while

60. See Stanford [2010].

instrumentalists are cynically presuming guilt unless innocence can be convincingly established. (Stanford [2021], p. 17)

So the question is how theoretical claims are ‘proven guilty’ for the realist, presumably something that hinges on ‘evaluation of the details of the *specific* evidence . . . in support of that particular belief’ (ibid., p. 15). If they’re guilty by virtue of being supported by exclusively abductive evidence, then realism collapses into instrumentalism, so it must take more than that. For me, at least, this middle-path realist remains a cloudy figure—perhaps she is conservative, as Stanford claims—but it’s hard not to suspect ‘guilty until proven innocent’ is simply a better match for actual scientific method than ‘innocent until proved guilty’.

However that may be, the middle path realist’s attitude is apparently not Perrin*’s. So despite my initial assumption that Perrin* must end up as a ‘realist’ of some variety, let’s leave the middle-path realist aside for the moment and continue from here on the assumption that Perrin* is a middle-path instrumentalist engaged in the ground-level argument that his evidence is as rationally compelling as fossil evidence. The question is just: does this particular evidence justify this qualitative change in our level of confidence?

Given what we’ve seen, perhaps it’s no surprise that this isn’t the discussion that interests Stanford. His focus is neither on the middle-path instrumentalist’s struggle with her conscience over whether some new, non-abductive evidence is strong enough to satisfy her scruples, nor on her ground-level disagreement with an imagined middle-path realist over the adequacy of the evidence in some specific case. Rather, for Stanford, the relevant opponent of the middle-path instrumentalist is the middle-path *selective* realist who offers not just confidence in a particular case but a *general* criterion for identifying which parts of any given scientific theory have been confirmed and which haven’t.⁶¹ Though Perrin* expresses confidence in the particulate structure of matter and its attendant randomness, he doesn’t offer a general criterion of confirmation, and I suggested earlier that he would reject this demand.

61. Assuming the criterion is taken to be normative for this version of realism, the charge of conservatism becomes somewhat more plausible.

This isn't to deny that interesting or illuminating generalizations about the kind of evidence that confirms in this area or that might be possible. For example, some have been tempted to seek out common features in Perrin*'s 'detection' of molecular motion and more recent cases like the 'detection' of gravity waves, and to put forward a specification of what counts as 'detecting' something. My second-philosophical thought on Perrin*'s behalf is that any such specification would most likely be descriptive, not normative—that is, if a new case came along with compelling evidence that looked as if it, too, should be classified as 'detection', and if this new case didn't fit the existing specification, I suspect that the specification would be altered, not the new evidence rejected. This is more or less what happened in Perrin's case: at the time, detection was observation, and the Brownian motion experiments prompted the move to a broader notion. When someone draws an inference in the form of *modus ponens*, it seems right to say that the reason the inference is good is because it has that form; *modus ponens* is the norm. But is it plausible to think that the epistemic force of Perrin*'s evidence ultimately depends on features it shares with other evidential situations?⁶²

Stanford sees this skepticism about the normative value of general criteria as 'a counsel of despair': we're being asked to content ourselves with the ordinary scientific evidence, specific to the individual case, but this is

presumably what scientists themselves have been doing all along, and many of their resulting sincere and carefully considered judgments . . . have repeatedly turned out to be spectacularly mistaken. (Stanford [2021], pp. 21–22)

Here we're returned to the familiar terrain of 'spectacularly mistaken' versus 'approximately true',⁶³ which once gave way to Uniformitarianism versus Catastrophism, and then to all-Uniformitarian middle path—which led to the conservativeness charge against the generalizing selective realist.

62. See [2007], pp. 402–403. Stanford [2021], p. 20, cites this passage in his characterization of 'particularism', below.

63. Both problematic retrospective assessments in light of (iii).

But what about the Uniformitarian, fallibilist Perrin*, who claims to have specific evidence that goes beyond the eliminative or abductive? His closest approximation in Stanford's taxonomy is the radical particularist:⁶⁴

The particularist thinks the very best we can do in deciding whether some particular claim or commitment is true . . . [⁶⁵] is to carefully evaluate the details of the specific evidence we have for and against that particular claim or commitment. That is, she thinks that the delicate and painstaking *scientific* work of evaluating particular claims and commitments *already* represents our most sophisticated efforts to determine the appropriate level of confidence we should have in particular claims about what things exist and how they interact. (Stanford [2021], p. 18)

This figure, it turns out, is rejected by both middle-path realists and middle-path instrumentalists:

middle path realists and instrumentalists are . . . united . . . in contrast . . . to the . . . expectations of those who defend radical forms of what we might call 'particularism'. (Stanford [2021], p. 17, see also p. 22)

Strikingly, this united front views the particularist, not as someone to be confronted, but as someone irrelevant to the discussion, because

Since its inception . . . the modern realism debate has been predicated on the assumption that there is some *point* to ascending to levels of abstraction at which we generalize about 'mature scientific theories' and their 'empirical successes' or 'approximate truth'. (Ibid., pp. 17–18)

64. I say 'approximation' because Stanford's radical particularist also holds that 'there is not now nor was there ever any point' in generalizing about, e.g., confirmation, or in examining the record of scientific success in the historical record (Stanford [2021], pp. 18–19). The former might be useful in some ways without being normative, and the latter is presumably part of what colors Perrin*'s Uniformitarianism and his fallibilistic sense of the riskiness of scientific theorizing. Stanford also describes a 'modest particularist', but I use the term without adjective to mean the radical version.

65. The excised phrase, 'and/or will be retained and ratified throughout the course of further inquiry', includes the predictive element we're avoiding whenever possible.

The philosophical debate over realism is concerned with ‘broad reflections on the scientific enterprise as a whole or patterns in the historical record’ (ibid., p. 19), not with the details or status of Perrin*’s evidential innovations. In other words, the ground-level debate of *Exceeding Our Grasp*, as the particularist understands it, is just beside the point.

I think it’s no longer possible to avoid the conclusion that there are, on Stanford’s account, two different realism/instrumentalism debates. The first is typified by a ground-level dispute—between the middle-path instrumentalist and a middle-path realist, or between the middle-path instrumentalist and her own conscience—for example, over the efficacy of Perrin*’s evidence in the case of atomic theory: should the line between the real and the merely instrumental be moved? This is a particularist debate, to be settled ‘on evidence or other considerations specific to the case in question’ (Stanford [2021], p. 15). The second is a more general or ‘abstract’ exchange between the middle-path instrumentalist and the middle-path selective realist over whether there’s a general criterion for picking out the confirmed parts in any scientific theory, or perhaps over the rationality of some brand of theoretical conservativeness. The important point for our purposes is that the latter—the general, ‘abstract’ debate—is what figures in ‘the modern realism debate’, and the former—the particularist debate—does not. (Stanford’s initial identification of the instrumentalist attitude with that of the realist toward, say, quantum mechanics, would seem to involve a conflation of these two distinct disputes.)

So, in the end, we see that Stanford, like van Fraassen, simply doesn’t engage Perrin* on the subject of atomic theory—or by extension, the Second Philosopher, either, with her cloud chambers and electron microscopes. For van Fraassen, Perrin* is a benighted Native, who persists in thinking that ordinary scientific evidence is relevant to the debate; for Stanford, he’s a middle-path instrumentalist struggling with a decision over where to draw the line between what he takes instrumentally and what he takes realistically, not an entrant in the more general, properly philosophical debate. Stanford is no transcendentalist—as an integrative naturalist, his conclusions are based on the historical record, not philosophical empiricism—but he, too, insists that Perrin* isn’t functioning at the appropriate level. For van Fraassen, the appropriate level is the seminar room, separate from ordinary scientific inquiry;

for Stanford, it's this higher level of generality or abstraction, still within a broadly empirical inquiry.⁶⁶ But either way, they both insist that believing in atoms isn't enough to qualify one as a 'realist' in the relevant sense. So, despite appearances, at least according to these two influential practitioners, the Second Philosopher never entered the realism/instrumentalism debate in the first place!

But labels and partisanship aside, I think this discussion has helped illuminate the nature both of Stanford's contribution and of our disagreement. By calling attention to his problem of unconceived alternatives in theoretical science, Stanford has shifted attention from the pessimistic induction and the status of unobservables to the shortcomings of purely abductive evidence. This is the same concern I raised for the special case of atomic theory in my own response to indispensability arguments and Quinean holism in the philosophy of mathematics, and I think the examples of Perrin, Einstein, and many others demonstrate that it's a concern shared by scientists themselves: it's not enough for atomic theory to explain and predict; we demand some kind of direct contact with its subject matter.⁶⁷ The trick for us as methodologists is to examine what this 'some kind of direct contact' comes to and why it's a rational basis for belief. Stanford ([2010]) has done something analogous for the case of fossil evidence (we have access to the process of fossilization); I've made some halting steps elsewhere⁶⁸ on doing so for the case of Perrin's evidence (we detect the random walk).

So again, labels and partisanship aside, where do we two integrative naturalists ultimately differ? We agree that abductive evidence isn't enough by itself to rationally compel belief; we agree that determining where the line falls between what is and isn't fully confirmed (and thus rationally believed) is to be undertaken on a case-by-case basis; we agree that the relevant evidence in each particular case is specific to that case; and we agree it's unlikely that there's a general, normative criterion for what's confirmed in this way. Where we disagree is on that old ground-level problem—is Perrin*'s evidence good enough?—but Stanford takes the truly 'philosophical' disagreement to be

66. Van Fraassen's seminar-room Empiricism is a stance, presumably freely chosen. Stanford seems to base his preference for the higher level of abstraction on tradition.

67. Again, as in (iii), above, historicism needn't come into it.

68. See [1997], pp. 135–143, [2007], pp. 404–407.

the one between the two of us and those who think there is a general, normative criterion. So in the end—setting aside local disputes over the efficacy of specific evidence in particular cases—it seems we disagree only on where philosophers’ efforts are most fruitfully deployed. Stanford remains focused on opposing those offering candidate criteria; I believe the real task is the methodological one of examining what has or might work in those local disputes and why. Could it be that, once the rhetoric is cast aside, the real difference between us is no more than one of emphasis?⁶⁹

69. Warm thanks to my reading group—Adam Chin, Alysha Kassam, Charles Leitz, Stella Moon, Evan Sommers, and especially Christopher Mitsch and Jeffrey Schatz—and to Anjan Chakravartty for their most helpful reactions to an early draft. I’m also indebted to David Malament and Mitsch, again, for help with the physics. Finally, Stanford was good enough to read the penultimate draft and correct some of my misunderstandings of his position; I’m grateful to him and regret those misunderstandings that inevitably remain.

PART II

SKEPTICISM

Hume and Reid

My topic here is meta-philosophy, the question of how philosophy is properly done. For some years now, I've been developing a particularly austere, roughly naturalistic¹ approach to philosophical questions that I call Second Philosophy. It's seemed to me that one effective way to convey the spirit of Second Philosophy is to compare and contrast it with other, more familiar methods, like transcendental or therapeutic philosophy.² Here I hope to pursue this sort of engagement with two other venerable schools of thought: Hume's *Science of Man* and Reid's *Philosophy of Common Sense*.

Hume presents a fitting starting point for any discussion of naturalism—especially when Reid is on the agenda—so my first pass at a portrait of the Second Philosopher traces her relations to the Scientist of Man. Of course, Hume's cheerfully industrious inquirer³ eventually lands on the barren rock of skepticism, so it's also worth taking a second-philosophical look at the kinds of considerations that led poor Hume to his shipwreck. This sets the stage for Reid.

1. HUME

The Second Philosopher is an entirely workaday inquirer, out to describe and understand the world in which she lives. She begins with ordinary

1. I use 'naturalism' here as a general term implying no more than a vague methodological kinship between philosophy and natural science. My goal is to trace how this idea plays out in Hume, Reid, and my own Second Philosophy ([2007]).

2. See [2007], [2011a]. [[See also [2014b], [2017a].]]

3. E.g., Stroud [1977], p. 1, sees in Hume 'the unbounded optimism of the enlightenment'.

perception, gradually develops more elaborate methods of observation, experimentation, theory formation and testing; she corrects her prior beliefs and refines her methods as she goes. She's idealized to the extent that she's equally at home in everything from physics, chemistry, and astronomy to psycholinguistics, biology, and botany. Along the way, she comes to appreciate the usefulness of elementary logic and arithmetic, and eventually of more advanced mathematics for formulating and elaborating her account of worldly phenomena. Given that she's always keen to understand how her methods work when they do, and don't work when they don't, she comes to wonder, to take just one example, about logic: why is logical inference reliable?, what is the proper way to pursue the study of logic?, what is the nature of its subject matter?, how do we come to believe it?, is there only one correct logic?, and so on. In these ways, she comes to ask questions traditionally regarded as philosophical—though she approaches them with the same tried-and-true methods she's been honing from the start—and the result is an example of Second Philosophy.⁴

Now compare this with Hume's description of the Scientist of Man,⁵ who proceeds by 'experience and observation' (intro., 7).⁶ He never 'go[es] beyond experience, or establish[es] any principles which are not founded on that authority' (intro., 10). As Scientists of Man, Hume concludes,

We must . . . glean up our experiments in this science from a cautious observation of human life, and take them as they appear in the common course of the world, by men's behavior in company, in affairs, and in their pleasures. Where experiments of this kind are judiciously collected and compar'd, we may hope to establish on them a science, which will not be inferior in certainty, and will be much superior in utility to any other of human comprehension. (intro., 10)

4. See [2007], Part III. [[See also Essay #8, [2014b].]]

5. I allow myself a slight reconstruction of Hume's opening pages, to bring them into a closer parallel with my presentation of the Second Philosopher, but I don't think this substantially distorts Hume's intentions.

6. All references to Hume come from his [1739], unless otherwise indicated. I use the common book/part/section/paragraph citations, with 'intro.' and paragraph number for the Introduction.

Though the Scientist of Man is an 18th-century figure and the Second Philosopher resides squarely in the 21st, though the Scientist of Man focuses on the human portion of the Second Philosopher's more all-encompassing sphere of interest, there are obvious parallels. Notice in particular the approach Hume takes to describing the inquiry he has in mind. Though he speaks of 'sciences' and draws his method from Newton, he doesn't seem inclined to offer what contemporary philosophers of science would call a 'demarcation criterion,' that is, a characterization, once and for all, of the methods that are proper to the Scientist of Man. Instead, Hume describes his inquirer's starting point—experience, observation, experimentation—draws our attention by example to the sort of thing he has in mind—the practices of the descendants of Lord Bacon, leading up to Newton—and uses 'science' as a rough-and-ready term for that sort of thing. My description of the Second Philosopher follows the same pattern.

Once he's introduced his Scientist of Man, Hume presumably begins the *Treatise* by adopting the persona and conducting his philosophizing in that mode. This has also been my own practice with the Second Philosopher; I adopt her perspective and attempt to generate some Second Philosophy. This approach invites the question: what stance is required for describing the Scientist of Man or the Second Philosopher in the first place? Was Hume functioning as a Scientist of Man even in the Introduction to the *Treatise*, or was some other perspective needed to get the project off the ground? Examining the text, we find Hume beginning with a series of historical and sociological observations about the factors that have produced a widespread disdain for metaphysics; suggesting it would be 'vain and presumptuous' to suppose that the truth should be 'easy and obvious' (intro., 3); proposing the loosely characterized empirical method as more promising than the ineffective methods of the past.⁷ None of this goes beyond the range of the Scientist of Man, as we've described him so far. My own presentation of the Second Philosopher is if anything more circumspect: I simply describe her behavior in rough-and-ready terms, without appeal to a demarcation criterion; I don't

7. He also spends some time eschewing hypotheses in good Newtonian fashion (intro., 8), which was certainly good scientific methodology at the time. In [2008] and [2011b], chapter 1, I argue that times have changed.

advocate for her approach, leaving that evaluation to the reader's own judgment; and I don't spend time 'decrying all those [systems of philosophy] which have been advanced before [this one]' (intro., 1).⁸

Given these broad similarities between the Scientist of Man and the Second Philosopher, between Hume's approach and one fairly strict version of contemporary naturalism, I think we can't help but feel a jolt of surprise at what turns up on the very next page, the first page of the body of the *Treatise*: instead of describing 'men's behavior in company, in affairs, and in their pleasures' (intro., 10), Hume begins with a detailed examination of impressions and ideas, apparently by introspection into his own mental experience. Obviously this is at odds with the sort of approach a Second Philosopher might take to a study of human perception: she'd look into everything from the composition of ordinary objects and their tendencies to absorb and reflect light to the structure of the eye, the neural pathways to the brain, and the processing that takes place there. Her assessment of the reliability of our perceptual beliefs would involve an analysis of the conditions under which our perceptual mechanisms tend to function poorly and well, conditions ranging from the state of the perceiving subject to the ambient lighting, the distances in question, and the composition of the scene itself. Somewhere in all this there would be a study of the phenomenology of perceptual experience, including no doubt introspective evidence, but this would not be the leading subject of the inquiry.

What accounts for this stark divergence? Part of the answer may lie in Hume's allegiance to the Theory of Ideas, but if so, this is only implicit at the outset of the *Treatise*. That theory becomes explicit only much later, at the end of Part 2, where Hume announces that

'Tis universally allowed by philosophers, and is besides pretty obvious in itself, that nothing is ever really present with the mind but its perceptions

8. Stroud has a somewhat different concern in his [2009a]: not that I, in describing Second Philosophy, use methods unavailable to the Second Philosopher; rather that I respond to and enter into discussions that a true Second Philosopher wouldn't be motivated to address. For some thoughts on this question, see [2011a], footnote 56. [[For a more recent and extended treatment of the Second Philosopher's array of methods and inquiries, see [2017a], where she's called the Plain Inquirer.]]

or impressions and ideas, and that external objects become known to us only by those perceptions they occasion. (1.2.6.7)

This passage refers us forward to the argumentation in Part 4, which we'll take up in a moment. Stroud describes Hume's attitude this way:

This is a precursor of what has come to be called the 'sense-datum' theory of perceiving, and it has been held in one form or another by most philosophers since Descartes . . . The legacy of Descartes, Locke and others made that part of the theory of ideas completely uncontroversial to Hume [and] Hume is not alone in this. There is very little argument for the basic principles of the theory of ideas in Locke, either, but Berkeley's *Dialogues* provide an almost complete catalogue of the familiar considerations in support of some such view. (Stroud [1977], p. 26)

Hume's largely unexamined belief in the Theory of Ideas must help explain why he is happy to start the *Treatise* on the topic of impressions and ideas.

But what about Hume's inquirer, the Scientist of Man—are we to understand him as implicitly presupposing the Theory of Ideas from the outset, from the very beginning of Part 1?⁹ If we understand the *Treatise* as 'an inquiry . . . a structured sequence of discoveries and reflections that Hume is narrating for his reader' (Broughton [1992], p. 166), indeed as a narrative that serves to track the progress¹⁰ of the Scientist of Man, then there's every reason to suppose that the full force of the Theory of Ideas isn't felt until the actual argumentation of Part 4: it isn't presupposed, it's discovered in the course of the Scientist's investigations.¹¹ On the basis of meticulous case studies, Broughton argues that

9. If so, Hume's Scientist of Man may suffer from a disorder similar to the one Broughton ([2002], pp. 28–32) diagnoses in Descartes's meditator: both begin their inquiries in ways that are motivated by conclusions they will reach only much later.

10. Broughton [2005], p. 175, distinguishes a progress, which successfully 'moves us from a less-good position to a better one', from an inquiry, which may not, and argues that the Scientist of Man traces an inquiry but not a progress. I don't intend 'progress' here to involve any more than her 'inquiry'.

11. Cf. Stroud [1977], p. 27: 'the fact that Hume . . . thinks [the theory of ideas] is 'pretty obvious in itself', should not suggest that he thinks he knows it in some way other than by observation or experience. For him, it is known by the same kind of 'cautious observation of human life' that informs him of most of the rest of his philosophical system'.

Hume begins the *Treatise* by studying our general pre-reflective beliefs, ideas and outlooks, *and by using them*, or at least using those among them that are, we agree, the most careful, attentively considered, and broadly speaking, reasonable. His method enjoins him to be an observer of human life, and we observe people's lives by listening to what they say and seeing what they do, both in their (our) relations to one another and in their (our) relations to various things and events . . . the scientist of man [is] perfectly entitled to observe people seeing, hearing (etc.) things. (Broughton [1992], pp. 160, 166)

Here we see the Scientist of Man apparently starting out from the same point as the Second Philosopher, and only subsequently being led to the Theory of Ideas.

Let's suppose then that Hume begins where he does because he knows where he's going to end up, but that the impressions and ideas described by the Scientist of Man are taken, at the outset, to be ordinary sensory perceptions of the world.¹² What this doesn't explain is what motivates the Scientist of Man to focus so resolutely on his own experience instead of embracing a broader investigation of how perception functions in the world, a study closer to the Second Philosopher's. In the paper just cited, Broughton floats some possible explanations:¹³ that topics like the physical composition of the objects we perceive, optics, and the physiology of sensation fall to the Natural Philosophers, not the Philosophers of Man, or that 'in general, early modern associationist psychologists neither found much explanatory use for physiology nor succeeded in using psychology to confirm the (largely speculative) physiology they did invoke' (Broughton [1992], p. 168). The first of

12. Siding, more or less, with the vulgar: 'In general all the unthinking and unphilosophical part of mankind, (that is, all of us, at one time or other) . . . suppose their perceptions to be their only objects, and never think of double existence internal and external, representing and represented' (1.4.2.36); 'Tis certain, that almost all mankind, and even philosophers themselves, for the greatest part of their lives . . . suppose, that the very being, which is intimately present to the mind, is the real body or material substance' (1.4.2.38).

13. Broughton also suggests that Hume may have ignored physiology because he took for granted 'All that experience can tell us is what sensations we have, what the physiology of sensation is, and what our states of consciousness are. Experience cannot tell us the further why or how of it' (Broughton [1992], p. 165)—an appeal to the Newtonian principle of eschewing hypotheses (see footnote 7). This is less directly relevant to the concern raised in the text.

these doesn't strictly help with our concern; if physics, optics, physiology fall outside the range of the Scientist of Man, our question becomes: how can the Scientist of Man expect to assess the reliability of perception without cooperation from the Natural Philosophers? How can you tell when perception is and isn't properly reporting the properties of what's perceived without consulting your best account of the actual properties of what's perceived? The second may help explain why Hume didn't appeal to physiology, but that leaves the question of physics, optics, and the rest.

It seems to me there's a more principled reason why Hume narrows the Scientist of Man's inquiry as he does, one that traces back to the opening strains of the Introduction. Before he even reaches his proposal that the 'experimental method of reasoning' be applied to 'moral subjects' (subtitle), he declares:

'Tis evident, that all the sciences have a relation, greater or less, to human nature; and that however wide any of them may seem to run from it, they still return back by one passage or another. Even *Mathematics*, *Natural Philosophy*, and *Natural Religion*, are in some measure dependent on the science of MAN . . . There is no question of importance, whose decision is not compriz'd in the science of man; and there is none, which can be decided with any certainty, before we become acquainted with that science. In pretending therefore to explain the principles of human nature, we in effect propose a compleat system of the sciences, built on a foundation almost entirely new, and the only one upon which they can stand with any security . . . the science of man is the only solid foundation for the other sciences.¹⁴ (intro. 4, 6, 7)

14. Unfortunately, I can't pretend to understand why Hume insists on this. His case for it appears to lie in the continuation of a sentence quoted above: 'Even *Mathematics*, *Natural Philosophy*, and *Natural Religion*, are in some measure dependent on the science of MAN; since they lie under the cognizance of men, and are judg'd of their powers and faculties' (intro., 4). Is the argument that we learn about the world by using our cognitive faculties, so our theory of the world is founded on our theory of our cognitive faculties? No doubt Hume has something better than this to offer, but I don't know what it is.

Only after all this does he continue, '[T]he only solid foundation we can give to [the Science of Man] itself must be laid on experiment and observation' (ibid., 7).

Now if the Science of Man is to provide an epistemic foundation for physics, optics, physiology, and so on, then there's no mystery as to why the Scientist of Man doesn't appeal to these sciences in the course of his analysis of perception: to do so would be to use these sciences in the course of developing their own foundation! Broughton makes this observation in her more recent 'Hume's naturalism about cognitive norms':

Notice that Hume does not describe any reciprocal relation between the science of man and other sciences: he holds out no hope that natural philosophy, for example, might lead to changes and improvements in the science of man. His naturalism, then, does not involve seeking concepts or results from well-established, empirically based disciplines. (Broughton [2003], p. 7)

Let me mention in passing a possible second-philosophical concern that Hume not only cuts off the Scientist of Man from the concepts and results of the other sciences, but that he can't even properly pose the question of the reliability of perception without appeal to his best theories of the things purportedly perceived and the physical world they (and we) inhabit.

For that matter, I think this foundationalism—if I may use that term for the position just described—is problematic even from Hume's own point of view. If the Scientist of Man begins from ordinary observations of human beings and their interactions with the people and objects around them, if careful observations of this sort are to be regarded as unproblematic at the outset, then what possible objection could there be to careful observations of falling bodies, the properties of pea plants, or human anatomy? And if the Scientist of Man is warranted from the outset in applying the methods of Bacon and Newton to his data, why shouldn't he be warranted in applying those same methods to data of these other sorts? And if all this is correct, how can physics or botany or genetics or physiology be any less well-founded than the Science of Man? Once the Science of Man becomes empirical, it's hard to see why it isn't just one empirical science among many, all capable of

correcting and improving each other. If a conflict arises between the Science of Man and some other science, it flies in the face of proper scientific method to think one can know ahead of time which side of the equation will need adjustment.

So it's hard to see how the Scientist of Man, as described by Hume in the remainder of the Introduction, could arrive at the foundationalist position that Hume himself embraces there. Returning to the question raised a moment ago, perhaps Hume doesn't behave entirely as a Scientist of Man ought in setting out the proper role and function of the Science of Man; contrary to our earlier tentative assessment, perhaps Hume in fact occupies some other perspective as he's describing the character and methods of the Scientist of Man, a perspective from which this strong foundationalism is defensible for some reason or other.¹⁵ This would raise a consistency problem for Hume himself—advocating one method while employing another—but it might manage to sustain the Scientist of Man as otherwise advertised, as the fully empirical inquirer we've been comparing with the Second Philosopher. The trouble with this line of thought is that it doesn't answer the question we're asking: why does the Scientist of Man begin with a narrow focus on his own perceptions? When he gets down to business, Hume's Scientist of Man is apparently hobbled by the very foundationalism Hume enunciates in the Introduction. So I think we must conclude that Hume *is* behaving as a Scientist of Man as he understands that figure, and that, for all his fine qualities, the Scientist of Man is less purely empirical in his practice than advertised, less fully naturalistic an inquirer than the Second Philosopher.¹⁶

In any case, though Hume begins his inquiry with a flourish of 'experience and observation . . . careful and exact experiments . . . explaining all effects from the simplest and fewest causes . . . [never] going beyond experience' (intro. 7, 8), we all know that he ends up 'on the barren rock' of skepticism (1.4.7.1):

15. The obvious candidate is a prior commitment to traditional empiricism, but I leave this aside here because I'm out to assess the prospects for a more naturalistic reading of Hume.

16. For that matter, the same goes for Quine, despite his having coined the phrase 'epistemology naturalized'. See [2007], I.6.

I begun this subject with premising, that we ought to have an implicit faith in our senses, and that this wou'd be the conclusion, I shou'd draw from the whole of my reasoning. But to be ingenuous, I feel myself *at present* of a quite contrary sentiment, and am more inclin'd to repose no faith at all in my senses . . . than to place in [them] such an implicit confidence. (1.4.2.56)

I am ready to reject all belief and reasoning, and can look upon no opinion even as more probable or likely than another. (1.4.7.8)

Despite their disagreements, the Second Philosopher should be troubled by this outcome: if the application of ordinary empirical methods, albeit without the aid of the special sciences, in fact leads to utter skepticism about the external world, she needs to figure out what's gone wrong.¹⁷ Notice that her reaction here differs from Hume's: he simply despairs of knowing the external world; her best guess is that we in fact have quite a number of reliable beliefs about the world, including many perceptual beliefs, so something must be wrong with the methods that brought Hume to his rock. If these methods are her methods, she figures they stand in urgent need of revision.

So how does Hume's line of thought lead him into skepticism? When he officially introduces and defends the Theory of Ideas, in 1.4.2.45, he alludes briefly to familiar considerations of perceptual relativity:

the seeming encrease and diminution of objects, according to their distance . . . the apparent alterations in their figure . . . the changes in their colour and other qualities from our sickness and distempers . . . (1.4.2.45)

and the venerable appeal to perceptual illusions:

When we press one eye with a finger, we immediately perceive all the objects to become double, and one half of them to be remov'd from their common and natural position. (1.4.2.45)

17. The intended contrast here is with ways of posing the skeptical challenge that explicitly adopt an extra-scientific viewpoint: we want an account of our knowledge of the world that doesn't depend on anything we now think we know about it (see, e.g., Stroud [1996]). I come back to this point in the concluding remarks below. [[For more, see [2017a].]]

He might also have mentioned hallucinations, with some standard case like the phantom pain in the severed limb. He concludes:

The only existences, of which we are certain, are perceptions, which [are] immediately present to us by consciousness. (1.4.2.47)

The argument has two threads: perceptual relativity shows that at least some of our percepts are ‘dependent on our organs, and the disposition of our nerves and animal spirits’ (1.4.2.45), which external objects are not;¹⁸ in the case of double vision, since ‘we do not attribute a continu’d existence to both these perceptions’ (1.4.2.45), at least one must not be an external object. Then, since all these percepts ‘have the same nature’ (1.4.2.45) and ‘from like effects we presume like causes’ (1.4.4.4), it follows that all of

our sensible perceptions are not possest of any distinct or independent existence. (1.4.2.45)¹⁹

So those items we perceive directly, of which we’re directly aware, are fleeting entities, internal to our individual minds, with no external existence. This is the Theory of Ideas.

18. Another argument from perceptual relativity begins in 1.4.4.3: ‘A man in a malady feels a disagreeable taste in meats, which before pleas’d him the most . . . That seems bitter to one, which is sweet to another . . . Colours reflected from the clouds change according to the distance of the clouds, and according to the angle they make with the eye and luminous body. Fire . . . communicates the sensation of pleasure at one distance, and that of pain at another’. Then, in 1.4.4.4, he concludes, ‘when different impressions of the same sense arise from any object, every one of these impressions has not a resembling quality existent in the object. For as the same object cannot, at the same time, be endow’d with different qualities of the same sense, and as the same quality cannot resemble impressions entirely different; it evidently follows, that many of our impressions have no external model or archetype’. He then argues via essentially Berkeleyan considerations that primary qualities suffer from the same difficulty. I ignore this argument in the text because it seems to me to raise a number of irrelevant issues, e.g., the thorny matter of what counts as ‘resemblance’. If a percept encodes information about the environment in a way that’s accessible to the subject’s computational processes, is this enough to count as resemblance? If not, what more is required, and why?

19. The shape of this argument is clearer in the 1.4.4.4 passage: ‘Many of the impressions of colour, sound, &c. are confest to be nothing but internal existences, and to arise from causes, which no way resemble them. These impressions are in appearance nothing different from the other impressions of colour, sound, & c. We conclude, therefore, that they are, all of them, deriv’d from a like origin.’

So, how does our sensory experience manage to deliver information about the world? Here, Hume tells us, philosophers have devised the ‘opinion of the double existence of perceptions and objects,’ according to which

the former are suppos’d to be interrupted, and perishing, and different at every return [and] the latter to be uninterrupted, and to preserve a continu’d existence and identity. (1.4.2.46)

This is what we now call a ‘representative theory of perception’: our percepts or ideas or sense-data are known immediately; external objects are known mediately, by their means. Hume then offers this objection, which he takes to close the case:

The only conclusion we can draw from the existence of one thing to that of another, is by means of the relation of cause and effect, which shows, that there is a connexion betwixt them, and that the existence of one is dependent on that of the other. The idea of this relation is deriv’d from past experience, by which we find, that two beings are constantly conjoined together, and are always present at once to the mind. But as no beings are ever present to the mind but perceptions; it follows that we may observe a conjunction or relation of cause and effect betwixt perceptions, but can never observe it betwixt perceptions and objects. ’Tis impossible, therefore, that from the existence of any of the qualities of the former, we can ever form any conclusion concerning the existence of the latter, or ever satisfy our reason in this particular. (1.4.2.47)

Here we have the radical skeptical conclusion that we can know nothing of the external world.

For the troubled Second Philosopher, two questions arise: do the arguments from perceptual relativity, illusion, and hallucination actually establish the Theory of Ideas?, and does the Theory of Ideas truly imply the skeptical conclusion? The first question will be a major theme in our discussion of Reid, so let me focus for now on the second question. The notion that the Theory of Ideas is the culprit in Hume’s descent into skepticism

is commonplace, so I'd hardly be bringing this up if I didn't intend to raise doubts. Let me try to sneak up on these.

To begin with, notice that Hume makes his case by arguing that causation—an association between ideas—can't connect an external object with an idea. Here it may appear that his controversial theory of causation is doing some essential work in undermining a representative theory of perception, but I think this can't be right. Granted, Hume has previously (in Part 3) traced causation to constant conjunction, but in these terms, there remain real causal relations in the world—the impact of a moving billiard ball on a stationary one causes the second to move—and he goes to some lengths to provide us with reliable principles of causal inference (1.3.15), including the rule 'same effect . . . same cause' that figured in the argument for the Theory of Ideas.²⁰ We're assuming it established, for the sake of argument, that we're immediately aware only of ideas of billiard balls and their constant conjunction, but to assume we aren't thereby mediately aware of the billiard balls themselves and *their* constant conjunction would be to beg the question currently at issue.

So we need to ask: why couldn't the Scientist of Man, on the basis of his ideas, take note of the placement of external objects, observe people's introspective reports of their experiences and their perceptions, put these together with what he knows of eyes, nerves, and brains, and devise a causal explanation of why perceptual beliefs are generally reliable, and even of when they tend to go astray? Assuming, as we are, that the Theory of Ideas is true, why couldn't he discover in this way that objects under certain conditions of position and lighting cause changes in the retinas of people suitably equipped and attentive, that these changes cause further changes in the nerves and brains of these individuals, that these changes cause those individuals to become aware of certain subjective percepts, to do some conscious computational processing—call it inferring, if you like²¹—then to report certain perceptual beliefs, and finally, that this process is generally reliable? I'm not supposing that this is what we *do* discover when we investigate human perception—this

20. See Broughton [1983].

21. Imagine, for example, that we were consciously aware of the calculations performed by the visual system to extract distance information by stereopsis from the slightly different stimulations of our two retinas.

is an imaginary empirical theory—but the question before us now is whether a theory of double existence necessarily leads to skepticism; the suggestion is that if the Theory of Ideas and a representative theory of perception were correct, there'd be no apparent obstacle to our finding this out empirically and assessing the reliability of our perceptual beliefs in the same way.

Now we know why Hume would reject this line of thought: because the Science of Man has to be in place before we can develop our theories of eye, nerves, brain, and so on, so we can't appeal to these in the course of our investigation of the Science of Man. Fair enough. But even without going into these various details drawn from the other sciences, why couldn't the Scientist of Man simply observe the conditions under which people tend to make perceptual reports and straightforwardly evaluate their overall reliability? Why couldn't he, in essence, regard the entire process described a moment ago as a black box and assess its performance in a rough-and-ready way by observing its track record? Perhaps Hume thinks he has already cast doubt on this kind of reasoning in the previous section of the *Treatise* (1.4.1), relying on arguments even sympathetic commentators find especially unconvincing.²² But it seems to me that the problem is more fundamental than this, that it goes beyond disallowing appeal to the sciences or to probabilistic inferences and extends even to individual perceptual beliefs of the most straightforward kind.

To see this, suppose the Scientist of Man forms the simple perceptual belief that 'there's a rose in my garden' upon seeing one there under optimal conditions. Hume tells us that he intends his inquirer 'to have an implicit faith in [his] senses' (1.4.2.56), so the Scientist takes this particular belief to be well justified, based on that implicit faith. Now suppose he's confronted with Hume's argument for the Theory of Ideas, which, once again, we're taking to be sound; what should we expect his reaction to be? If he is, in fact, taking ordinary perception as a given, then it seems to me he would say, 'How surprising! I only directly perceive my subjective ideas; this is all that's accessible to my introspective gaze. So my perception of the rose in my garden must involve some added element, an indirect sort of perception that's somehow mediated by my idea of the rose. I wonder how that works?'

22. See, e.g., Fogelin [2009], chapter 3.

In other words, he comes to a representative theory of perception. Finally, suppose he's then confronted with Hume's claim that a representative theory leads to radical skepticism. I submit that we might expect him to weigh his well-justified perception against the conclusion that we know nothing of external objects and to critically assess the steps leading up to the apparent conflict, just as the Second Philosopher would recommend. But, as Hume's line of thought actually plays out, his simple perceptual beliefs have enough *psychological* force to keep him from joining Berkeley in some form of phenomenalism, but not enough *epistemic* force to prompt him to re-examine his argumentation and the assumptions on which it rests.²³ Instead he simply draws his skeptical conclusion.

This suggests a more nuanced portrait of Hume's inquirer. The Scientist of Man doesn't in fact begin with implicit faith in his senses, or with the Theory of Ideas, either. Rather, he begins with the stipulation that all his studies must be founded on the Science of Man, and that the Science of Man itself not only begins from introspection of his experiences, but begins from such introspection alone, unaided by any other knowledge of the world, particular or general. He takes for granted, at the beginning, that what he introspects are straightforward perceptions of that world, so that ordinary perceptual beliefs are available to him and the restriction isn't much felt. But when the Theory of Ideas is finally established in Part 4, the connection between introspection and simple perceptual belief is severed: perception tells us directly only about ideas, not about roses, so our beliefs about roses now stand in need of defense. Since simple perceptual beliefs have no independent standing, since there is in fact no 'implicit faith,' that defense must proceed purely from introspection, and thus we're effectively caught behind the Veil of Perception. If this is right, once again, it isn't the Theory of Ideas that leads Hume to his skepticism; it's his underlying foundationalism²⁴—and

23. As Broughton observes ([2004], p. 547): 'A different philosopher might have questioned the authority of [his cognitive] norms rather than accept such a negative outcome.'

24. Perhaps Hume's account of how our idea of external existence is derived from impressions also plays a role in motivating his skepticism. (See, e.g., Fogelin [1993], p. 93: 'A system of beliefs can be discredited by revealing its disreputable provenance . . . when *empirical investigation* lays bear the actual mechanisms that lead us to embrace [our belief in the external world], we are immediately struck by their inadequacy'.) Broughton [2003], p. 19, takes the opposite position: 'Note that Hume is not troubled by the confusions involved in the *production* of the vulgar idea of body; what troubles him is the

that foundationalism is far more traditional and pervasive than it first appears.²⁵

No doubt this reading of Hume is debatable, at least as much as they all are and probably considerably more so, but by now I think it should at least be clear that Hume's inquirer is a far cry from the Second Philosopher and that she needn't worry that her own methods will lead her inevitably down Hume's road to a skeptical conclusion. Even if there were second-philosophically persuasive evidence for the Theory of Ideas, an assumption we haven't questioned for now, Hume's route from there to radical skepticism depends on auxiliary premises she has no reason to accept, all stemming from Hume's insistence on the foundational role of the Science of Man and its introspective starting point. Let's now turn to Reid, Hume's admirer and critic, the self-proclaimed Philosopher of Common Sense.

2. REID

Reid begins his *Inquiry into the Human Mind, on the Principles of Common Sense*, sounding very much like Hume in the opening pages of the *Treatise*:

Wise men now agree, or ought to agree in this, that there is but one way to the knowledge of nature's works; the way of observation and experiment. By our constitution, we have a strong propensity to trace particular facts and observations to general rules, and to apply such general rules to account for other effects . . . This procedure of understanding is familiar to every human creature in the common affairs of life, and it is the only one

failure of the vulgar and philosophical beliefs to meet our norms of consistency and causal reasoning'. As we read through the tortuous twists and turns Hume pursues in his quest for the source of our idea of external existence, it's not hard to imagine Kant and Reid drawing a decidedly un-Humean conclusion from the proceedings—if it's not in the pure sensory data, and it obviously is there in our perceptual beliefs, then there must be some other element involved—and to feel some considerable sympathy for their position! [[See Essay #1.]]

25. Even if something like this correctly diagnoses the sources of Hume's descent into skepticism, it doesn't explain why so many contemporary philosophers, including those who disagree with Hume on many points, still find his transition from the Theory of Ideas to radical skepticism so easy to accept. This is a question for another occasion. [[See [2017a], II.]]

by which any real discovery in philosophy can be made. (Reid [1764], I.1, pp. 11–12)

He makes the familiar appeal to Newton:²⁶

The man who discovered that cold freezes water, and that heat turns it into vapour, proceeded on the same general principles, and in the same method, by which Newton discovered the law of gravitation and the properties of light . . . he who philosophizes by other rules . . . mistakes his aim. (Reid [1764], I.1, p. 12)

In the end,

[a]ll that we know of the body, is owing to anatomical dissection and observation, and it must be by an anatomy of the mind that we can discover its powers and principles. (Ibid.)

In all this, Reid's *Philosopher of Common Sense* sounds like a close cousin of the *Scientist of Man*.²⁷

There's considerably more methodological reflection in the first chapter of Reid's *Inquiry* ([1764]) and the first essay of his later *Essays on the Intellectual Powers of Man* ([1785]), but what must strike us most forcefully after our experience with Hume is this: when we turn the page to the second essay, 'On the powers we have by means of our external senses,' we find not introspective reports, but a detailed discussion of 'the organs of sense,' of

[t]he effluvia of bodies drawn into the nostrils with the breath . . . the medium of smell; the undulations of the air . . . medium of hearing; and the

26. As with Hume (see footnote 7), the Newtonian disdain for hypotheses is present here, too (see Reid [1764], I.1, p. 12). Reid cites Descartes's vortices, a particularly telling example (see [2008], pp. 21, 27, or [2011b], pp. 9–10, 18).

27. This may be just a stylistic variation, but given how differently they end up, I can't help noting that where Hume says we should never go 'beyond experience' (intro., 8), Reid says we should never go beyond 'just induction from facts' (Reid [1764], I.1, p. 12).

rays of light passing from visible bodies to the eye . . . the medium of sight.
(Reid [1785], II.2, p. 74; see also Reid [1764], VI.21)

We hear of ‘the impressions made on the organs of sense’ being ‘communicated to the nerves, and by them to the brain’ (Reid [1785], II.2, p. 75); we receive summaries of all that’s currently known about these processes, and of the wayward hypotheses philosophers have added and their ill effects. Unlike most philosophical treatments, the chapters of the *Inquiry* take up smell, taste, hearing, touch, and sight in that order, patiently sorting out the differences and inter-relations. The discussion of perception in the second *Essay* makes confident appeal to the geometry and optics of vision, to the structure of the eye, to experiments in human perception, even to comparisons with animals or ‘brutes’. There are even passages leaning toward something very like developmental psychology.²⁸

When it comes to investigating of the operations of the mind itself,²⁹ Reid carefully reviews the available methods. Given the state of knowledge in the late 18th century, it’s again unsurprising that he doesn’t include the

28. See, e.g., Reid [1785], II.22, pp. 248–249. In the poignant Reid [1764], I.2, p. 15, this is more a dream than a reality: ‘Could we obtain a distinct and full history of all that hath passed in the mind of a child, from the beginning of life and sensation, till it grows up to the use of reason; how its infant faculties began to work, and how they brought forth and ripened all the various notions, opinions, and sentiments, which we find in ourselves when we come to be capable of reflection; this would be a treasure of natural history, which would probably give more light into the human faculties, than all the systems of philosophers about them since the beginning of the world.’ Alas, Reid despairs: ‘It is in vain to wish for what nature has not put within the reach of our power. Reflection, the only instrument by which we can discern the powers of the mind, comes too late to observe the progress of nature, in raising them from their infancy to perfection’ (ibid.). Of course, this is just the sort of thing contemporary cognitive scientists have been able to chart (see [2007], III.5 and IV.2.ii, for summaries and references to some of this work).

29. Reid is a confirmed dualist—see, e.g., [1785], II.4, pp. 87–88: ‘[S]ome Philosophers . . . imagined that man is nothing but a piece of matter so curiously organized, that the impressions of external objects produce in it sensation, perception, remembrance, and all the other operations we are conscious of. . . . This [is a] foolish opinion . . . If one should tell of a telescope so exactly made as to have the power of seeing . . . it is the same absurdity to think that the impressions of external objects upon the machine of our bodies, can be the real efficient cause of thought and perception’—but this doesn’t keep him from regarding his study as an empirical one. In his account of 18th-century psychology, Hatfield ([1995], pp. 187–188) notes, ‘Many considered it to be a natural science based on experience, including those who considered themselves to be studying an immaterial substance . . . if one believes that immaterial entities exist and that some of them inhabit human bodies, it makes good sense to seek to determine the powers and capacities of such substances empirically, by studying the manifestation of the mind in the behavior of others, and in one’s own experience of mental phenomena.’

Nichols [2007], chapter 1, points out that Reid’s focus on the operations of the mind often allows him to sidestep questions of substance (e.g., pp. 20–21), but worries that Reid will ultimately be forced ‘to place the mind outside the realm of science’ (p. 37). In his unpublished replies to Priestley’s

hope of information from physiology, let alone neuroscience. For the experiential side of the mind, 'the chief and proper source . . . of knowledge is accurate reflection upon the operations of our own minds' (Reid [1785], I.5, p. 56). Reid emphasizes that this introspection is far from infallible; his discussion comes with repeated cautions on the various difficulties and pitfalls, and admonitions to care and practice in cultivating and improving one's powers of observation (Reid [1785], pp. 59–64).³⁰ But, unlike Hume, Reid doesn't limit our range of evidence even here to introspection; he describes two other sources as well.

The first comes as a surprise to those of us who imagine that ordinary language philosophy was invented in Oxford in the mid-20th century:

The language of mankind is expressive of their thoughts, and of the various operations of their minds. The various operations of the understanding, will, and passions, which are common to mankind, have various forms of speech corresponding to them in all languages, which are signs of them, and by which they are expressed: And a due attention to the signs may, in many cases, give considerable light to the things signified by them. (Reid [1785], I.5, p. 56)

Of course not all features of the mind are on display in language:

We can only expect, in the structure of languages, those distinctions which all mankind in the common business of life have occasion to make. (*Ibid.*)

This might easily be mistaken for a more recent declaration from Austin:

materialism, Reid deploys the Newtonian admonition against hypotheses, without notable success (see Tapper [2003]).

30. Reid distinguishes our 'consciousness' of the mind's operations—'all men are conscious of the operations of their own minds, at all times, while they are awake' (Reid [1785], I.6, p. 58)—from 'reflection', which involves active attention to those operations (see Reid [1785], VI.1, pp. 421–422). Smith [1990] argues that consciousness is the interior counterpart to sensation, and reflection the counterpart to perception. In any case, it's reflection that plays a central role in Reid's investigation of the mind, and accurate reflection requires both skill and extensive practice (see Reid [1785], I.6, p. 59). (This point comes up below, in connection with Reid's first objection to the Theory of Ideas.) I use the term 'introspection' more or less for Reid's 'reflection' on the acts of one's mind.

Our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking, in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters, than any that you or I are likely to think up in our arm-chairs of an afternoon—the most favored alternative method. (Austin [1956a], p. 182)

Even the dig at a priori philosophizing might resonate with Reid. In his scholarly commentary, Lehrer remarks, ‘I have no evidence of J. L. Austin having read Reid. I conjecture that Moore was the conduit of Reidian philosophy to twentieth-century Cambridge’ (Lehrer [1989], p. 93).³¹

The similarities here run deep. Both Reid and Austin begin from a keen appreciation for the many ways mis-use of ordinary words and failure to define technical terms can distort our understanding:³² for examples, see Reid on ‘idea’ and ‘impression’³³ or Austin on ‘real’ and ‘direct’.³⁴ Both bemoan the ill effects of false dichotomies;³⁵ both regard the use of mockery as often the most appropriate response to views that fly in the face of common sense.³⁶

31. In this passage, Lehrer is discussing speech act theory, which also has its precursor in Reid, not the more general recommendation that philosophers attend to ordinary language, but the resemblance is clear in both cases. Lehrer traces the direct influence of Reid on Moore in his [1976].

32. E.g., Reid [1785], I.1, p. 38: ‘When we use common words, we ought to use them in the sense in which they are most commonly used by the best and purest writers in the language; and, when we have occasion to enlarge or restrict the meaning of a common word, or give it more precision than it has in common language, the reader ought to have warning of this, otherwise we shall impose upon ourselves and upon him! E.g., Austin [1962], p. 19: ‘Thus, it is quite plain that the philosophers’ use of ‘directly perceive’, whatever it may be, is not the ordinary, or any familiar use . . . But we are given no explanation or definition of this new use—on the contrary, it is glibly trotted out as if we were all quite familiar with it already’.

33. For a start, see the relevant passages of ‘Explication of words’ (Reid [1785], I.1, pp. 27–36), but discussions of these words run throughout Reid’s writings.

34. E.g., see Austin [1962], pp. 14–19, 62–77.

35. E.g., Reid [1785], I.7, p. 67: ‘There is not a more fruitful source of error in this branch of philosophy, than the division of things which are taken to be complete when they are not really so’. E.g., Austin [1962], p. 4: ‘The question, do we perceive material things or sense-data, no doubt looks very simple—too simple . . . There is no *one* kind of thing that we ‘perceive’ but many *different* kinds, the number being reducible if at all by scientific investigation and not by philosophy’.

36. With Reid, this is an official doctrine: ‘It is in vain to reason with a man who denies [common sense] . . . To discountenance absurdity, Nature hath given us a particular emotion, to wit, that of ridicule . . . the novelty of an opinion to those who are too fond of novelties; the gravity and solemnity with which it is introduced; the opinion we have entertained of the author; its apparent connection with

Both quite reasonably allow that ordinary language is sometimes shaped by ignorance and prejudice, but both are optimistic that we have the tools necessary to uncover these lapses.³⁷ Both also look unabashedly to science.³⁸ The main difference, it seems to me, is that Reid is interested in the broad picture of the human mind—of how it perceives the world—so he attends mostly to what we nowadays call ‘linguistic universals’:

There may be peculiarities in a particular language, of the causes of which we are ignorant, and from which, therefore, we can draw no conclusion. But whatever we find common to all languages, must have a common cause; must be owing to some common notion or sentiment of the human mind. (Reid [1785], I.5, p. 56)

In contrast, Austin focuses on narrower issues, often in moral philosophy, which leads him to a more fine-grained examination, for example, of the notion of excuses, conducted almost exclusively in English.³⁹ But this appears

principles already embraced, or subserviency to interests which we have at heart; and, above all, its being fixed in our minds at a time of life when we receive implicitly what we are taught; may cover its absurdity, and fascinate the understanding for a time. But if ever we are able to view it naked, and stripped of those adventitious circumstances from which it borrowed its importance and authority, the natural emotion of ridicule will exert its force’ (Reid [1785], VI.4, pp. 460, 462, 463). With Austin, the method is simply used: e.g., ‘[I]f we are going to talk about ‘real’, we must not dismiss as beneath contempt such humble but familiar expressions as ‘not real cream’; this may save us from saying, for example, or seeming to say that what is not real cream must be a fleeting product of our cerebral processes’ (Austin [1962], pp. 63–64).

37. E.g., Reid [1785], I.1, pp. 26–27: ‘A Philosopher is, no doubt, entitled to examine even those distinctions that are to be found in the structure of all languages; and, if he is able to shew that there is no foundation for them in the nature of the things distinguished; if he can point to some prejudice common to mankind which has led them to distinguish things that are not really different; in that case, such a distinction may be imputed to a vulgar error, which ought to be corrected in philosophy.’ (Reid points out such a vulgar error, a false analogy, in our tendency to treat the mind as ‘some subtle matter, like breath or wind . . . in almost all languages’ (Reid [1785], II.17, p. 205; for another example, see I.1, pp. 25–26). E.g., Austin [1956a], p. 185: ‘[I]t must be added too, that superstition and error and fantasy of all kinds do become incorporated in ordinary language and even sometimes stand up to the survival test (only, when they do, why should we not detect it?).’

38. Of course, Reid appeals to science throughout. For Austin, see, e.g., his [1956a], p. 189: ‘[T]he third source-book is psychology, with which I include such studies as anthropology and animal behavior . . . some varieties of behavior, some ways of acting or explanations of doing actions, are here noticed and classified which have not been observed or named by ordinary men and hallowed by ordinary language.’

39. E.g., see the priceless Austin [1956a], p. 185: ‘You have a donkey, so have I, and they graze in the same field. The day comes when I conceive a dislike for mine. I go to shoot it, draw a bead on it, fire: the brute falls in its tracks. I inspect the victim, and find to my horror that it is *your* donkey. I appear on

to be a difference of focus, not of method, and the method itself seems perfectly amenable to the Second Philosopher.⁴⁰

Reid's final source of information about the operations of the mind—after introspection and attention to ordinary language—is careful observation of our fellow humans. First their behavior:

The actions of men are effects: Their sentiments, their passions, and their affections, are the causes of those effects; and we may, in many cases, form a judgment of the cause from the effect. (Reid [1785], I.5, p. 57)

For example:

It is obvious, from the conduct of men in all ages, that man is by his nature a social animal; that he delights to associate with his species; to converse, and to exchange good offices with them. (Ibid.)

Here we find the behavioral study of men in the marketplace that was so unexpectedly missing in Hume. But evidence from their conduct isn't all Reid gathers from observation of his fellows; their opinions can also be illuminating:

The opinions of men may be considered as the effects of their intellectual powers . . . Even the prejudices and errors of mankind, when they are general, must have some cause no less general; the discovery of which will throw some light upon the frame of human understanding. (Ibid.)

Indeed, Reid suggests, this is one of the main benefits of studying the history of philosophy.

your doorstep with the remains and say—what? 'I say, old sport, I'm awfully sorry, etc., I've shot your donkey *by accident*'? or '*by mistake*'? Then again, I go to shoot my donkey as before, draw a bead on it, fire—but as I do so, the beasts move, and to my horror yours falls. Again the scene on the doorstep—what do I say? 'By mistake'? Or 'by accident'? To be fair, Reid is capable of some subtle distinctions of his own (see, e.g., Reid [1785], I.1, pp. 34–35).

40. For a more complete second-philosophical discussion of Austin's method, see [2011a]. [[Or better, [2017a]. Also Essay #6.]]

After our recent encounter with Hume, I think we can all agree that here, with Reid, we find ourselves in a completely different world, undertaking a completely different type of inquiry—and one dramatically more congenial to the Second Philosopher's sensibilities. Given that Hume and Reid begin with such similar-sounding rhetoric, we want to know how Reid manages to navigate so different a course. Part of the answer must lie in his well-known critique of the Theory of Ideas. Reid relates with gusto how Descartes embraced this theory, how Locke and Berkeley followed in his footsteps, and how Hume drew the ultimate conclusion of radical skepticism: Descartes 'did what he could to shut it out,' Berkeley gave up the material world to preserve 'the world of spirits,' but Hume 'wantonly sapped the foundation of this partition, and drowned all in a universal deluge' (Reid [1764], I.7, p. 23).⁴¹ How should we respond to this calamity? Reid answers:

A traveler of good judgment may mistake his way, and be unawares led into a wrong track; and while the road is fair before him, he may go on without suspicion, and be followed by others; but when it ends in a coal-pit, it requires no great judgment to know that he hath gone wrong, nor perhaps to find out what misled him. (Reid [1764], I.8, p. 23)

With the Second Philosopher, Reid suggests that the skeptical conclusion should inspire us to re-examine the methods that landed us there, and Reid further proposes that the culprit is the Theory of Ideas.

Now we raised some doubts in §1 about the notion that the Theory of Ideas by itself is enough to generate skepticism, but let's set this aside for the moment to focus on how Reid builds his case against what he sometimes calls 'the ideal system' (Reid [1764], I.7, p. 23, and elsewhere). This will eventually return us to the question left hanging in our exploration of Hume: do the arguments from perceptual relativity, illusion, and hallucination actually

41. Kant wasn't the only one awakened from a dogmatic slumber by Hume. In a letter to Hume (reproduced in Reid [1764], p. 264), Reid writes, 'Your system appears to me not onely coherent in all its parts, but likewise justly deduced from principles commonly received among Philosophers: Principles, which I never thought of calling in Question, untill the conclusions you draw from them in the treatise of humane Nature made me suspect them.' [[See Essay #1 for more on Reid and Kant.]]

establish the Theory of Ideas? So, to begin, what are Reid's objections to the Theory of Ideas?

Reid describes the theory this way:

By the impressions made on the brain, images are formed of the object perceived . . . the mind, being seated in the brain . . . immediately perceives those images only, and has no perception of the external object but by them. This notion of our perceiving external objects, not immediately, but in certain images or species of them conveyed by the senses, seems to be the most ancient philosophical hypothesis we have on the subject of perception, and to have with small variations retained its authority to this day. (Reid [1785], II.4, p. 90)⁴²

We should note that when Reid speaks of 'impressions made on the brain,' he's using the term in his own way:

There is therefore sufficient reason to conclude, that, in perception, the object produces some change in the organ [of sense]; that the organ produces some change upon the nerve; and that the nerve produces some change in the brain. And we give the name of an *impression* to those changes. . . . Whatever may be the nature of those impressions upon the organs, nerves and brain, we perceive nothing without them. Experience informs that it is so . . . In the constitution of man, perception, by fixed laws of nature, is connected with those impressions. (Reid [1785], II.2, pp. 75–76)

By contrast, 'perception' for Reid, is a mental act:⁴³

42. See also Reid [1785], I.1, p. 31: 'When . . . in common language, we speak of having an idea of any thing, we mean no more by that expression, but thinking of it. The vulgar allow, that this expression implies a mind that thinks; an act of that mind which we call thinking, and an object about which we think. But, besides these three, the Philosopher conceives that there is a fourth, to wit, the *idea*, which is the immediate object. The idea is in the mind itself and can have no existence but in the mind that thinks; but the remote or mediate object may be something external, as the sun or moon; it may be something past or future; it may be something which never existed.'

43. Cf. footnote 29. [[Reid [1764], VI.20, p. 168, observes that 'the perception of an object implies both a conception of its form, and a belief in its present existence . . . moreover . . . this belief is not the effect of argumentation and reasoning' (see also his [1785], II.5, p. 96). Contemporary commentators often take this as an entry in the debate over the proper analysis of 'perception', but just one page

In speaking of the impressions made on our organs in perception, we build upon facts borrowed from anatomy and physiology, for which we have the testimony of our senses. But being now to speak of perception itself, which is solely an act of the mind, we must appeal to some other authority. (Reid [1785], II.5, p. 96)

He alludes here to introspection, but as we've seen, he also countenances other avenues for investigation of the mind. In any case, the Theorist of Ideas posits that perception of an external object is indirect; that is, it proceeds by means of a mental image or representation of the object that is itself directly perceived.

Reid raises three main objections to this theory, apart from the claim that it leads to skepticism, perhaps the strongest counter-argument of all. The first objection is that the theory conflicts with common sense:

It is directly contrary to the universal sense of men who have not been instructed in philosophy. When we see the sun or the moon, we have no doubt that the very objects which we immediately see, are very far distant from us, and from one another. We have not the least doubt that this is the sun and moon which God created some thousands of years ago, and which have continued to perform their revolutions in the heavens ever since. But how are we astonished when the Philosopher informs us, that we are mistaken in all this; that the sun and moon which we see are not, as we imagine, many miles distant from us, and from each other, but that they are in our own mind; that they have no existence before we saw them, and will have none when we cease to perceive and to think of them; because the objects we perceive are only ideas in our own minds, which can

previous, in the introductory paragraph of this section, 'Of perception in general,' Reid cautions: 'The most simple operations of the mind, admit not of a logical definition: all we can do is to describe them, so as to lead those who are conscious of them in themselves, to attend to them, and reflect upon them: and it is often very difficult to describe them so as to answer to this intention' (ibid., p. 167). Here Reid simply notes features he typically finds when he introspects on ordinary cases of perception and invites the reader to do the same: 'I can only appeal to the reader's own thoughts, whether such facts are not agreeable to what he is conscious of in his own mind' (Reid [1785], II.5, p. 96). He isn't proposing necessary and sufficient conditions.]]

have no existence a moment longer than we think of them. (Reid [1785], II.14, p. 172)

Here Austin might interject that the plain man has no opinion whatever about what we do or don't see 'immediately' or 'mediately,' that these ordinary words are being used in a special philosophical sense that hasn't been clearly specified.⁴⁴ In any case, I see no obstacle to the Theorist of Ideas employing the more circumspect terminology of our imaginary empirical representation theorist, saying that we perceive the moon by a process of inference that begins from our awareness of an inner representation. This needn't conflict with the commonsensical notion that we see the moon.

But it might conflict with common sense in another way, or perhaps we should say, with ordinary introspective evidence: we don't seem to be aware of any such representation. In his own introspective analysis, Reid observes:

When I smell a rose . . . the agreeable odor I feel, considered by itself, without relation to any external object, is . . . a sensation . . . This sensation can be nothing else than it is felt to be. Its very essence consists in its being felt; and when it is not felt, it is not. There is no difference between the sensation and the feeling of it; they are one and the same thing . . . in sensation, there is no object distinct from that act of the mind by which it is felt; and this holds true with regard to all sensations. (Reid [1785], II.16, p. 194)

Here Reid insists that the sensation is identical with the act of mind of feeling it,⁴⁵ but whether it's an object or an act, the important point is that it doesn't function as an Idea is supposed to do. First, we're most often entirely unaware of our sensations:

44. E.g., Austin [1962], pp. 14–15: 'Philosophers, it is said, 'are not, for the most part, prepared to admit that such objects as pens or cigarettes are ever directly perceived'. Now of course what brings us up short here is the word 'directly'—a great favorite among philosophers, but actually one of the less conspicuous snakes in the linguistic grass. We have here, in fact, a typical case of a word, which already has a very special use, being gradually stretched, without caution or definition or any limit, until it becomes, first perhaps obscurely metaphorical, but ultimately meaningless. One can't abuse ordinary language without paying for it'. He elaborates on pp. 15–19.

45. See also Reid [1764], VI.20, pp. 167–168.

The operations of the mind, from their nature, lead the mind to give its attention to some other object. Our sensations . . . turn our attention to the things . . . so much, that most [sensations], and those most frequent and familiar, have no name in any language. In perception . . . there is an object distinct from the operation itself; and, while we are led by a strong impulse to attend to the object, the operation escapes our notice. (Reid [1785], I.5, pp. 60–61)^{46,47}

Second, sensations are non-representational, non-conceptual:

Sensation, taken by itself, implies neither the conception nor belief of any external object. It supposes a sentient being, and a certain manner in which that being is affected, but it supposes no more. (Reid [1785], II.16, p. 199)

Finally, though these content-less sensations figure in a causal process that ends in perception, our perceptual beliefs are immediate, not inferred from anything:

The clear and distinct testimony of our senses carries irresistible conviction along with it, to every man in his right judgment . . . this conviction is

46. For sensations of touch in particular, see Reid [1764], V.5, pp. 63–64, or [1785], II.16, pp. 195–196; for visual sensations, see Reid [1764], VI.3, pp. 82–83 (also VI.8, p. 102): ‘I cannot . . . entertain the hope of being intelligible to those readers who have not, by pains and practice, acquired the habit of distinguishing the appearance of objects to the eye, from the judgment which we form by sight of their colour, distance, magnitude and figure. The only profession in life wherein it is necessary to make this distinction is that of painting. The painter hath occasion for an abstraction, with regard to visible objects, somewhat similar to what we require; and this indeed is the most difficult part of his art. For it is evident, that if he could fix in his imagination the visible appearance of objects, without confounding it with the things signified by that appearance, it would be as easy for him to paint from life, and to give every figure its proper shading and relief, and its perspective proportions, as it is to paint from a copy’.

47. One might argue that on occasions when we *are* aware of our sensations, they do function as Ideas do in the Theory of Ideas, that is, as the inferential basis for our knowledge of external objects. Given the capriciousness of our attending or not, it seems more likely that sensations are performing the same function in all perceptions, whether they’re noticed or not. Indeed, Reid distinguishes two roles for the senses: ‘to make us feel, and to make us perceive’ (Reid [1785], II.17, p. 210). While perception gives us information about the world, ‘the painful sensations . . . are admonitions to avoid what would hurt us; and the agreeable sensations . . . invite us to those actions that are necessary to the preservation of the individual, or of the kind’ (Reid [1785], II.16, p. 198). Nichols [2007], chapter 5, develops this ‘teleological’ interpretation of Reid on sensations.

not only irresistible, but it is immediate; that is, it is not by a train of reasoning and argumentation that we come to be convinced of the existence of what we perceive; we ask no argument for the existence of an object, but that we perceive it; perception commands our belief upon its own authority, and disdains to rest its authority upon any reasoning whatsoever. (Reid [1785], II.5, p. 99)

Thus Reid takes the non-existence of Ideas in the relevant sense to be evident on careful introspection into the operations of our minds.

Of course, as we've noted, Reid departs from many of his predecessors and successors in that he denies introspection is perfectly reliable.⁴⁸ In fact, he describes it as a difficult thing to do properly:

[T]he number and quick succession of the operations of the mind make it difficult to give due attention to them . . . (Reid [1785], I.6, p. 60)

[W]hen [an operation of the mind] is exerted, we are conscious of it; but then we do not attend to the operation, but to its object. When the mind is drawn off from the object to attend to its own operation, that operation ceases, and escapes our notice. (Reid [1785], I.6, p. 61)

The same precision in the use of words; the same cool attention to the minute differences of things; the same talent for abstraction and analyzing, which fits a man for the study of mathematics, is no less necessary in this. But there is this great difference between the two sciences, that the objects of mathematics being things external to the mind, it is much more easy to attend to them, and fix them steadily in the imagination. (Reid [1785], I.6, p. 61)

48. See Hurlebert and Schwitzgebel [2007] for a recent discussion. Another of Reid's examples involves the case of double vision, to be considered below ([1764], VI.13, pp. 134–135): when we attend to a nearby object, those in the background appear double; nevertheless, 'you may find a man that can say with a good conscience, that he never saw things double all his life . . . in order to see things double, at least in order to have any reflection or remembrance that he did so, it is necessary that he should look at one object, and at the same time attend to the faint appearance of other objects which are within the field of vision. This is a practice which perhaps he never used, nor attempted; and therefore he does not recollect that ever he saw an object double.'

Nevertheless introspection is one important type of evidence, and here, Reid claims, it counts against the Theory of Ideas.⁴⁹

Reid's second objection against the Theory is that it doesn't account for the phenomenon it sets out to explain:

We are at a loss to know how we perceive distant objects . . . Ideas in the mind seem to account for [this] . . . by the means of ideas, [perception is] reduced to . . . a kind of feeling or immediate perception of things present, and in contact with the percipient; and feeling is an operation so familiar, that we think it needs no explication.

But this feeling or immediate perception, is as difficult to be comprehended, as the things which we pretend to explain by it. Two things may be in contact without any feeling or perception; there must therefore be in the percipient a power to feel or to perceive. How this power is produced, and how it operates, is quite beyond the reach of our knowledge. (Reid [1785], II.14, p. 185)

This point also seems well-taken: we don't understand how we perceive the moon; how does it help to say instead that we perceive an idea of the moon?

(Notice that our imaginary empirical Theory of Ideas needn't leave the matter in this unsatisfactory state. It would include a chapter on what constitutes 'awareness' of the internal representation—perhaps an account growing out of the study of phenomena like blindsight,⁵⁰ an account that isolates the neural underpinning that makes some brain activity conscious and some not—and since we're imagining here, let's add an explanation of the evolutionary advantage to our being aware of some mental activities and unaware of others.⁵¹ Its explanation of the reliability of perception would involve a range of external considerations; for example, the accuracy of our calculations of distance from the disparity between our two retinal images

49. Including, of course, my imaginary empirical version. My interest isn't in the truth of the imaginary theory, but in the claim that it doesn't, by itself, lead inevitably to skepticism.

50. See, e.g., Weiskrantz [1997].

51. This last makes the imaginary empirical theory even more unlikely, because we're probably better off not being aware of the bulk of the computational processes that underlie perception.

would depend on optics, geometry and so on.⁵² All this, if it came to pass, Reid would presumably applaud and embrace, but the Theory of Ideas he was presented with enjoyed no such virtues. And, as we've seen, if Hume had his way, it never would, because the sort of scientific investigation involved would be out of order in his Science of Man.)

Finally, Reid's third objection to the Theory of Ideas is that the arguments offered for it are ineffective.⁵³ Here he gives pride of place to an argument from perceptual relativity found in Hume's *Enquiry*:

The table, which we see, seems to diminish, as we remove farther from it:
But the real table, which exists independent of us, suffers no alteration:
It was, therefore, nothing but its image, which was present to the mind.
(Hume [1748], 12.1.9)

Reid responds by distinguishing the real magnitude of the table from its apparent magnitude, which 'is measured by the angle which an object subtends at the eye' (Reid [1785], p. 181). He then reconsiders Hume's case:

The table we see seems to diminish as we remove farther from it; that is, its apparent magnitude is diminished; but the real table suffers no alteration,

52. Incidentally, Reid followed Berkeley on perception of distance; e.g., see Reid [1785], II.19, pp. 235–236: 'By sight . . . we perceive visible figures to have extension in two dimensions . . . By touch . . . we perceive originally [bodies'] three dimensions . . . by experience, we learn to perceive [three dimensions] by sight'. In his historical overview, Crone [1992] argues that the rise of empiricism delayed the discovery of distance perception by stereopsis (i.e., from the disparity between the two retinal images) until Wheatstone in 1838. Reid's willingness to posit elements of perception not present in sensation (see the next footnote) made him a likely candidate to have figured this out, but he didn't. (Ironically this very Berkeleyan error led him to develop a sort of non-Euclidean geometry in his [1764], VI.9. See van Cleve [2002] for a recent discussion.)

53. I consider here the discussion in Reid [1785], pp. II.14, pp. 174–184. In his [1764], V.7, pp. 69–70, Reid treats only one *experimentum crucis*—are there ideas that aren't derived from sensation?—and he argues that our conceptions of hardness, extension, figure, and so on, are not. Earlier work from which the *Inquiry* largely derives apparently includes a fuller array of arguments; commentators consider the possibility that Hume's reactions to the book manuscript moved Reid to delete this material (see Brookes [1997], pp. xvii–xx). In any case, the argument of the *Inquiry* speaks more directly to the empiricist assumption that all ideas originate in impressions (in Hume's terms) or sensations (in Reid's terms) than to the more familiar arguments for what we now call sense-data. Reid gets quite caught up in his analysis of what is or isn't present in sensation, partly as a result of his Berkeleyan holdover touched on in the previous footnote. Nowadays it seems less clear than it must have in the 18th century that there is such a thing as a pure sensation, unaffected by belief, etc.

to wit, in its real magnitude; therefore it is not the real table we see: I admit both the premises in this syllogism, but I deny the conclusion. The syllogism has what the Logicians call two middle terms: Apparent magnitude is the middle term in the first premise; real magnitude in the second. Therefore, according to the rules of logic, the conclusion is not justly drawn from the premises. (Reid [1785], p. 182)

Indeed Reid takes these phenomena as evidence for his own conclusion:

The real table may be placed successively at a thousand different distances . . . and it can be determined demonstratively, by the rules of geometry and perspective, what must be its apparent magnitude . . . in each of those distances . . . Let the table be placed successively in as many of those different distances . . . as you will . . . open your eyes and you shall see a table precisely of that apparent magnitude . . . which the real table must have in that distance . . . Is not this a strong argument that it is the real table that you see? (Reid [1785], p. 183)

In other words, it's not that what I really see is a percept of the table; it's that the thing I really see, the real table, subtends a different angle on the retina and consequently looks different at different distances, and in entirely predictable ways. My perception of the apparent magnitude is accurately registering an entirely objective feature of the relation between the object and the eye.⁵⁴

Reid tells us this is 'all I have found in Mr. Hume's writings upon this point' (Reid [1785], II.14, p. 179), so he must have missed or forgotten the argument from illusion based on double vision noted a moment ago from the *Treatise*, as well as various arguments from hallucination.⁵⁵ Nonetheless,

54. See Nichols [2007], chapter 4, for a subtle discussion of the complexities of Reid's 'visible figure'. Here, incidentally, is another point of contact with Austin, who also insists that many so-called illusions aren't illusory at all: 'What is even faintly surprising . . . in the idea of a stick's being straight but looking bent sometimes? Does anyone suppose that if something is straight, then it jolly well has to look straight in all circumstances?' (Austin [1962], p. 29). De Bary [2002], pp. 54–56, expands on this parallel between Austin and Reid.

55. I should mention that Reid ([1785], II.14, pp. 175–177) also considers arguments of the form: how could an external object act where it is not (i.e., in the mind)? Strangely, Reid replies by doubting that 'when we perceive objects, either they act upon us, or we act upon them' (p. 176). As Van Cleve points out ([2004], p. 102), Reid could have replied that the external object needn't act directly, that

he discusses both double vision and the phantom pain for their independent interest. On the subject ‘of seeing objects single with two eyes’ (Reid [1764], VI.13), he reports a series of experiments he’s performed and concludes among other things that we see an object double only when the two impressions don’t fall on the corresponding areas of the two retinas. Thus, when we ‘press one eye with a finger’ (Hume [1739], 1.4.2.45) as Hume proposes, one of the retinal impressions is displaced, so we should expect to see objects double. His results, Reid assures us, hold

[i]nvariably in all perfect human eyes, as far as I am able to collect from innumerable trials of various kinds made upon my own eyes, and many made by others at my desire . . . This general phenomenon appears therefore to be founded upon a very full induction. (Reid [1764], VI.13, p. 137)⁵⁶

Thus seeing an object double is also a perfectly predictable phenomenon, but it needn’t follow that it isn’t the object that we see.

As for phantom pain, Reid notes that disagreeable and agreeable sensations are unlike many others in that we do tend to notice them. Unfortunately, we also tend to conflate the sensation with its cause:

When we consider the sensation of pain by itself, without any respect to its cause, we cannot say with propriety, that the toe is either the place, or the subject of it. But it ought to be remembered, that when we speak of a pain in the toe, the sensation is combined in our thought, with the cause of it, which really is in the toe. The cause and effect are combined in one complex notion, and the same name serves for both. It is the business of the Philosopher to analyse this complex notion, and to give different names to its different ingredients. He gives the name of *pain* to the sensation only, and the name of *disorder* to the unknown cause of it. Then it is

a chain leads from the object to the perceiver, like the lighting of the fuse here to the explosion of the bomb over there. This response coheres far better with Reid’s understanding of perception, as is obvious even from the few passages quoted in the text (though Reid would prefer to say that one link in the chain ‘produces’ or ‘occasions’ rather than ‘causes’ the next, for reasons unrelated to the point at issue here), so I assume Reid would have viewed it as a friendly amendment.

56. Passages like this in Reid’s writings make Hume’s reference to ‘experiments’ at [1739], 1.4.2.45, sound a bit flat.

evident that the disorder only is in the toe, and that it would be an error to think that the pain is in it. (Reid [1785], II.18, p. 213)

So what happens when ‘a man who has had his leg cut off, many years after feels a pain in a toe of that leg’ (Reid [1785], II.18, p. 214)?

Nature has connected our perception . . . with certain sensations. If the sensation is produced, the corresponding perception follows even when there is no object, and in that case it is apt to deceive us. In like manner, Nature has connected our sensations with certain impressions that are made upon the nerves and the brain: And, when the impression is made, from whatever cause, the corresponding sensation and perception immediately follows. Thus, in the man who feels pain in his toe after the leg is cut off, the nerve that went to the toe, part of which was cut off with the leg, had the same impression made upon the remaining part, which, in the natural state of his body, was caused by a hurt in the toe: And immediately this impression is followed by the sensation and perception which Nature connected with it. (Reid [1785], II.19, p. 214)

Here, Reid is happy to agree that the perception (though not the sensation) in such cases is ‘fallacious’ (ibid.). But this hardly shows that the false perception, or any perception, is inferred from a sensation.

In all this, we find Reid offering the beginnings of an empirical account of how and why the world appears to us as it does—including those experiences normally cited in the standard arguments—that doesn’t appear to involve anything playing the role of an idea or sense-datum. (And, as noted, my understanding is that a contemporary Second Philosopher of vision would agree on this last point.) So clearly one way in which Reid’s world differs dramatically from Hume’s is that it does without ‘the ideal theory’. But I think it would be a mistake to identify this as the most important difference, as you might expect from my attempt to downplay the role of the Theory in generating Hume’s conclusions. Furthermore, I think that Reid himself sometimes sees it this way, despite his much more prevalent and conspicuous denunciations of the Theory of Ideas. To help highlight what I take to

be Reid's fundamental contribution, let me return for a moment to Stroud's discussion of Hume.

Stroud sees Hume's naturalism as his greatest achievement:

He was interested in human nature, and his interest took the form of seeking extremely general truths about how and why human beings think, feel and act in the ways they do . . . These questions were to be answered in the only way possible—by observation and inference from what is observed. Hume saw them as empirical questions about natural objects within the sphere of human experience, so they could be answered only by an admittedly general, but none the less naturalistic, investigation . . . Of all the ingredients of lasting significance in Hume's philosophy I think this naturalistic attitude is of greatest importance and interest. (Stroud [1977], p. 222)

In this, Stroud describes Hume as rejecting the traditional philosophical approach to the study of human nature:

Hume's theory sees every aspect of human life as naturalistically explicable. It places man squarely within the scientifically intelligible world of nature, and thus conflicts with the traditional conception of a detached rational subject. (Stroud [1977], p. 13)

Of course something goes badly wrong in this sunny picture. Stroud sums up the state of play after the barren rock like this:

If we remain within the traditional philosophical theory we will inevitably regard ourselves as worse off in ordinary life than we would have originally supposed. But if the real discovery comes not with the philosophical conviction itself, but in an appreciation of the source of the instability and transience of that philosophical conviction [here Stroud alludes to the fact that Hume can't maintain his skepticism over the backgammon table], then we might no longer regard ourselves as so badly off. If we see that we simply do not, and cannot, operate according to the traditional philosophical conception of reasonable belief and action, it is just possible that our

dissatisfactions will then be directed onto that conception itself, and not onto our ordinary life which is seen not to live up to it. Of course, as long as the Cartesian philosophical model [of a detached rational subject] is thought to embody the conception of reasonableness we actually try to fulfill in everyday life and science, that result will not be forthcoming.^[57] What is needed, then, and would be completely in the spirit of Hume's 'experimental' examinations of human nature, is an alternative description of how we actually proceed in everyday life, and what we regard as essential to the most reasonable beliefs and actions we find there.^[58] Hume does not suggest even the beginnings of such a quest, probably because the theory of ideas makes it unthinkable to him, but once we escape the theory of ideas there is nothing in Hume's general picture of the proper study of man that would rule out an alternative to the traditional philosophical picture. (Stroud [1977], p. 117)

What I've suggested here is that Hume fails to realize his naturalistic goals not because he embraces the Theory of Ideas but because he presupposes a kind of foundationalism that surely qualifies as a 'traditional philosophical picture' and as 'Cartesian': all our reasonings in the Science of Man must begin from introspection of our current experience.

Toward the end of the *Essay*, Reid himself tells a similar story, again beginning with Descartes:

There is, no doubt, a beauty in raising a large fabric of knowledge upon a few first principles. The stately fabric of mathematical knowledge, raised upon the foundation of a few axioms and definitions, charms every beholder. DES CARTES, who was well acquainted with this beauty in the mathematical sciences, seems to have been ambitious to give the same beautiful simplicity to his system of philosophy; and therefore sought

57. 'We will lament our ordinary 'failures' to live up to the picture, rather than the artificiality and irrelevance of the only picture we have'.

58. 'The Cartesian picture is certainly more than a mere *a priori* prejudice; there are powerful considerations in its favor. But it remains to be seen that there is no alternative that will accommodate those considerations while providing a more naturalistic and hence more palatable conception of how and why we think and act as we do'.

only one first principle as the foundation of all our knowledge, at least of contingent truths.

And so far has his authority prevailed, that those who came after him have almost universally followed him in this track. This, therefore, may be considered as the spirit of modern philosophy, to allow of no first principles of contingent truths but this one, that the thoughts and operations of our own minds, of which we are conscious, are self-evidently real and true; but that everything else that is contingent is to be proved by argument. (Reid [1785], VI.7, p. 516)

This method of philosophizing is common to DES CARTES, MALEBRANCHE, ARNAULD, LOCKE, NORRIS, COLLIER, BERKELEY and HUME; and as it was introduced by DES CARTES, I call it the Cartesian system. (Reid [1785], VI.7, p. 525)

The result of this Cartesian system, Reid continues, is Hume's shipwreck:

From this single principle of the existence of our own thoughts, very little, if anything, can be deduced by just reasoning, especially if we suppose that all our other faculties may be fallacious . . . Mr HUME seem[s] to me to have reasoned more consequentially from DES CARTES[s] principle than he did himself; and indeed I can't help thinking, that all who have followed DES CARTES[s] method . . . have escaped the abyss of scepticism by the help of weak reasoning and strong faith more than by any other means. (Reid [1785], VI.7, p. 518)⁵⁹

Here, unlike the similar passage quoted a while ago, the culprit isn't the Theory of Ideas itself.⁶⁰ Even if we held instead that we aren't aware of any

59. Cf. Reid [1785], VI.7, p. 526, which follows directly on the previous quotation in the text: 'Some of these have gone to the utmost length in skepticism, leaving no existence in Nature but that of ideas and impressions. Some have endeavoured to throw off the belief of a material world only, and to leave us ideas and spirits. All of them have fallen into very gross paradoxes, which can never sit easy upon the human understanding, and which, though adopted in the closet, men find themselves under a necessity of throwing off and disclaiming when they enter into society . . . Indeed, in my judgment, those who have reasoned most acutely and consequentially upon this system, are they that have gone deepest into skepticism.'

60. The suggestion that Reid is more fundamentally opposed to 'foundationalism' than to 'representationalism' is not uncommon in the literature—see, e.g., Loeb [2007], p. 86, and his references

intermediary, that our perceptual beliefs are immediate, this wouldn't remove the concern that all such beliefs could be false: if all we have to appeal to is the evidence of introspection, if our senses aren't to be trusted until they're proved reliable on that basis, then—to put it colorfully—an Evil Demon could presumably manipulate those immediate beliefs as easily as he does our sense-data; the Veil of Belief is just as devastating as the Veil of Ideas. What does the work here, what rules out any empirical case for the reliability of our perceptual beliefs on either model, is the underlying foundational principle, the restriction of trustworthy evidence to introspection alone—just as we found in our examination of Hume's foundationalism.

So, if this Cartesian model is the true culprit, and Hume was unable to offer a viable alternative, what does Reid offer in its place? He begins, as Stroud would have him, from the observation that we don't, in fact, live up to the Cartesian model:

It is, no doubt, the perfection of a rational being to have no belief but what is grounded on intuitive evidence, or on just reasoning: But man, I apprehend, is not such a being; nor is it the intention of Nature that he should be such a being. (Reid [1785], II.21, p. 238)

And, again as Stroud would have him, he sees this as the fault of that model:

We come into the world without the exercise of reason; we are merely animal before we are rational creatures; and it is necessary for our preservation, that we should believe many things before we can reason. How then is our belief to be regulated before we have reason to regulate it? has Nature left it to be regulated by chance? By no means. It is regulated by certain principles which are parts of our constitution; whether they ought to be called animal principles, or instinctive principles, or what name we give to them, is of small moment; but they are certainly different from the faculty of reason: They do the office of reason while it is in its infancy, and

to Wolterstroff [1987] and de Bary [2002]—but I think the specifics of what these terms are taken to entail will vary from one commentator to the next.

must as it were be carried in a nurse's arms, and they are leading strings to it in its gradual progress. (Reid [1785], II.21, pp. 238–239)

Of course, as we've seen, Reid freely admits that the senses are fallible:

Deceptions of sense . . . proceed from some disorder or preternatural state, either of the external organ, or of the nerves and the brain, which are internal organs of perception. (Reid [1785], II.22, p. 251)

But the Cartesian's focus on the shortcomings of sense perception ignores the equally obvious fact that our other faculties are also fallible. I quoted earlier from Reid's catalogue of the obstacles to reliable introspection, but he also notes that

in a delirium, or in madness, perception, memory, imagination and our reasoning powers, are strangely disordered and confounded. (Ibid.)

He concludes:

We must acknowledge it to be the lot of human nature, that all the human faculties are liable, by accidental causes, to be hurt and unfitted for their natural functions, either wholly or in part: But as this imperfection is common to them all, it gives no just ground for accounting any of them fallacious . . . It appears, I think, from what has been said, that there is no more reason to account our senses fallacious, than our reason, our memory, or any other faculty of judging which Nature hath given us. They are all limited and imperfect; but wisely suited to the present condition of man. We are liable to error and wrong judgment in the use of them all; but as little in the informations of sense as in the deductions of reason. (Reid [1785], II.22, pp. 251–252)

Thus the preference in the Cartesian model for introspection and reason over the senses is groundless.

Where does this leave us? One option would be to do the Cartesian one better:

If a Sceptic should build his skepticism upon this foundation, that all our reasoning and judging powers are fallacious in their nature . . . he must be left to enjoy his scepticism. (Reid [1785], VI.5, p. 480)

To such a sceptic I have nothing to say. (Reid [1764], V.7, p. 71)

But in response to the Cartesian himself, the ‘semi-sceptic,’ Reid presents his alternative picture:

Why, Sir, should I believe the faculty of reason more than that of perception; they came both out of the same shop, and were made by the same artist; and if he puts one piece of false ware into my hands, what should hinder him from putting another? (Reid [1764], VI.20, p. 169)

The faculties of consciousness,⁶¹ of memory, of external sense, and of reason, are all equally gifts of Nature. (Reid [1785], VI.4, p. 463)

In place of the restricted Cartesian perspective, Reid proposes to begin by trusting to the general reliability of all his faculties—‘that furniture which nature hath given to the human understanding . . . a part of our constitution . . . *the common sense of mankind*’ (Reid [1764], VII, p. 215). His disagreement with Hume over the Theory of Ideas is a mere footnote to this fundamental shift; indeed I’ve been suggesting that the Theory of Ideas, in the form of that imagined empirical theory, could arise in this setting with no skeptical consequences (though as far as we know, it would probably be false).

From this perspective, the debates over the Argument from Illusion, the Theory of Ideas, and the route from there to skepticism are all red herrings; what’s doing the real work is the underlying preference for introspection alone. Starting from this Cartesian point of departure, none of Reid’s discussions of the standard cases—perceptual relativity, illusion, hallucination—is admissible, depending as they do on our ordinary empirical ways of finding out about the world. On the other hand, from the Cartesian perspective, the route

61. Here I think ‘reflection’ would be more appropriate (see footnote 30 above). In any case, I assume that Reid’s ‘reflection’ on the operations of the mind (what I’ve been calling ‘introspection’) also belongs on this list.

to skepticism doesn't actually require the details of the standard arguments: if all we have to rely on is introspective evidence, then the bare possibility of any perceptual error is enough, because we have no grounds on which to determine that one perceptual belief rather than another is likely to be veridical. Reid's revolution undercuts the Cartesian presupposition by dragging it into the daylight and challenging its credentials. Once inner reflection, as Reid calls it, is accurately described and seen to be one fallible but still valuable faculty among several, his *Philosophy of Common Sense* stands as the logical alternative to the Cartesian model.⁶²

There's much more to be said on these topics, but let me stop here and try to draw a pair of morals from this line of thought. The first concerns the nature of the radical skeptical challenge to our knowledge of the external world. One way of posing this challenge is to begin with an appeal to some hypothesis like the Evil Demon or extraordinarily lifelike dreaming. From a second-philosophical perspective, such hypotheses are colorful ways of highlighting a particular methodological desideratum: I'm to defend my perceptual belief in that rose in my garden without appeal to any of my usual ways of defending my beliefs, that is, without appeal to any facts about the world. This 'from scratch' challenge is perfectly coherent—the Second Philosopher would welcome a satisfying response—but she herself has no idea how to go about producing one. Still, she doesn't see that her inability to produce evidence of this extraordinary kind in any way undercuts the force of the ordinary evidence she is able to provide.⁶³

62. To be fair, there are less naturalistic-sounding aspects of Reid's position, most prominently his apparent claim that the reliability of our faculties is a self-evident First Principle, not subject to or in need of proof. When Reid sounds these foundationalist, rationalistic themes, he's motivated sometimes by a familiar regress of justifications (e.g., Reid [1785], IV.4, p. 455) and sometimes by the desire to cut off the skeptic's argument that perception can't be effectively defended (e.g., Reid [1785], I.2, p. 41). The latter at least seems misguided: once the Cartesian starting point has been rejected, perception can be explored and evaluated in the ordinary ways that Reid and the Second Philosopher recommend. Indeed, Reid's declared foundationalism is at odds with his actual practice of investigating the mechanisms and shortcomings of vision (e.g., in his treatment of double vision, his account of how a visual figure registers objective relational information, his observations about the 'blind spot' created by the optic nerve, and two long chapters of the *Inquiry* on 'squinting' (where the two eyes are improperly aligned)). I'm tempted to think that Reid's position here is more a function of his discouragement over the limitations of then-current studies of perception, but this is a story for another day. For now, let me just say that I don't believe this side of Reid substantially undermines the second-philosophy-friendly portrait sketched here.

63. See [2007], I.2, [2011a]. [[This theme is developed at length in [2017a].]]

As we've noted, the skeptical argumentation of the naturalistic Hume at first appears to take a different form: beginning from our ordinary ways of finding out about the world, we're inexorably led to skepticism; our best empirical methods eventually undermine themselves. What we've now seen is that this isn't so: our ordinary empirical methods don't lead to the Theory of Ideas, because (as Reid shows) we can explain perceptual relativity, illusions, and hallucinations without appeal to sense-data; and even if we do embrace a representative theory of perception, we aren't thereby condemned to skepticism. What actually does the work in Hume's argument is the Cartesian starting point that rules out all but introspective evidence. Even from this starting point, it isn't clear that the Theory of Ideas follows, but in any case, it isn't needed: the simple fact of any perceptual error is enough to jeopardize my belief in the rose—whether that belief is immediate or inferred—unless I can show that I'm seeing it in conditions that don't predispose my senses to err, and I can't show that without appealing to various features of the perceptual situation and to my previous experiences in similar situations. In the end, it's the Cartesian restriction that's doing the work, and the work it's doing is essentially the same as expedients like the Evil Demon hypothesis: it's blocking access to my tried-and-true ways of finding out about the world. In this way, the apparent distinctiveness of Hume's argument evaporates.

Incidentally, notice that Reid's response to the skeptic bears some resemblance to the Second Philosopher's. Both recognize that the skeptic is challenging me to defend my perceptual belief in the rose without appealing to my usual methods and well-confirmed beliefs about the world, and both admit that this challenge is one they can't meet.⁶⁴ However, both also reject the notion that our inability to meet this challenge undercuts our ordinary evidence, that my belief in the rose is unjustified until the skeptic's challenge has been removed. Instead both hold that our ordinary evidence retains its force, despite our lack of extraordinary evidence.⁶⁵

64. It could be that Reid regards this as impossible in principle, while the Second Philosopher simply admits that she sees no way of doing it.

65. Of course, Reid adds an argument to the effect that the semi-skeptic's challenge is silly—the 'same shop' argument.

This comparison leads to the second moral: if you want to pinpoint the historical juncture when the Cartesian starting point was rejected, when the study of man took its proper place as a part of a larger study of the natural order, when something like contemporary naturalism took root, you'd do better to bypass Hume's *Science of Man* and focus on Reid's *Philosophy of Common Sense*. If the contours of my own second-philosophical version of naturalism have been clarified along the way, so much the better!⁶⁶

66. My thanks to the members of my 2009–2010 seminar on philosophies of common sense, to Sean Greenberg and Jim Weatherall, and to helpful audiences at UCLA, University of Toronto, University of Colorado at Boulder, University of Pennsylvania, Simon Fraser, the Naturalism conference at University of Cincinnati, and the Analytic Philosophy conference at University of Texas at Austin.

Moore's Hands

One of the most enigmatic and widely discussed of history's many attempts to respond to the radical skeptic is G. E. Moore's 'Proof of an external world' (Moore [1939]). Interpretations vary widely. At one extreme, Moore's name has been associated with a 'dogmatic' refutation of the skeptical conclusion.¹ At the other, commentators contend that despite how the paper has been read by philosophers from Wittgenstein² to Stroud³ and beyond, it's in fact an argument against idealism, not skepticism.⁴ The former can hardly be the actual Moore, who explicitly denies that he can prove he's not dreaming (ibid., p. 149).⁵ As for the latter, there may well be something to the idea that the Moore is concerned with idealism, but surely anyone who can prove the existence of an external world is epistemically better off than the skeptic imagines.⁶ Somewhere between these extremes lie readers who take Moore to be addressing the skeptic, but not fully appreciating how radical the challenge is.⁷ In fact, I think that Moore was quite well acquainted with his skeptical adversary and that he presents a cogent anti-skeptical response (even if he's also concerned with idealism). These claims will be my focus here, with a meta-philosophical glance now and then at Moore's method.

1. See, e.g., Pryor [2000], Wright [2002].

2. See Wittgenstein [1949–1951].

3. See Stroud [1984].

4. See, e.g., Baldwin [1990]; Coliva [2010], [2018]; my [2018].

5. These writers admit as much. See, e.g., Pryor [2000], p. 541: 'What follows is not a serious attempt at Moore exegesis. I'm just using Moore to introduce some ideas I want to develop.'

6. Cf. Stroud [1984], pp. 136–137.

7. E.g., see part one of Wittgenstein [1951] or chapter II of Stroud [1984].

1. SETTING UP THE PROOF

As is well-known, Moore spends the bulk of ‘Proof’ on a lengthy, occasionally baffling, preamble that begins with the quotation from Kant:

It still remains a scandal to philosophy . . . that the existence of things outside of us . . . must be accepted merely on *faith*, and that, if anyone thinks good to doubt their existence, we are unable to counter his doubts by any satisfactory proof. (Moore [1939], p. 127)⁸

Moore agrees that this is scandalous, then turns his attention to a detailed analysis of the notion of ‘a thing outside of us’ that will occupy him until the proof itself finally appears in the concluding few pages. He proposes that another Kantian notion, ‘thing to be met with in space,’⁹ provides a fairly good gloss of what’s intended:

my body, the bodies of other men, the bodies of animals, plants of all sorts, stones, mountains, the sun, the moon, stars, and planets, houses and other buildings, manufactured articles of all sorts—chairs, tables, pieces of paper, etc. (Ibid., p. 130)

He avoids the philosopher’s term ‘material object’, despite the fact that all his examples qualify, because there are things ‘to be met with in space’ that don’t fit, like shadows.

Pushing on, he notes that Kant tends to equate ‘thing to be met with in space’ with ‘thing presented in space’. At this point, Moore makes a small foray into empirical science, going so far as to quote and cite a psychology

8. This is the Kemp Smith translation. The more recent version of Guyer and Wood has it this way: ‘it always remains a scandal of philosophy . . . that the existence of things outside us . . . should have to be assumed merely **on faith**, and that if it occurs to anyone to doubt it, we should be unable to answer him with a satisfactory proof’ (B xxxix). For a brief discussion of Kant’s ‘proof’, see [2011a], pp. 126–134.

9. A373. Guyer and Wood have ‘things **that are to be encountered in space**’. In terms of the Kantian distinction between ‘empirical object’ and ‘thing in itself’, this is the former. Moore dismisses the latter out of hand (see Moore [1939], p. 139), one reason I tend to doubt that idealism is his main target (see [2018], p. 232).

manual and a physiology text on the subject of 'negative after-images,' and even to conduct a few experiments of his own:

I took the trouble to cut out of a piece of white paper a four-pointed star, to place it on a black ground, to 'look steadfastly' at it, and then to turn my eyes to a white sheet of paper: and I did find that I saw a grey patch for some little time [so far the textbook description]—I not only saw a grey patch, but I saw it *on* the white ground, and also this grey patch was of roughly the same shape as the white four-pointed star at which I had 'looked steadfastly' just before—it was also a four-pointed star. I repeated this simple experiment successfully several times. (Moore [1939], p. 131)

While most of what goes on in this long preamble to the proof would presumably qualify as 'conceptual analysis,' just what we would expect from Moore, it's striking to find him here engaged in a bit of active experimentation. This is the only instance I know of in Moore's opus, but empirical investigation occurs throughout the work of Thomas Reid, Moore's predecessor in the 'philosophy of common sense.'¹⁰ In any case, this after-image, Moore concludes, is clearly 'presented in space,' but isn't 'to be met with in space,' because the latter term implies 'that *anyone*, who had been in the room at the time, and who had normal eyesight . . . might have seen . . . it' (ibid., p. 132). Conversely, a thing could be 'to be met with in space' without ever actually being 'presented' to anyone, so Moore leaves 'presented in space' behind and sticks with 'to be met with in space.'

Despite all this, Moore doesn't regard 'to be met with in space' as fully determinate—what about the sky or mirror images?—but he figures there are examples uncontroversial enough—stones, planets, chairs—that showing things of that sort to exist will be enough to show that 'things to be met with in space' exist. He then embarks on an equally detailed, if perhaps a bit trickier discussion of the relations between 'things to be met with in space' and 'things outside of us,' which he takes to mean 'things external to our

10. See, e.g., Reid [1764]. Commentators speculate that Moore was influenced by Reid during his stay in Edinburgh from 1904 to 1908, but the evidence is circumstantial. (Cf., Essay #3, footnote 31 and surrounding text.)

minds'. In the end, to say that something is 'external to our minds' is to say that its existence doesn't imply that any of us is having experiences. He then argues that if such-and-such is 'to be met with in space', it must also be 'external to our minds'. So, finally, in order to prove that there are 'things outside of us', as Kant thought we should, it will be enough to prove that there are things to be met with in space.

2. THE PROOF

By means of this prolonged run of conceptual analysis, with a naturalistic dash of empirical science, Moore has arrived at the climactic moment:

If I can prove that there exist now both a sheet of paper and a human hand, I shall have proved that there are 'things outside of us'; if I can prove that there exist now both a shoe and sock, I shall have proved there are now 'things outside us'; etc.; and similarly I shall have proved it, if I can prove that there exist now two sheets of paper, or two human hands, or two shoes, or two socks, etc. Obviously, then, there are thousands of different things such that, if, at any time, I can prove any one of them, I shall have proved the existence of things outside us. Cannot I prove any of these things? (Moore [1939], p. 145)

And, famously, yes he can:

I can prove now, for instance, that two human hands exist. How? By holding up my two hands, and saying, as I make a certain gesture with the right hand, 'Here is one hand', and adding, as I make a certain gesture with the left, 'and here is another'. (Ibid., pp. 145–146)

Even years of familiarity hardly dim the blunt audacity of this passage. What does Moore think he's doing?

Much of what's left of the paper is taken up with Moore's argument that the proof is 'rigorous', at a tempo reverting to the slow and deliberate unfolding of the preamble. Finally, in the concluding three paragraphs, he

admits that he's 'perfectly well aware that, in spite of all that I have said, many philosophers will . . . feel that I have not given any satisfactory proof' of the existence of external things (Moore [1939], p. 148), and he tries to figure out what could be bothering them. What they want, of course, is a proof of the premises of the argument, and in fact, not even quite this:

What they really want is not merely a proof of these two propositions, but something like a general statement as to how *any* propositions of this sort may be proved. (Moore [1939], p. 149)

What the skeptic wants, Moore suggests, is an explanation of how we could prove (and presumably thereby come to know) any proposition at all that's 'of the same sort as' his premise, 'here's a hand'—in other words, any proposition at all about the external world.

Here Moore demonstrates that he fully understands the skeptic's challenge. All his beliefs about the world are being called into question at once, and he's being asked to defend his claim that 'here's a hand' from that standpoint. Any of the ordinary evidence he would be inclined to appeal to—that he clearly sees his hand directly in front of him, that he can manipulate it at will, that no unusual conditions appear to be present¹¹—all this is being ruled inadmissible. The skeptic is demanding not ordinary, but extraordinary evidence, a kind of evidence whose trustworthiness doesn't depend on any information Moore may think he has about the world. He's being asked to defend his premises from scratch, so to speak.

So, with the radical skeptical challenge clearly in view, what does Moore offer us? Though the line of thought gets a bit murky in the final two paragraphs of 'Proof', two points stand out. First, he freely admits that he has no such extraordinary evidence to offer:

This, of course, I haven't given; and I do not believe it can be given: if this is what is meant by proof of the existence of external things, I do

11. And if the circumstances were unusual, they could be checked: 'If one of you suspected that one of my hands was artificial he might be said to get a proof of my proposition 'Here's one hand, and here's another', by coming up and examining the suspected hand close up, perhaps touching and pressing it, and so establishing that it really was a human hand' (Moore [1939], p. 149).

not believe that any proof of the existence of external things is possible.
(Moore [1939], p. 149)

Why not? Because

in order to do it, I should need to prove for one thing, as Descartes pointed out, that I am not now dreaming. (Ibid.)

So far, this seems familiar. Cartesian dreaming is commonly understood as functionally equivalent to the Evil Demon hypothesis, and to refute that, one would need evidence that survives when all our ordinary evidence is ruled out; in other words, one would need extraordinary evidence. This, reasonably, Moore doesn't claim to possess.

But then Moore says something unusual:

I have, no doubt, conclusive reasons for asserting that I am not now dreaming; I have conclusive evidence that I am awake: but that is a very different thing from being able to prove it. (Moore [1939], p. 149)

If he has evidence, indeed conclusive evidence, that he isn't dreaming, then he can't be speaking of Cartesian dreaming or the kind of extraordinary evidence that would be required to refute it. What he must have in mind is ordinary dreaming, the everyday dreaming we're all familiar with, and the ordinary evidence we all have when we're awake: my experience is continuous and coherent, I'm aware of the past and can plan for the future, I can read and follow a train of thought, and so on.¹² But evidence of this sort is futile in the face of Cartesian dreaming or the from-scratch challenge.¹³ In other words,

12. Those familiar with lucid dreaming will add that I can read a clock and use my smartphone, I can turn lights on and off, I can't put my hand through a solid object, or take great leaps as in low-gravity conditions.

13. Appealing to both Cartesian dreaming and the from-scratch challenge is redundant. If Cartesian dreaming is what we're out to refute, then we need the very sort of extraordinary evidence that the from-scratch challenge demands, so Cartesian dreaming is enough by itself. Similarly, in the face of the from-scratch challenge, not even ordinary dreaming can be refuted, so the from-scratch challenge is also enough by itself. (Cf. [2017a], pp. 70–71, footnote 49.)

as Moore says, having such evidence 'is a very different thing from being able to prove it'.

The second point that stands out in this final passage is that he doesn't think the cogency of his proof of an external world requires him to give this kind of proof, to provide him extraordinary evidence, for his premises:

some people . . . think that, if I cannot give such extra proofs [of his premises], then the proofs that I have given [of the existence of external things] are not conclusive proofs at all. And this, I think, is a definite mistake. (Moore [1939], pp. 149–150)

These people, apparently including Kant, think

that if I cannot prove that there is a hand here, I must accept it merely as a matter of faith—I cannot know it. (Ibid., p. 150)

Moore thinks this view can 'be shown to be wrong', but the catch is

though shown only by the use of premises which are not known to be true, unless we do know of the existence of external things. (Ibid.)

I can know there's a hand here—I needn't accept it on faith—on the basis of ordinary evidence, evidence that involves external things, notwithstanding the fact that I have no extraordinary evidence. In other words, Moore holds that my lack of extraordinary evidence doesn't undercut the epistemic force of the ordinary evidence that I do have.¹⁴

This reading helps explain why Moore begins his paper with that long, somewhat tedious recitation of commonplaces: his lists of everyday objects, his straightforward discussions of what's to be met with in space, even his little experiments with after-images. The point of all this is to ease us into his argument by calling on our familiar ways of thinking, drawing on our habitual trust in ordinary evidence, accustoming us to the everyday, despite the

14. Readers of [2017a] will recognize this as the position of the Plain Inquirer (a.k.a. the Second Philosopher).

philosophical context. Once we've settled into this commonsense approach, he gestures with his hands and rests his case. In effect, he's turned the tables on the skeptic: the dialectic usually begins with the skeptic drawing his radical conclusion and challenging the anti-skeptic to reply; instead Moore establishes his mundane outlook and challenges the skeptic to unseat him. The surprise is that it turns out to be harder than we might expect. Why should our failure to achieve extraordinary evidence undermine our ordinary evidence?

3. THE LECTURES

I can't pretend that the enigmatic text of 'Proof' compels this reading, but I think there's strong evidence for it in a remarkable set of Moore's lecture notes from the academic year 1933–1934. These lectures were delivered just a few years before the appearance of 'Proof of an external world' (1939), on the topic, 'what is one trying to do if one tries to do philosophy' (Moore [1933–1934], p. 153)—clearly signaling a focus on meta-philosophy.

He begins, unsurprisingly, with a discussion of conceptual analysis as one of the things philosophers try to do,¹⁵ but in the third lecture he turns to questions of another kind:

Questions about reality as a whole, to which one particular answer is 'assumed' or 'presupposed' (in a sense I have yet to define) by *us all* both in ordinary life and in the special sciences. (Moore [1933–1934], p. 174)¹⁶

As an example, he offers the question of the truth or falsity of this claim:

Whatever else may be true about the Universe, one thing that's true of it is that there have existed in it in the past very large numbers of material things [or] 'bodies' in . . . the sense in which the moon is a body, my body is a body, this desk and blackboard are bodies. (Ibid., p. 175)

15. He also considers other types of analysis, like Russell's theory of descriptions.

16. I take the liberty of spelling out the abbreviations Moore uses in these notes, replacing '&' with 'and' and so on.

In what sense do we 'presuppose' an answer in ordinary life?

That there have been many bodies is 'presupposed' by us all in ordinary life, in the sense that we have all, in ordinary life, often *observed* facts from which, taken together, it strictly *follows* that there have been many bodies. For instance, I at this moment *observe* that there are a good many human bodies in this room, a good many desks, a good many articles of clothing, a blackboard. And I have very often in the past made similar observations. I've observed in walking down a street that there were a good many houses in it; in looking at a wood that there were a good many trees in it; in looking at [a] sea-beach that there were a good many pebbles on it. It seems not too much to say that *all* adult human beings have frequently observed facts of this sort: so that in the case of each of us, it follows from many facts that he has observed, taken together, that there have been large numbers of bodies in the Universe, whatever else may be true of it. (Ibid., p. 176)

Here Moore adopts our ordinary standpoint and argues that there are many material things. Obviously, this foreshadows 'Proof', but there is more.

The sense in which science 'presupposes' an answer to the question of the existence of bodies is slightly different:

They're entailed by *established conclusions* of the sciences. One science, for example, has established that thousands of years ago there existed large numbers of enormous reptiles, ichthyosauri and such like. This is a genuine conclusion of a science and an established one; and from it there follows that there have been many bodies in the Universe. Another science has established that more than thousands of years ago just as now there were millions of stars. This is a genuine conclusion of astronomy; and from it there follows that there have been many bodies in the Universe. (Moore [1933–1934], p. 177)

Here Moore goes beyond his elementary experiments with after-images and appeals directly to natural science. For now, let's just note that this is neither commonsense philosophizing nor conceptual analysis.

What puzzles Moore is how any *philosophical* posing or answering of the question—are there bodies?—would relate to these scientific claims. On the one hand, the claims imply an answer; on the other hand, no scientist is likely to put the question in so stark a form. This slender reed brings Moore no solace:

[I]t's true that no special sciences raise these abstract questions themselves; but if the sciences do raise and settle questions from which one particular answer to them *follows*, it seems to me obviously rather a subterfuge to say that in raising them [the philosopher] is raising questions that scientists don't raise. (Moore [1933–1934], p. 178)

If the philosopher denies the existence of bodies, 'his answer will *conflict* with, be inconsistent with, the results of the sciences' (ibid.). On the other hand, if the philosopher aims to prove the existence of bodies, 'what's the use of his going over the ground again, and proving again what the scientists have already proved?' (ibid.). It isn't even clear that the question falls in the philosopher's purview:

It's admitted that it's not the business of [the philosopher] to discuss whether millions of stars existed a million years ago: that must be left to astronomy. But if so, how can it be his business to discuss whether it's true that many material things have existed in the Universe or none have? If he comes to the conclusion that they have, he will be merely saying something which the astronomers have already proved; if he says none have, he will be contradicting the astronomers. (Ibid.)

What room can we find for the philosopher even to address the question of the existence of the external world?

One possible opening is the thought that the philosopher will approach the question with a different method of the scientists, but

then the question arises whether the method used by the sciences isn't the best and the only proper method for settling such questions. (Moore [1933–1934], p. 178)

Here our earlier meta-philosophical hint has become an explicit question: should we join the naturalist in taking science to be our best way of finding out about the world?¹⁷

Moore returns to this question again and again in the course of the lectures. In connection with the philosophy of mathematics, he remarks that Russell's discussions of mathematical truths raise

the same sort of puzzling question as to the relation of philosophy to mathematics, which I said was raised by the fact that philosophers do raise questions to which the sciences imply an answer. Surely it's the business of the mathematicians to decide whether particular mathematical propositions are true? And if so what's the use of the philosopher discussing whether *any* mathematical propositions are true? Suppose he decides they are, can he give better reasons than the mathematicians give? Suppose he decides they aren't. He's contradicting the mathematicians. And aren't they better judges? (Moore [1933–1934], p. 185)

Another important case for our purposes is the nature of evidence:

The sciences *do* say not only *p*, . . . but there's *good evidence for p*: and it has happened that *p* belongs to a class of propositions with regard to which philosophers have concluded: We *never* have good evidence for a proposition of that sort. Isn't the fact that the scientists say: Such-and-such *is* good evidence for so-and-so, a reason for saying: It *is* good evidence? (Ibid., p. 189)

Again and again, Moore asks the same questions. They sound rhetorical, he seems to edge toward naturalism, but he doesn't make this explicit in any of these passages.

In the concluding pages of the lectures, Moore finally asks what methods philosophers *have* used to argue for one side or the other on the existence of bodies—and also what methods they *ought* to use. One example of a method

17. I have in mind here what's often termed 'methodological naturalism', the idea that science is the best way to find out about the world. Given the impossibility of specifying once and for all what counts as 'science', I prefer the standpoint of the Plain Inquirer in [2017a] and the Second Philosopher in [2007], but Moore isn't concerned with such niceties.

philosophers have used to argue that there are no bodies is the attempt to show that ‘here is a hand’ is self-contradictory. This would do the job, so the argument ought to be examined, but Moore doesn’t think it survives that examination. On the methods philosophers have used to support the existence of bodies, Moore cites Descartes, but again he thinks the argument doesn’t hold up to scrutiny. In the end, at last, he returns to the theme we’ve been tracing:

I’ve insisted the sciences give indirect proofs of [the existence of bodies], by proving that particular material things have existed: I instanced the proof of the existence of Plesiosauri, Ichthyosauri, etc.; and of the existence of millions of stars 10,000 years ago. (Moore [1933–1934], p. 194)

Moore’s intent is clear: he regards the deliverances of science as relevant to the question of the existence of the external world and indeed as proving its existence.

It seems reasonable, then, to see ‘Proof of an external world’, written only a few years later, as continuing in this same general direction. Rather than declare his confidence in science directly, Moore subtly eases the reader into the intended point of view, first, that of common sense, and then—with his gestures toward scientific results and experimentation—that of the naturalist. Many readers take Moore to regard common sense as sacrosanct, but the tone of the lectures suggests that when it comes into conflict with science, it’s common sense that gets corrected: isn’t science ‘the best and only proper method for settling such questions?’ (Moore [1933–1934], p. 178).¹⁸ For purposes of ‘Proof’, though, he’s come to realize that he doesn’t need esoteric examples of dinosaurs and stars—his hands will do just as well.

A direct link between the lectures and the essay surfaces in the single caveat Moore raises. In ‘Proof’, as we’ve seen, he acknowledges that his argument relies on ‘premisses which are not known to be true, unless we do know

18. This view doesn’t seem special to the 1933/1934 lectures. See, e.g., Moore [1910–1911], pp. 12–13: ‘[T]he sphere of [some] of the special sciences may be defined as being to give us detailed knowledge . . . about material objects . . . In the case of all these sciences, there are, we believe, an immense number of things which are now definitely known to be facts; a great many, which were formerly believed, but are now definitely known to be errors.’ In this early lecture, Moore includes these conclusions of science under the heading of Common Sense. (I’m indebted to Lycan [2007], p. 97, footnote 35, for this reference.)

of the existence of external things' (Moore [1939], p. 150). In the lectures, this same sentiment appears immediately after the passage quoted above about the 'indirect proofs' provided by science:

All scientific proofs of the existence of particular kinds of material things seem to me to assume as premisses the existence of others: the proof of ichthyosauri assumes the existence of fossils; that of stars 10,000 years ago the existence of stars *now*. (Moore [1933–1934], p. 194)

In other words, in both cases, Moore denies having met the from-scratch challenge. What's new in 'Proof'—though perhaps implicit in the lectures—is his explicit denial that a from-scratch proof is required.

So I rest with the reading of Moore's argument in §2, above. He understands the skeptic's from-scratch challenge full well; he takes it as the demand that he prove 'here's a hand' without appeal to any of his other beliefs about the world. This effectively neutralizes all the ordinary evidence he thinks he has for the existence of his hand—that he can see it, feel it, clench and unclench it, and so on—and requires him to produce some new brand of extraordinary evidence, evidence that somehow works all by itself. This he can't do. He can't answer the from-scratch challenge, but he doesn't see that his lack of extraordinary evidence undercuts his ordinary evidence. He inhabits his naturalistic standpoint and stands by his visual, tactile, and proprioceptive experiences, supported, if need be, by further elaboration: given his current surroundings, given how those sensory modalities work, there is every reason to expect that they are reliable in this case.

So, 'here's a hand' and 'here's another'.

Wittgenstein on Hinges

Our topic today is hinge epistemology.¹ I feel like something of an interloper here for a number of reasons, among them that I'm not a hinge epistemologist or a participant in the contemporary debate on its structure and merits. But I do have some thoughts on the role of hinge propositions in Wittgenstein's discussion—in particular, I don't think he was a hinge epistemologist, either. Given his standing as founding father of the view, this reading is no doubt somewhat heretical, but I hope it might still be of some interest. One preliminary note, though: I honestly don't believe there's only one correct reading, or only one interesting or fruitful reading of the Wittgensteinian opus early, middle, or late,² so it would be more accurate to say that I'm presenting what I take to be an illuminating option for your consideration.

Our ultimate goal, of course, is *On Certainty*, but for reasons we'll get into in a moment, it's even more treacherous to interpret than the *Tractatus* or the *Investigations*. My thought, then, is to get a running start by drawing a trajectory through those early and middle works, then to try to extend this line of development into his final writings on skepticism.

The story begins, then, with the *Tractatus*.³ Of course there are many schools of thought on this book, each with many variations⁴—from the logical atomism of Russell and the logical positivism of the Vienna Circle to the

1. This essay was delivered as a talk at the UCI Hinge Epistemology Conference in September 2017. I'm grateful to Annalisa Coliva and Duncan Pritchard for the invitation.

2. For the record, I'm using 'early' for the *Tractatus* and surrounding manuscripts, 'transitional' for the *Philosophical Remarks*, *Philosophical Grammar*, *Big Typescript*, etc., 'middle' for the *Investigations* and its ilk, and 'late' or 'final' for *On Certainty* and other writings of 1949–1951.

3. Wittgenstein [1921]. For more on the reading sketched here, see [2014b], chapters 3 and 4.

4. See Stern [2003] for a helpful taxonomy.

new, therapeutic reading of Diamond and Conant. Whatever each of these may have to offer, here I propose a straightforward so-called ‘metaphysical’ reading familiar from the likes of Anscombe, Stenius, and Pears: Wittgenstein is offering an account of logic and representation. To my mind, this is among the most fascinating accounts of these notions ever proposed.

Wittgenstein’s project can be seen as beginning on analogy with Kant’s: Kant asks, ‘what must the world be like that we can cognize it as we do?’; Wittgenstein asks, ‘what must the world be like that we can represent it as we do?’ What makes Kant’s project so difficult is that he thinks our cognition of some aspects of the world—for example, its geometry—is a priori. Analogously, Wittgenstein assumes that the sense of a representation must be independent of any contingent facts about the world.⁵ This is a natural idea: we have to know what a claim says, what it would take for it to be true, before we look to the world to determine whether or not it is true. Armed with this simple idea—let me call it the ‘priority of sense’—Wittgenstein produces many of the book’s wonders, beginning with its confounding ontology.

The argument is familiar to students of the *Tractatus*. If p is a statement apparently about a complex object, C , does p have sense if C doesn’t exist? There seem to be two options. If the answer is no, if the non-existence of C implies that there is no way the world would be if p were true, then whether or not p has sense would depend on the contingent facts, in particular, on whether or not C happens to exist. That can’t be right by the priority of sense—sense is independent of contingent facts!—so the answer must be yes, p has sense even if C doesn’t exist. How can this be? Well, part of what p claims must be that the constituents of C are so arranged as to constitute C . But then, on pain of infinite regress, this sort of analysis must eventually bottom out in some fundamental, non-composite, non-contingent ur-constituents. So, in order for us to represent the world as we do, our language must include simple names of necessarily existing simple objects.

But naming isn’t all there is to representing. Faced next with the challenge of explaining how Othello can represent Desdemona loving Cassio when in fact she doesn’t, he comes up with the picture theory: the parts of

5. In fact, neither philosopher distinguishes clearly between a priority and necessity, so the analogy is actually stronger than it may sound to our post-Kripkean ears.

the picture stand for parts of the world; the picture represents those parts of the world as related to one another just as the corresponding parts of the picture are related to one another. In this way, an elementary proposition—a concatenation of names—can represent a state of the world even if that state doesn't actually obtain. Again, the sense of the proposition—what the world would be like if it were true—is prior to any facts about how the world actually is; the proposition represents a possible way the world could be.

When Wittgenstein extends this account of representation to all propositions—that is, to all logical combinations of elementary propositions—he wants to preserve the idea that every such proposition describes a possible situation. This requires that any conjunction of elementary propositions is possible, in other words, that the elementary propositions are logically independent and the possible states of the world correspond to rows in a big truth table (except for all T's and all F's). At this point, the various strong theses of the *Tractarian* view fall into place. But why should we think that the elementary propositions are logically independent?

This is a difficult question to answer, but let me just observe that here, too, the priority of sense is implicated. To see how, recall the familiar apparent anomaly from 6.3751: 'the statement that a point in the visual field has two different colours at the same time is a contradiction.' This means that 'red here now' isn't elementary, and by the same reasoning, neither are 'this is three centimeters long' or 'this particle is moving with velocity v ' or anything else we might look to as an example of an elementary proposition. The *Tractarian* response, of course, is to agree that these aren't elementary, to conclude that 'this point in the visual field is red' is in fact a complex proposition, which—when analyzed down to a logical combination of truly elementary propositions about truly simple objects—will be revealed to be contradictory, senseless. My point is that this recovery makes essential use of simples; if we stick to ordinary, familiar objects, there would be no hope for logical independence. So the priority of sense is playing a role here, too—indirectly, as guarantor of the non-standard ontology that saves logical independence from immediate refutation.

Now, as is well-known, the color exclusion problem was one of several puzzles that soon led Wittgenstein to re-think the *Tractarian* position on representation. Various ideas flit through the transitional period, until a new and

very different vista opens up in the *Investigations*. By then, Wittgenstein has rejected the ‘illusion’ that ‘proposition’, ‘sense’, and ‘truth’ are ‘*super*-concepts’, realizing that they are just words with ‘a use . . . as humble . . . as . . . the words “table,” “lamp,” “door”’ (*PI* §97):

The sense in which philosophy . . . speaks of sentences and words is no different from that in which we speak of them in ordinary life when we say, for example, ‘What is written here is a Chinese sentence’ . . . We’re talking about the spatial and temporal phenomenon of language, not about some non-spatial, atemporal non-entity. (*PI*, after §108)

Our whole approach to meaning

must be turned around, but on the pivot of our real need. (*PI* §108)

What we really need is an understanding of how we ordinary humans actually manage to represent the world.

As we all know, this middle-period Wittgenstein proposes that in many cases, ‘the meaning . . . is its use in language’ (*PI* §43), so the pressing question becomes: how can I grasp the meaning, if it encompasses the entire use? The problem is starkly posed with the case of the recalcitrant student who’s taught the meaning of ‘plus-2’ with apparent success, but who goes on to write ‘1004’ after ‘1000’. The numerical example might suggest that the challenge arises only at the relatively sophisticated level of grade-school arithmetic, but of course it can be posed just as easily for new applications of even the very first words a child acquires, like ‘cat’, ‘dog’, ‘house’, ‘tree’, ‘red’, ‘square’, and so on—words based on pre-linguistic classificatory abilities, many of which we share with other animals. So, how do we manage to mean ‘plus-2’ so that ‘1002’ comes after ‘1000’ and ‘1004’ doesn’t? More simply, how do we manage to represent cats rather than cats-plus-the-Eiffel-Tower?

Here again, what Wittgenstein offers in response has been interpreted in a variety of ways, but I think it’s safe to say that many of these count as therapeutic readings. The line I favor is among these.⁶ In the welter of cross-talk

6. See [2014b], pp. 63–73, 95–98, [2017a], pp. 181–187, drawing on Stern [2004].

that makes up the rule-following discussion there are at least three voices. One is the rule-following skeptic who ultimately claims that

no course of action could be determined by a rule, because every course of action can be brought into accord with the rule. (*PI* §201)

Then there is the voice often called the ‘interlocutor’ or (by David Stern, following Cavell) the ‘voice of temptation’ that attempts to block the skeptic’s conclusion by appealing to a litany of potential determiners—an intuition, a feeling, an image, a paradigm case, the instructor’s intentions, a reformulation of the rule, and so on—none of which does the job. Still another is the ‘voice of correctness’ (again Stern) or sometimes ‘Kripkenstein’ (from Kripke [1982]; see also Fogelin [1976], [1987]) who accepts the skeptical conclusion that nothing pre-determines the correct or incorrect applications, but charts an account of how the community’s shared inclinations and practices support our assertion that ‘1002’, not ‘1004’, follows ‘1000’ in our debate with the recalcitrant student.

So, where is the therapeutic Wittgenstein in all this? Each of these three voices stakes out a controversial philosophical position: the meaning skeptic presents the ‘paradox’ that there is no such thing as following a rule; the voice of temptation attempts a so-called straight solution; while the voice of correctness offers a so-called skeptical solution. But Wittgenstein tells us that when philosophy is done properly

[i]f someone were to advance *theses* . . . it would never be possible to debate them, because everyone would agree to them. (*PI* §128)

So, apparently none of these voices represents Wittgenstein himself. If we’re to do philosophy properly, then, what are we to say about the case of the recalcitrant student?

Wittgenstein counsels us to ‘stick to matters of everyday thought’ (*PI* §106), to ‘speak the language of everyday’ (*PI* §120):

The aspects of things that are most important for us are hidden because of their simplicity and familiarity. (One is unable to notice something—because it is always before one’s eyes.) (*PI* §129)

If we return to the case in this spirit, we realize that we aren’t actually at a loss. When the student puts ‘1004’ after ‘1000’, we see that we were too hasty in thinking he understood; we give him more examples, more training, more testing. If he continues stubbornly in his error—resisting all our efforts to help him understand—we conclude that he’s in for a difficult life. What we emphatically do not do is doubt that ‘1002’ is the right answer. As long as this poor fellow is an outlier, we retain our confidence in our general ability to teach and learn, our general ability to use our language with meaning. Our actual methods remain well suited to what Wittgenstein has called our ‘real need’.

What’s gone wrong, then, with the various other voices is that they somehow don’t feel the force of these perfectly ordinary considerations. Somehow they

get on the wrong track where it seems [they] have to describe extreme subtleties, which [they] are unable to describe with the means at [their] disposal. [They] feel as if [they] had to repair a torn spider web with [their] fingers. (*PI* §106)

Thus begins the therapeutic diagnosis: they’ve lost sight of what actually happens in everyday life. The only hope of a cure is to figure out what’s causing this persistent distortion. ‘If I am convinced that a mouse cannot come into being . . . by spontaneous generation out of grey rags and dust’, then examining ‘those rags very closely . . . will perhaps be superfluous’:

what it is in philosophy that resists such an examination of details, we have yet to come to understand. (*PI* §52)

What blocks our attention to the ordinary facts in this case?

The recalcitrant student is portrayed as arising in our midst, but in fact he’s unlike anything we typically face in our classrooms or tutoring sessions. Presumably he hasn’t displayed any of the usual signs of cognitive deficit,

and known cognitive deficits don't tend to be so strictly localized. Perhaps we rule out that he's playing a joke or acting on a dare or some other sort of put-on. If we regularly came upon inexplicable cases like this, it would be a problem, but we don't—and that's the key. Our troubled philosophers are being led to question the effectiveness of our ordinary methods by bringing them up against a kind of case that's not found in practice. That we couldn't force this wayward student to get the hang of what we're trying to teach him doesn't change the fact that our ways of training actual students tend to work just fine. Meditating on his case blinds us to what 'would in fact, *does* in fact, *do* to explain to someone how to go on', as Cora Diamond puts it. We come to think we need to convince 'someone on whose uptake, on whose responses, we are not at all depending' (Diamond [1986], pp. 68–69). We think we need an explanation of our meaning that can't possibly be misunderstood, so that no sincere student could be as recalcitrant as the skeptic imagines.

But in fact we *don't* need this, and recognizing that we don't is to turn our inquiry around 'on the pivot of our real need' (*PI*, §108). Getting the voices of temptation and correctness to see this would effect a cure, would 'show the fly the way out of the fly-bottle' (*PI* §309). I think it has to be acknowledged that this therapy seldom works in practice—the rule-following paradox remains a topic of much philosophical debate—but if it did, the outcome would be as Wittgenstein hopes: 'the one that gives philosophy peace' (*PI* §133).

Before moving on to *On Certainty*, let me make one last observation that might indirectly support this reading of the rule-following argument: the distorting presupposition it targets—that we need an account of meaning that works no matter how bizarre the circumstance—is none other than the priority of sense, so central to the *Tractatus*—the assumption that our statements must have the meanings they do quite independently of any contingent facts. Obviously this is a tendency of thought to which Wittgenstein himself was deeply susceptible, a tendency at the very heart of his early philosophy, so perhaps it wouldn't be surprising if exorcising that tendency would be equally central to his next phase, to the middle period.

OK, so we now have a fairly straightforward therapeutic reading of middle-period Wittgenstein on meaning and representation. My plan, as suggested earlier, is to use this as a template for a reading of the treatment

of skepticism in *On Certainty*, but first I must acknowledge that many acute observers, including most prominently our co-symposiast Professor Moyal-Sharrock,⁷ have argued in detail that the final Wittgenstein of 1949–1951 was no longer a therapeutic philosopher, that he reverted to proposing and defending philosophical theses. As it happens, a former student here, Brian Rogers, wrote a provocative dissertation a few years back, arguing that there was no such reversion, that Wittgenstein continued to pursue his therapeutic method in *On Certainty*.⁸

Evaluation of these opposing claims is difficult, given that the book wasn't prepared for publication by Wittgenstein himself, but selected from a series of notebooks by his literary executors, Anscombe and von Wright. They explicitly note that 'it is all first-draft material, which [Wittgenstein] did not live to excerpt and polish', but they nevertheless maintain that 'it constitutes a single sustained treatment of the topic' (OC, preface). In a remarkable philological effort, Rogers challenges this characterization. Drawing on evidence as varied as the colors of ink, the neatness (or not) of the handwriting, vendors' stamps on the notebooks, and of course contemporaneous letters and reports, he argues persuasively that only §§1–65 (which I'll call part 1) and §§300–676 (part 3) are 'sustained' treatments. Part 1 comes from late 1949—either during or shortly after Wittgenstein's discussions in Ithaca with Malcolm about Moore, when he was ill but not yet diagnosed with prostate cancer—and part 3 from the last six weeks of his life in early 1951, after the hormone and X-ray treatments were abandoned and he felt better for a short time before his death. The middle portion of the book as we know it today, §§66–299 (part 2) was selected and organized by the editors from several notebooks mixed in with material on color and philosophy of psychology as well as skepticism, material sometimes difficult even to order and often difficult to date.

Rogers also establishes that Wittgenstein was consistently unhappy with the work in parts 1 and 2, though more so for the latter than the former. In February 1951, just before those last six weeks, he wrote to Malcolm that 'my

7. See Moyal-Sharrock [2004].

8. Rogers [2011]. See also Rogers [2010].

head [is] empty' although 'it was half empty already when I was in Ithaca.' Two months later, in April 1951, he wrote, again to Malcolm:

An extraordinary thing has happened to me. About a month ago I suddenly found myself in the right frame of mind for doing philosophy. I had been absolutely certain that I'd never again be able to do it. It's the first time after more than 2 years that the curtain in my brain has gone up.⁹

This stands to reason, as Rogers reports, because Wittgenstein was increasingly weak and ill in much of 1949 and mental confusion is a well-known side effect of the hormone treatments he received from November 1949 until March 1951.

All this means that *On Certainty* can't be approached with the confidence appropriate to Wittgenstein's finished works. A great deal is now known about the many stages of selection, revision, and re-ordering that went into the composition of the *Investigations*. If Wittgenstein's final thoughts on skepticism were ultimately intended as therapeutic in some similar sense, then extrapolating from first-draft remarks would involve a considerable measure of speculation. Wittgenstein's disapproval of parts 1 and 2 and approval of part 3 adds an additional hurdle for any reading:

No matter what characteristics interpreters believe that Wittgenstein hoped his most polished writings would achieve (and there is a wide range of opinions on this subject), I believe that their readings of *On Certainty* would benefit from comparing the remarks written in 1950 with those from 1951, trying to determine what characteristics the later material has that the earlier material lacks, and explaining why they believe *Wittgenstein* would think that the later work was of higher quality. (Rogers [2011], p. 128, footnote 2)

Rogers himself meets this challenge by identifying traces of therapy (as he understands it) in part 1, none in part 2, and consistent appearance in part 3.

9. These letters are quoted in Rogers [2011], pp. 74 and 1, respectively, and form only one small bit of the case made there. See also Rogers [2010], p. 360.

He concludes that Wittgenstein never gave up his therapeutic approach to philosophy, since it was fully present in the final writings that he valued most.

My own understanding of the structure of Wittgensteinian therapy differs from Rogers's. He sees it as loosely analogous to Freudian analysis,¹⁰ addressed to the particular needs of the individual philosopher being treated, so it's important that Moore is addressed only minimally in part 1,¹¹ not at all in part 2,¹² and more personally in part 3 (during which time Wittgenstein discussed these matters directly with Moore in Cambridge). The therapeutic model I've sketched here treats intellectual tendencies more generically, diagnosing how an apparently insuperable puzzle can arise when philosophers, even those with apparently opposing views, are somehow blind to the perfectly ordinary answers to their questions. And often the philosophical views offered, as well as the presuppositions that blind their proponents, are tendencies of thought Wittgenstein himself finds, or has in the past found, captivating. If the final writings on skepticism are to be fitted to this template, the interpretive speculation mentioned a moment ago will involve imagining how this first-draft material might have been sorted, revised, and arranged into these various voices.

To begin at the beginning, the passages in part 1 come not from bound notebooks but from five loose sheets of large paper, each folded in half to form four pages, that were discovered in Anscombe's home in 1967—long after Wittgenstein's death and just in time to be included in the original publication of *On Certainty*. It's unknown whether Wittgenstein left them behind deliberately or just forgot about them when he moved to his doctor's home for those final productive months after treatment had ended. The dating of the sheets is also subject to dispute. Rogers argues persuasively that they were written either during or shortly after Wittgenstein's stay with Malcolm, as his illness was progressing but before its diagnosis and the beginning of treatment.¹³ Under the circumstances, it's not surprising that there

10. See Rogers [2011], chapter 4, p. 111: 'While Wittgenstein certainly didn't model his procedure on psychoanalysis, aspects of his philosophical project can be illuminated through comparison with some of Freud's descriptions of the therapeutic process.'

11. E.g., to get him to see that he hasn't answered the skeptic's question (*OC* §§19, 20).

12. E.g., *OC* §137: 'Moore's assurance that he knows . . . does not interest us. The propositions, however, which Moore retails as examples of such known truths are indeed interesting.'

13. See Rogers [2011], pp. 54–62.

are conspicuous traces of Malcolm-like thinking about the use and misuse of the word ‘know’.¹⁴ For a serious ordinary language examination of ‘know’, there’s always Austin [1946], but in any case this isn’t the therapeutic strain we’re tracing here. Let’s look instead at the direct confrontation with Moore in these passages.

The manuscript begins, of course, with a reference to Moore’s ‘Proof of an external world’:

If you do know that *here is one hand*, we’ll grant you all the rest. (OC §1)

Moore, famously, offered

holding up my two hands, and saying, as I make a certain gesture with the right hand, ‘Here is one hand’, and adding, as I make a certain gesture with the left, ‘and here is another’ (Moore [1939], p. 146)

as proof that there are ‘things outside of us’, an external world.¹⁵ If there’s a voice of temptation to be found in *On Certainty*, it’s a good guess that it belongs to Moore—or better, to Moore as Wittgenstein understands him.¹⁶ To fill in the analogy, much as the discussion in the *Investigations* can be seen as addressing the question ‘is there such a thing as following a rule (or meaning something by one’s words)?’, the discussion in *On Certainty* addresses ‘is there an external world?’ In each case, the voice of temptation answers ‘yes’: 1002 is what follows after 1000 by the rule ‘plus-2’; here is a hand and here is another. So far so good.

In the *Investigations*, the skeptical voice then challenges the voice of temptation with the case of the recalcitrant student, which leads to the long dialogue between them. There is less of this in *On Certainty*, which stands to reason given that rule-following skepticism is a new breed, while the

14. Cf. Coliva [2010], p. 2: ‘Wittgenstein’s reception of [Moore [1925] and [1939]] was mediated by the reading of them afforded by one of Moore’s first commentators, viz. Malcolm. Yet Malcolm’s reading of Moore itself was mediated by his early attendance of Wittgenstein’s lectures (1938/1939) . . . This complex entanglement poses a number of historically difficult questions’.

15. See Essay #4.

16. From now on, I use ‘Moore’ to mean ‘Moore as Wittgenstein understands him’ unless otherwise indicated.

considerations that motivate skepticism about our knowledge of the external world are so venerable and so well-known as to need no rehearsal. In the entire book, a few passing references to dreaming¹⁷ and occasional allusions to the sense-datum inference¹⁸ are about all we hear from the traditional skeptical arsenal. What we do hear from the skeptic's perspective in part 1 are protests that 'here is a hand' doesn't solve the problem:

The statement 'I know that here is a hand' may then be continued, 'for it is *my* hand that I'm looking at.' Then a reasonable man will not doubt that I know.^[19]—Nor will the [skeptic];²⁰ rather he will say that he was not dealing with the practical doubt which is being dismissed, but there is a further doubt *behind* that one.—That this is an *illusion* has to be shewn in a different way. (OC §19)

My believing the trustworthy man stems from my admitting that it is possible for him to make sure. But someone who says that perhaps there are no physical objects makes no such admission. (OC §23)

Faced with the question 'is there an external world?', we have the voice of temptation insisting that 'here is a hand' and the skeptic, presumably backed by one or another skeptical scenario, responding that she has not addressed the sort of doubt he's raising.

17. These occur only in part 3. The final entry in the book includes the claim that 'Someone who, dreaming, says 'I am dreaming' . . . is no more right than if he said in his dream 'it is raining', while it is in fact raining' (OC §676). (See also §383.) Perhaps Wittgenstein never experienced a lucid dream (that is, a dream in which the dreamer is aware that the current experience is a dream). Researchers studying this phenomenon have devised a method for the sleeper to signal a lucid dream state by voluntary eye movements (see [2017a], pp. 25–26).

18. See OC §90: "I know" is supposed to express a relation, not between me and the sense of a proposition (like "I believe") but between me and a fact. So that the *fact* is taken into my consciousness. (And here is the reason why one wants to say that nothing that goes on in the outer world is really known, but only what happens in the domain of what are called sense-data.)' See also §426.

19. Incidentally, Wittgenstein sometimes appears to believe that we can't be mistaken about our own hands: '[I]t is not true that a mistake merely gets more and more improbable as we pass from the planet to my own hand. No: at some point it has ceased to be conceivable' (OC §54). In fact there is a condition (xenomelia) whose sufferers believe that some limb is not part of themselves. Cf. §360: 'I KNOW that this is my foot. I could not accept any experience as proof to the contrary'. Apparently some experiences, for some people, do count against. Coliva [2010], 3.1.1, requires that doubts 'must have consequences in practice' (p. 104). Grizzly as it must seem, sufferers from this condition often seek to have the offending limb amputated.

20. Wittgenstein writes 'idealist', but it's clear from the context that he means what we usually call the 'skeptic'.

Part 2 of *On Certainty* is the collection of material selected by the editors from various notebooks all dating to the period of Wittgenstein's cancer treatments. These passages contain most of the discussion of so-called 'hinge propositions',²¹ which include Moore's 'here is a hand':

I should like to say: Moore does not *know* what he asserts . . . but it stands fast for him. (OC §151)

Can't an assertoric sentence . . . be isolated from doubt? (OC §87)

It may be . . . that *all enquiry on our part* is set so as to exempt certain propositions from doubt . . . They lie apart from the route travelled by inquiry. (OC §88)

At times, Wittgenstein seems motivated by a fairly straightforward foundationalism:

To be sure, there is justification; but justification comes to an end. (OC §192; see also §110)

Propositions like Moore's are these endpoints, claims for which no further justification can be offered:

I want to say: my not having been on the moon is as sure a thing for me as any grounds I could give for it.

And isn't that what Moore wants to say, when he says he *knows* all these things?—But is his knowing it really what is in question, and not rather that some of these propositions must be solid for us? (OC §111–112)

In fact, these propositions are constitutive of the very practice of offering justifications, of confirming or disconfirming:

The *truth* of certain empirical propositions belongs to our frame of reference. (OC §83)

21. The earlier §§35–36 might be considered foreshadowings.

All testing, all confirmation and disconfirmation of a hypothesis takes place already within a system. And this system . . . belongs to the essence of what we call an argument. The system is not so much the point of departure, as the element in which arguments have their life. (OC §105)

[These propositions are] the inherited background against which I distinguish between true and false . . . their role is like that of rules of a game. (OC §§94–95)

Regarding [them] as absolutely solid is part of our *method* of doubt and enquiry. (OC §151)

In addition to these metaphorical comparisons to ‘frames of reference’ and ‘rules of a game’, we hear of ‘the axis around which a body rotates’ (OC §152) and the ‘river-bed of thoughts’ (OC §97), which consists

partly of hard rock, subject to no alteration or only to an imperceptible one, partly of sand, which now in one place now in another gets washed away, or deposited. (OC §99)

Perhaps it’s some confirmation of Rogers’s analysis that the now-standard terminology comes only in part 3:

That is to say, the *questions* that we raise and our *doubts* depend on the fact that some propositions are exempt from doubt, are as it were like hinges on which those turn. (OC §341)

That is to say, it belongs to the logic of our scientific investigations that certain things are *indeed* not doubted. (OC §342)

But it isn’t that the situation is like this: We just *can’t* investigate everything, and for that reason we are forced to rest content with assumption. If I want the door to turn, the hinges must stay put. (OC §343)

In sum, then, though Moore seems to affirm an ordinary empirical claim—that here’s a hand—this is actually a proposition with ‘a peculiar logical role in the system of our empirical propositions’ (OC §136).

As has probably been obvious for some time now, my suggestion is that the theory of hinge propositions comes from the voice of correctness, analogous to Kripkenstein’s community view of rule-following and meaning. Let me hasten to add that to say these aren’t Wittgenstein’s own considered views is not at all to suggest they lack philosophical interest: the rule-following paradox and the prospects for its solution continue to draw attention in philosophy of language, and I don’t have to convince this audience that hinge epistemology has an important role in the contemporary debate there! But my goal here is to lay out this particular therapeutic reading of the Wittgensteinian opus, so I hope you’ll bear with me.

Notice that the appeal to hinge propositions is a ‘skeptical solution.’²² Moore, as the voice of temptation, has insisted that ‘Here’s a hand’, partly on the evidence of his senses.^{23,24} The skeptic counters one way or another that this purported evidence has no force. Much as the Kripkensteinian voice of correctness accepts the skeptical conclusion that nothing about me or my intentions or my training or anything else pre-determines that ‘1002’ is the correct step after 1000, the hinge theorist agrees with the skeptical conclusion that Moore has no evidence for his claim. Much as the Kripkensteinian goes on to offer a communitarian account of what nevertheless supports

22. Something like this may underlie Pritchard’s sense that the Wittgensteinian component of his ‘biopic’ solution to the skeptical paradoxes leaves us with a form of ‘epistemic vertigo’ (Pritchard [2016], pp. 184–188). It’s unsettling to learn that ‘Here’s a hand’ isn’t a straightforwardly justified belief.
23. See Coliva [2010], p. 34, where we are referred to Moore [1941]: in connection with claims like ‘I have in my hand some sheets of paper with writing on them’ (p. 227), Moore writes, ‘I had . . . at the time I made it, *the evidence of my senses*. I do not mean by this that the evidence of my senses was the *only* evidence I had for [it]: I do not think it was. What I mean is that . . . I was seeing or hearing or feeling things (or, if that will make my meaning clearer, ‘having visual, auditory, tactile or organic sensations’) . . . such that to see or hear or feel those things *was* to have evidence (not necessarily *conclusive* evidence) for part at least of what I asserted when I asserted the proposition in question’ (p. 243). It should be acknowledged that Moore published this piece only when he collected his papers in 1959 and that he wrote in the introduction ‘There are bad mistakes in it which I cannot yet see how to put right’ (Moore [1959], p. 13). Still, it seems likely that his reservations concerned the discussion of dreaming in the final few pages, not the passage just cited. (See Weatherall [2017] for an interesting take on this.)
24. Moyal-Sharrock ([2004], pp. 122–123) argues that Moore doesn’t have the evidence of his senses: a belief may be caused by experience without being based on or derived from experience. But even if I don’t note my experience first, then infer the existence of my hands, I might legitimately respond to a challenge to that belief by saying that I see them, can move them at will, etc.

our assertion that ‘1002’ is correct, the hinge theorist goes on to offer an account of hinge propositions that supports our right to them: they ‘stand fast’ for us; they form the necessary backdrop against which presenting evidence becomes possible. So the analogy clearly suggests that this strain in Wittgenstein’s thought, so prominent in part 2, would eventually have found its way to the voice of correctness. In partial response to Rogers’s challenge: Wittgenstein may well have felt frustrated with his work during this period at least partly because he felt he himself was unable to shake off this particular philosophical theory.

So, what happens in part 3 of *On Certainty*, after a six-month hiatus, when the curtain in his brain goes up?²⁵ One notable point, to begin, is that the debate between the skeptic and the voice of temptation takes a new form. Part 1 left off with the skeptic insisting that ‘here’s a hand’ doesn’t establish the existence of an external world, presumably on the basis of one or another familiar skeptical challenge, and schooling the innocent Moore, the voice of temptation, on the Doubt Behind the Doubt. In the fairly early going of part 3, a more sophisticated Moore appears to have learned this lesson, now uttering his anti-skeptical claim—‘I know that’s a tree’²⁶—with a crucial new inflection:

What I am aiming at is . . . the difference between the casual observation ‘I know that that’s a . . .’, as it might be used in ordinary life, and the same utterance when a philosopher makes it. (OC §406)

25. Rogers, with his more individually targeted notion of therapy, places the emphasis on the character of Wittgenstein’s engagement with Moore in part 3. Rather than presenting a theory of meaning that ‘I know this is a hand’ violates (as in part 1), he pressures Moore to give sense to this claim: ‘After repeatedly giving various definite senses . . . yet only producing sentences that are philosophically impotent, Moore may come to [see] that this sentence only seems to become clear once it is given a practical, non-metaphysical use’ (Rogers [2011], p. 157; see also Rogers [2010], p. 362). This is a development of the Malcolm-like line about the use of ‘know’ that’s been set aside here.

26. During this period, Wittgenstein had discussions with Moore in Moore’s garden: ‘I am sitting with a philosopher in the garden; he says again and again “I know that that’s a tree,” pointing to a tree that is near us. Someone else arrives and hears this, and I tell him: “This fellow isn’t insane. We are only doing philosophy”’ (OC §467). It’s hard not to think that a little dose of Grice would have helped immeasurably, as Wittgenstein himself occasionally foresees: ‘Do I know that I am now sitting in a chair? . . . In present circumstances no one is going to say that I know this . . . But now, even if one doesn’t say it, does that make it *untrue*?’ (OC §552). Coliva [2010], 2.4 considers Searle’s version of this concern.

For when Moore says ‘I know that that’s . . .’ I want to reply ‘you don’t *know* anything!’—and yet I would not say that to anyone who was speaking without philosophical intention. (OC §407)

This theme recurs a few weeks later:

When one hears Moore say ‘I *know* that that’s a tree’, one suddenly understands those who think that that has by no means been settled. (OC §481)

It is as if ‘I know’ did not tolerate a metaphysical emphasis. (OC §482)

Wittgenstein doesn’t fill in what it is to speak with ‘philosophical intention’ or ‘metaphysical emphasis’, but it must be to assert ‘here’s a hand’ or ‘I know that’s a tree’ on some basis that’s purportedly immune to skeptical challenge, on some incontrovertible grounds.

Meanwhile, the voice of correctness remains much as it was in part 2,²⁷ but accompanied now by some hints of hesitation. For example, the entries for March 19th begin:

Here I am inclined to fight windmills, because I cannot yet say the thing I really want to say. (OC §400)

There follow four entries that sound the hinge proposition theme:

I want to say: propositions of the form of empirical propositions, and not only propositions of logic, form the foundation of all operating with thoughts (with language). (OC §401)

[T]he statements in question are statements about material objects. And they do not serve as foundations in the same way as hypotheses which, if they turn out to be false, are replaced by others. (OC §402)

27. Occasionally, the range of hinge propositions seems to stretch to new lengths: ‘We may say we know that water boils and does not freeze under such-and-such circumstances. Is it conceivable that we are wrong? Wouldn’t a mistake topple all judgment with it? More: what could stand if that were to fall?’ (OC §558). (See also §§599, 613.)

To say of [a] man, in Moore's sense, that he *knows* something; that what he says is therefore unconditionally the truth, seems wrong to me.—It is the truth only inasmuch as it is an unmoving foundation of his language game. (OC §403)

I want to say: it's not that on some points men know the truth with perfect certainty. No: perfect certainty is only a matter of their attitude. (OC §404)

Then Wittgenstein abruptly returns to the hesitation at the start:

But of course there is still a mistake even here. (OC §405)

This hesitation appears to arise when Wittgenstein asks himself if the hinge propositions are truly constitutive of confirming and disconfirming, if the language game would truly collapse without them:

What if it *seemed* to turn out that what until now has seemed immune to doubt was a false assumption? Would I react as I do when a belief has proved to be false? or would it seem to knock from under my feet the ground on which I stand in making any judgments at all? . . . Would I simply say 'I should never have thought it!'—or would I (have to) refuse to revise my judgment—because such a 'revision' would amount to the annihilation of all yardsticks? (OC §492)

Perhaps the answer is still supposed to be the latter, but consider:

I look at an object and say 'That is a tree,' or 'I know that that's a tree.'—Now if I go nearer and it turns out that it isn't, I may say 'It wasn't a tree after all' or alternatively I say 'It *was* a tree but now it isn't any longer'. But if all the others contradicted me, and said it never had been a tree, and if all the other evidences spoke against me—what *good* would it do me to stick with my 'I know'? (OC §503)

Even in connection with 'I believe my name is L.W.', he denies that 'I am infallible about it' (OC §425). These notes of concern lend at least some support

to the notion that hinge propositions would have ended up in the voice of correctness.

Once the voices of temptation and correctness are in conflict, the next step is to return to our ordinary practices, to the grey rags and dust. In the rule-following case, this meant revisiting the recalcitrant student and reminding ourselves of what would actually happen if someone were to behave that way in an everyday learning situation. The strict analogy in the current case would return us to the skeptical scenarios, to point out that we have perfectly good ways of determining that we aren't dreaming, that we have no reason, phenomenological or experimental, to suppose there are such things as sense-data, and so on.²⁸ Given that the skeptical scenarios haven't played any explicit role so far, it's perhaps not surprising that this sort of thing doesn't turn up in *On Certainty*.

Still, Wittgenstein does 'assemble some reminders' of our ordinary practices, as he does in the passages just cited: in the face of surprising evidence, we respond 'I should never have thought it!' or 'It wasn't a tree after all!' In more mundane cases:

Someone with bad sight asks me: 'do you believe that thing we can see there is a tree?' I reply 'I *know* it is; I can see it clearly and am familiar with it'.—A: 'Was [N.N.] at home yesterday?' I: 'Yesterday he was—I know he was; I spoke to him.' (OC §483)

In these cases, then, one says 'I know' and mentions how one knows, or at least one can do so. (OC §484)

If someone believes something, we needn't always be able to answer the question 'why he believes it'; but if he knows something, then the question 'how does he know?' must be capable of being answered. (OC §550)

And if one does answer this question, one must do so according to generally accepted axioms. *This* is how something of this sort may be known. (OC §551)

28. See the first two lectures of [2017a].

I am taught that under *such* circumstances *this* happens. It has been discovered by making the experiment a few times. Not that that would prove anything to us, if it weren't that this experience was surrounded by others which combine with it to form a system. Thus, people did not make experiments just about falling bodies but also about air resistance and all sorts of other things.

But in the end I rely on these experiences, or on the reports of them, I feel no scruples about ordering my own activities in accordance with them.—But hasn't this trust also proved itself? So far as I can judge—yes. (OC §603)

So, presumably, I know this is a hand, because I see it, feel it, manipulate it quite as usual.

The further diagnostic question, then, is what keeps the voices of temptation and correctness from appreciating the force of these ordinary methods of appealing to evidence in support of our claims? Presumably, as in the rule-following case, something in the skeptic's challenge has thrown them off the scent. There, the student's endless recalcitrance made both voices think they had to provide explanations of meaning that couldn't be misunderstood, rules that couldn't be misinterpreted—guaranteeing that this sort of recalcitrance could not arise. This left them unable to rely on our natural reactions to training, on the effectiveness of our practices in pursuit of our interests, on the very general features of the world that make all this work. In the present case, the skeptic's scenarios make us think we need evidence that guarantees that the radical skeptical scenarios are false, that there is no Evil Demon, no radical mismatch between sense-data and the world, and so on. Suddenly our ordinary evidence, our ordinary methods, no longer seem good enough; we need an extraordinary kind of evidence that ratifies our belief in our hands without appeal to anything else we think we know about the world—without appeal to any facts, for example, about what dreaming is like, or what our visual systems are like, or any other of the ordinary realities that underlie our everyday ways of finding out about the world.²⁹

29. Cf. Stroud's 'special philosophical reflection on our knowledge of the world': whose 'special features . . . make it impossible to rely on the reassuring kinds of answers we find it so easy to give to what sound like similar questions in everyday life' (Stroud [1996], pp. 132–133).

In this way, the skeptic alerts the voice of temptation to the doubt behind the doubt, the doubt that can't be turned away by observing that 'here's a hand' on the basis of perfectly ordinary evidence. This is what inspires that voice to repeat her observation but with 'philosophical intent' or 'metaphysical emphasis'—which must mean replacing the ordinary evidence with extraordinary evidence that would guarantee the truth of 'here's a hand' all by itself, without appeal to any contingent facts about the world. The voice of correctness, in contrast, accepts the skeptic's conclusion that there is no such extraordinary evidence. In response, he goes further: there is no evidence at all for 'here's a hand' because it's one of those propositions that must stand fast before there can be any possibility of evidence, of confirming or disconfirming.

The point is that either way, the ordinary evidence has been left behind; the skeptic has convinced both voices that the only acceptable evidence must be extraordinary. The voice of temptation attempts to find such evidence; the voice of correctness gives up on evidence entirely and hopes for another way. The cure, then, would be to get both parties to realize that we don't need evidence that works in splendid isolation, without appeal to any contingencies; we just need evidence that works for us in this world:

Whether I *know* something depends on whether the evidence backs me up or contradicts me. (OC §504)

It is always by favor of Nature that one knows something. (OC §505)

Both skepticism and the rule-following paradox illustrate a general philosophical tendency that Wittgenstein diagnoses in many contexts: the impulse to insist on the ideal—which doesn't work because it can't be found—leaving behind the real—which does work, for us in our world, and can be found.³⁰

That's the main story I want to tell about the role of hinge propositions in Wittgenstein's discussion of skepticism, but there was one last twist to the therapeutic reading of the rule-following discussion: the distorting

30. In response to Rogers's challenge, then: Wittgenstein was only semi-satisfied with part 1 because he was feeling his way into a new (to him) debate under the influence of Malcolm; he was dissatisfied with part 2 because he felt trapped by the voice of correctness; he was happier while writing part 3 because he was beginning to find his way back to the grey rags and dust.

presupposition shared by the voices of temptation and correctness was a doctrine Wittgenstein once embraced in the *Tractatus*, the priority of sense. A final touch for the analogy, then, would be to locate the demand for extraordinary evidence in the *Tractatus*. Unfortunately the *Tractarian* characterization of epistemology as ‘the philosophy of psychology’ (4.1121)—and thus not relevant to the book’s concerns—leaves little hope of anything conclusive on this score.³¹ If the picture theory were given a verificationist interpretation, then this, in combination with the priority of sense, would immediately imply that evidential relations are part of the system of representation, independent of any contingencies, but this reading of the book has been regarded as implausible for some time now, quite deservedly it seems to me.³²

Still, whatever may be true about the *Tractatus*, there’s no doubt that Wittgenstein was attracted to verificationism during his talks with the Vienna Circle and in early post-*Tractarian* writings like the *Philosophical Remarks*:

How a proposition is verified is what it says . . . The verification is not *one* token of the truth, it is *the* sense of the proposition. (PR §166(5–6))

In addition, it appears that the rejoinder

And then the sense of one proposition depends on whether another is true or false (PR §113(3))

31. There may be something relevant in the discussion of probability in the 5.15s, but if so, I’ve been unable to find it.

32. Still there is this odd passage from part 2 of *On Certainty*: ‘If everything speaks for a hypothesis and nothing against it, is it objectively *certain*? One can *call* it that. But does it *necessarily* agree with the world of facts? At the very best it shows us what ‘agreement’ means . . . What does this agreement consist in, if not in the fact that what is evidence in these language games speaks for our proposition?’ (OC §203). Is the view being floated here that ‘agreeing with the world of facts’—truth—is just having all evidence in favor and none against? Does a proposition’s sense—what it is for a proposition to be true—consist of its confirming evidence? The passage concludes, after the final sentence quoted above, with the parenthetical notation (*Tractatus Logico-Philosophicus*) as if the view in question is being attributed to the earlier book. (The editors remark, by the way, that the first part of the paragraph displayed here was crossed out in the manuscript, but the final sentence, with the parenthetical attribution, was not.) In contrast, Coliva [2010], pp. 162–163, takes the parenthetical to mark a departure from the *Tractatus*, not a worry that ‘it sounds all too reminiscent of the *Tractatus*’. (For what it’s worth, when Carnap assimilates what he takes to be the *Tractarian* lesson that some aspects of our theory are parts of the linguistic framework, not descriptive of the world—analytic, not synthetic—he includes the evidential rules on the analytic side of the ledger.)

is still intended as a *reductio*, on an implicit appeal to the priority of sense. So it seems likely that here, too, the distorting presupposition was one that appealed to Wittgenstein earlier in his career.

In sum, then, I propose that this therapeutic Wittgenstein wasn't himself a hinge epistemologist, whatever the independent merits of that position might be. Operating as we should with the grey rags and dust of our ordinary epistemic practices, we take ourselves to have compelling evidence for our belief that 'here's a hand'—we can see it, feel it, manipulate it as usual—in stark disagreement with the notion that it's among the unconfirmed hinge propositions of the voice of correctness. On the other hand, we also don't feel the need for the voice of temptation's 'metaphysical emphasis.' Wittgenstein's diagnosis is that the skeptic has lulled both voices into thinking they need a form of extraordinary evidence that confirms our claims from scratch, with no help from any contingencies. Once again he reminds us that we don't need methods that work in a vacuum; we only need methods that work for us, given our natural ways of reacting, our interests and goals, and the world in which we find ourselves.

One last remark. On this reading, notice that Moore as Wittgenstein understands him in part 1—the innocent Moore who simply answers the question 'is there an external world?' by appeal to his two hands—is actually right. This is a perfectly good reply. In fact, I think this is the real Moore, except for the implication in part 1 that he doesn't understand the doubt behind the doubt and needs to be instructed. I would argue that Moore understands perfectly well what the skeptic wants—understands that he wants extraordinary evidence (what Moore calls 'proof')—and straightforwardly admits that this is something he can't provide. The key is that this Moore, whom I take for the real Moore, insists that his failure to produce extraordinary evidence doesn't undermine the ordinary evidence, the compelling evidence, he has for the existence of his hands. In other words, it seems to me that this Moore and this therapeutic Wittgenstein are actually in agreement.³³ But that's a topic for another time.³⁴

33. See [2017a], pp. 158–175, Essay #4. Perhaps it's worth noting that despite his defense of common sense (Moore [1925]), Moore apparently didn't regard claims like 'here's a hand' as special, as uniquely effective in response to the skeptic. In his lectures of 1933–1934, he takes scientific evidence for the existence of stars and dinosaurs as 'indirect proofs' of the existence of material objects.

34. Thanks to Annalisa Coliva and Duncan Pritchard for organizing such a lively and enjoyable conference. I'm grateful to several participants for helpful discussions, especially Adam Leite and Genia Schönbautsfield.

PART III

LOGIC AND LANGUAGE

A Note on Truth and Reference

As she pursues her wide-ranging investigation of the world and our place in it, the Second Philosopher will soon come upon the question of human language use. The speaking, hearing, writing, reading, signing of natural languages obviously plays a fundamental role in human life. How does this work? Why is it so effective? To take an example, suppose my mind wanders while waiting on a traffic light; the light changes; my passenger utters the sounds ‘the light is green’; I snap to attention and proceed cautiously into the intersection. What just happened? My companion’s action has corrected and improved my behavior, to the benefit of us both. How did a mere series of sounds accomplish this?¹

It’s natural to think that at least part of the complex answer to this question will be that ‘the light’ in this situation referred to the traffic signal in our vicinity and that ‘the light is green’ was true at the time it was uttered. This is at least one conspicuous way in which the Second Philosopher comes face-to-face with questions of truth and reference.²

1. Hilary Putnam posed this challenge in his [1976], pp. 100–101: ‘We say certain things, conduct certain reasonings with each other, manipulate materials in a certain way, and finally we have a bridge that enables us to cross a river that we couldn’t cross before. And our reasoning and discussion is as much part of the total organized behavior-complex as is our lifting of steel girders with a crane . . . [What needs explaining is] the *contribution* of our linguistic behavior to the success of our total behavior.’
2. The goal of this essay is to extract the view of truth and reference in [2007], Part II, and present it as simply and straightforwardly as I can. (E.g., §II.6 of [2007] adopts an extensive, albeit friendly amendment to the position in §II.4: ‘indication relations’ are subsumed under Wilson’s more wide-ranging and flexible machinery of ‘directivities’ and ‘supports’. Here the amended version is simply presented from the start.) More history of and background to the debate is included there and used to illustrate how the Second Philosopher’s methods differ even from those of others typically regarded as ‘naturalists’. Of course, when one revisits writings over a decade old, one tends to see matters with fresh eyes. The main novelty here comes with the Austinian material toward the end and related discomfort with the label ‘disquotationalism’.

To begin with, she takes it to be obvious that the truth of ‘the light is green’ isn’t a matter of what’s useful or rational to believe or of her experiences or ways of checking; it’s a matter of the traffic light and its color, quite independently of her. She’s also happy to endorse the commonplace that

‘the light is green’ is true if and only if the light is green,

but this obviously isn’t much help in explaining what she hopes to explain. To go beyond this, she assembles some everyday observations:

We’re trained, explicitly or implicitly, to apply the term ‘traffic light’ under certain conditions: they’re shaped in certain ways, display certain colors; they generally appear in certain locations; ideally, they have certain effects on traffic flow. In the customary circumstances, the guidelines for the term’s use correlate with the presence of traffic lights, and this correlation explains the effectiveness of that usage. Given this background, your utterance leads me to expect that you’re calling my attention to the traffic light salient in our vicinity.

In his detailed studies of word-world connections, Mark Wilson has coined the terms ‘directivity’ for the guidelines we use in applying a term and ‘support’ for the worldly phenomena whose correlation with our directivities accounts for our successful usage.³ Our directivities aren’t fully determinate⁴—if a government functionary stands in the intersection and selects from a trio of colored flashlights to shine toward oncoming cars, are the flashlights an unusual kind of traffic light or a tool that can be used in place of traffic lights?—but this presents no problem for our everyday usage.

Simple cases like this might tempt us to pursue an account of truth and reference based solely on directivities and supports, on our assessment of what makes language effective, but a review of philosophers’ less straightforward

3. Readers of Wilson [2006] will realize that the language of ‘directivities’ and ‘supports’ is only the beginning of his illuminating analytical vocabulary: supports can be ‘direct’ or ‘distributed’, seemingly unified supports can actually display ‘façade’ structures of various sorts, and so on.

4. Though we sometimes imagine otherwise. See Austin’s abusive cat (Austin [1940], pp. 67–68) or Wilson’s Druids (Wilson [1982], [2006], pp. 34–37, summarized in my [2007], pp. 186–188).

cases councils otherwise. Consider, for example, Hartry Field's story of the ancient Greek whose directivities for 'Zeus is throwing thunderbolts' correlate neatly with the worldly support of lightning; his utterance would serve as a warning to others.⁵ Or Wilson's more colorful account of the medieval Japanese swordsmith who prepares a new metal blade by plunging it into the belly of a particularly brave prisoner (believing this to endow the blade with more than usual strength). One of his directivities for 'superior weapon' is that it has been so plunged, which just happens to correlate with the worldly support of rapid temperature-lowering to lock in the ferrite grain.⁶ These usages are no doubt effective, but perhaps we'd hesitate to say that the Greek is referring to lightning or that the swordsmith's 'a superior weapon gains strength from the bravery of the noble foe' is true.

So, if we're to build an account of truth and reference on the machinery of directivities and supports, at the very least we'll need to be selective about which correlations we're willing to count as referential. A similar challenge has long faced partisans of the causal theory of reference: which causal chains are 'appropriate', that is, which establish reference and which don't? This has turned out to be a devilishly difficult question to answer. One of the most sophisticated attempts comes from Philip Kitcher and Kyle Stanford (Kitcher and Stanford [2000]); building on Kitcher's earlier work on Priestley (Kitcher [1978]), they aim to produce a more refined (and more defensible) version of the causal theory. Obviously, oxygen was correlated with, served as worldly support for, Priestley's effective use of the term 'dephlogisticated air', but did he refer to oxygen?

As Kitcher acknowledges (Kitcher [1978], p. 533), there's certainly a temptation to think that he did and that he said some true things about it; a quick Google search on 'who discovered oxygen?' turns up this from the educational pages of the website of the American Chemical Society (www.asc.org):

The world recalls Priestley best as the man who discovered oxygen . . . In a series of experiments culminating in 1774, Priestley found that 'air is not

5. Field [1994], p. 100.

6. Wilson [2006], pp. 228–229.

an elementary substance, but a composition', or mixture, of gases. Among them was the colorless and highly reactive gas he called 'dephlogisticated air', to which the great French chemist Antoine Lavoisier would soon give the name 'oxygen'.

At the other extreme, contemporary instrumentalists in the philosophy of science often include Priestley's account among the successful but irredeemable false theories of the past:

The historical cases briefly described [which include Priestley's] illustrate a variety of grounds on which even very successful past scientific theories have been judged fundamentally mistaken rather than even approximately true by the lights of contemporary theoretical orthodoxy. (Stanford [2018], p. 218)

How are these to be reconciled?

Kitcher and Stanford⁷ set out to thread this needle by producing a historically motivated refinement of the causal theory, then adding two new elements: different tokens of the same expression can refer in different ways (including descriptively, not just causally), and a single token might partially refer.⁸ The upshot for Priestley is varied. Some of his uses of 'dephlogisticated air' attempt to refer descriptively, relying on the assumption that there's a substance called phlogiston that's given off in combustion, and thus fail to refer, but other of his uses, for example in his claim that 'dephlogisticated air' facilitates combustion and eases respiration, have their reference fixed to air rich in oxygen, in accord with Kitcher and Stanford's refined causal theory. They conclude that, in these and other ways, despite numerous reference failures, many of Priestley's uses refer to air rich in oxygen, or to pure oxygen, or partially refer to both.

When the details of all this are filled in, the authors admit that 'the resultant view of scientific language . . . is extremely complex' and that previous

7. Stanford tempers his support for this joint position in his [2006], p. 161, footnote 11, and more emphatically in his [2018], suggesting something in the direction of the account below. Essay #2, 2.iii, explains why I think this view is in tension with his case for instrumentalism.

8. In Field's sense (Field [1973]).

simpler theories have been replaced ‘with something much more messy’.⁹ Their hope is ‘that greater accuracy justifies the mess’ (Kitcher and Stanford [2000], p. 126). I’d like to suggest that what accounts for the difficulties Kitcher and Stanford (and many others) face in their attempts to devise a substantive theory of reference isn’t that the situation is messy, but that they’re actually chasing a chimera. Let me explain.

What Kitcher and Stanford want is to isolate some feature of Priestley, his physical environment, and his usage that constitutes his referring to oxygen (or to air rich in oxygen or . . .). Notice that in this case and in our other real or imagined examples, we’re out to evaluate referential facts in languages other than our own: ancient Greek, medieval Japanese, 18th-century English. If the fundamental feature of truth is given by the general version of the commonplace the Second Philosopher endorsed at the start, that is,

‘. . .’ is true iff . . . ,

with its referential counterpart,

‘—’ refers to —s (if there are any),

then there’s an obvious problem: ‘true’ and ‘refers to’ apply only to expressions in our current language. I can assess the truth or falsity of Pierre’s ‘Le feu de signalisation est vert’ or the referential achievements of his ‘le feu de signalisation’ only by first translating them into my language.¹⁰

Following this line of thought, consider two observations about the act of translation. First, whatever Kitcher and Stanford may have intended,

9. For a sense of the complexity involved in the causal component alone, consider: ‘A natural kind terms is associated with (a) a range of samples, (b) a range of foils, (c) an open sentence $\Phi[x]$ in conjunctive form $\phi_1 x \ \& \ \phi_2 x \ \& \ \dots \ \phi_n x$, such that each sample satisfies $\Phi[x]$ and each foil fails to satisfy at least one of the ϕ s. Each of the ϕ s can be used by members of the community into which the natural kind terms is being introduced. The term refers to the set of those things having the inner constitution that is a common constituent in the total causes of the presence of each of the ϕ -properties in each of the samples, and that is absent from all of the foils’ (Kitcher and Stanford [2000], p. 114).

10. This kind of translation, or maybe better ‘interpretation’, isn’t necessarily restricted to remote languages. E.g., it might occur within contemporary English when a hospital’s neurologist interprets its acupuncturist’s ‘inserting a needle at this point increases the flow of Qi to relieve pain in the elbow’ as ‘inserting a needle at this point stimulates the central nervous system to release these chemicals which in turn affect pain reports from the elbow’.

any responsible effort to render Priestley's 18th-century English into our modern idiom would properly draw on just the sort of factors they've been so painstakingly identifying. Imagine, for example, the range of subtle environmental and causal factors they'd call on to uncover the facts of reference when Cavendish 'exploded' samples of 'dephlogisticated air' and 'inflammable air' to generate water. For that matter, Wilson would most likely isolate the surrounding directivities and conclude that the worldly support for 'dephlogisticated air' is oxygen and for 'inflammable air' is hydrogen. My point is that the facts surrounding Priestley and Cavendish that turn up in the analyses of Kitcher, Stanford, and Wilson would influence any conscientious translation, and thus any conscientious evaluation of the reference of their terms and the truth of their claims.

The second observation is that translation is, as the saying goes, more an art than a science. If 'feu de signalisation' isn't rendered as 'traffic light' or something close, then something has gone objectively wrong, but in more subtle cases, like the ones under consideration here, reasonable constraints are not so tight. For the purposes of the American Chemical Society, where the goal is to highlight the way Priestley was interacting with the world as we now understand it, translating 'dephlogisticated air' as oxygen is perfectly appropriate, but in a scholarly paper on the history of scientific ideas, where the goal is to trace the nuances of Priestley's thought, 'air supposedly bereft of the in-fact non-existent substance phlogiston' would be better. Even our more fanciful earlier cases would likely show some such variation. For example, Hesiod's poem *Works and Days* is frequently described as a sort of farmer's almanac: a history of meteorology might well regard it as containing useful truths about weather prediction, perhaps even about lightning, despite the term 'Zeus's thunderbolts', while an account of its treatment of Greek mythology would take the fact that Zeus doesn't exist to imply that his thunderbolts don't either.¹¹ Choices

11. I put this in terms of lightning to match Field's presentation, but Adam Chin, who's actually read some Hesiod, reports that Zeus doesn't produce lightning in *Works and Days*. He does, however, control rainfall: 'When the keen sun's strength stops scorching and sweltering, after mighty Zeus begins the autumn rains, . . . then timber is freest from the worm when hewn by the iron, when it sheds its leaves to the ground and stops putting out shoots' ([700 BC?], 414, p. 49). So a better example might be

like these depend crucially on the context and goals of the particular translation.¹²

The point—again, following this particular line of thought—is that the facts about Priestley and his circumstances don't come with a pre-existing reference relation that we can set out to detect; instead, reference is parochial to our language, so translation of Priestley's 'dephlogisticated air' into our language is what assigns it a referent. This means that the reference of his terms and the truth of his claims aren't just a matter of his linguistic practices and his environment; they also involve our language and usage—and even the goals in a specific context of translation. This is the sense in which Kitcher and Stanford were chasing a chimera: there simply is no fact of the matter about reference and truth within the range of phenomena they take to be relevant. This would explain why a satisfactory account has been so difficult to find and why even partial successes are so 'complex' and 'messy'. Giving up this fool's errand would free us to explore word-world relations in their case-by-case detail, without the ill-fated distraction of trying to determine which of these rich correlations deserve to be called 'reference'.

But wait—comes the objection—if 'truth' and 'reference' are to play the role assigned them in the explanation of successful human language use, they have to be more substantial than this; at the very least, they shouldn't be the sort of thing that depends on the vagaries of translation! Surely, we need more than those trivial commonplaces, the R-sentences and T-sentences. Reverting to our original example, it can't be enough to point out that

'traffic light' refers to traffic lights

and that

Hesiod's belief that 'Zeus tends to send rain when the seasons change from summer to autumn.' (Thanks to Chin for this delightful dip into Hesiod.)

12. A nice real-world example comes from Jeremy Avigad's translation of Dedekind: 'Since my goals are not primarily historical, I will generally use contemporary terms to describe the mathematical substance of the developments. For example, where I speak of the ring of integers in a finite extension of the rationals, Dedekind refers to the 'system' of integers in such a field. Readers interested in terminological nuances and historical context should consult the sources cited above' (Avigad [2006], footnote 4).

‘The traffic light is green’ is true iff the traffic light is green;

we need an account of how that term links to this referent and how that sentence comes to have these truth conditions!

I think what’s happened here is a kind of mission creep, a shift of focus that I slipped rather disingenuously into the setting of the problem in my opening paragraphs:¹³ we started out with the challenge to explain why my companion’s utterance ‘the light is green’ was so effective in calling me back to the task at hand and initiating our progress into the intersection; we ended up asking, not what makes the utterance effective, but what makes it true? Obviously, these questions aren’t the same, as the cases of Zeus and the swordsmith amply demonstrate; even if we were to decide, for whatever purposes, to regard their usages as referential or true, our reasons wouldn’t coincide with the answers we would offer to the more straightforward question of their effectiveness. The objector is right that we need more than the R- and T-trivialities—we need something substantive about word-world relations—but it isn’t obvious that what’s called for is more robust theories of truth and reference.

In fact, what we need to answer the question we started with is precisely Wilson’s directivities and supports—as a matter of simple definition. (Recall: directivities are guidelines for usage, and supports are the worldly circumstances, correlation with which makes the usage effective.) Wilson’s style of inquiry is aimed at precisely the original question—why is human language use effective?—and it’s designed to cover all cases, from oddities like Zeus and the swordsmith to familiarities like our little exchange at the traffic light. Finally, as far as I can tell, it does so without any need to specify which correlations amount to ‘reference’ and which claims qualify as ‘true’.

To test this suggestion, consider Kitcher’s sequel to his [1978] on Priestley and his [2000] (with Stanford) on the causal theory: in his [2002], he argues that when bioengineers manufacture new organisms, their activity

13. Readers of Wittgenstein will recall his haunting observation: ‘The first step is the one that altogether escapes notice . . . The decisive movement in the conjuring trick has been made, and it was the very one that seemed to us quite innocent’ (Wittgenstein [1949], §308). On a similar note, Austin is rumored to have observed that ‘in most philosophy papers, it’s all over by the bottom of the first page’. (I’ve been unable to locate a source.)

can't be explained without a robust account of truth as correspondence. He takes as his opponents various deflationists who believe that a minimal notion of truth is all that's needed to explain human goal-oriented behavior. Kitcher mounts a convincing case that this isn't right, that a more detailed and penetrating investigation of the relevant word-world relations is required. Everything we've seen here supports Kitcher on this point—understanding cases from Zeus to traffic lights demands more than R- and T-sentences—but once again, his fascinating story of the bioengineers' processes of trial and error seems more amenable to analysis in the terms of directivities and supports than in terms of truth and reference. (The strain of insisting on the latter shows up in the need to introduce awkward and troublesome notions of 'partial reference' and 'approximate truth'.) Here again, close examination of word-world relations is crucial, but the labels 'true' and 'refers' are not.

My tentative conclusion on behalf of the Second Philosopher is that truth and reference play no working role in explaining the success of human language use—'tentative', obviously, because not every potentially relevant context has been examined.¹⁴ Of course, there remain the handy syntactic uses of 'true' familiar from writings of the disquotationalists—to affirm or deny without repeating ('What Mary said was true', 'That's false!'), to affirm or deny an indefinite or even infinite number of items at once ('His speech was littered with falsehoods', 'The axioms of ZFC are true')—but these require no substantive theory, no explanation of the T- or R-sentences.¹⁵ Still, it seems unlikely that every natural language use of 'true' or 'refers' falls neatly into one or the other of these categories (irrelevant or disquotational).

In his famous essay 'A plea for excuses', J. L. Austin commends the methods of ordinary language philosophy—the examination of *'what we should say when'*, and so why and what we should mean by it' (Austin [1956a],

14. If someone were to uncover a more substantive role for 'true' or 'refers' in some other context, of course, the Second Philosopher would change her mind about this. She'd then want to explore how these notions function where they do function and whether this could be generalized in some way to other cases. But until that happens, she stands by this tentative conclusion drawn from what are, after all, the most likely seeming candidates.

15. [2007], pp. 130–135, offers two diagnostics to separate correspondence from disquotational accounts of truth: do the R- and T-sentences require substantive explanations, or are they trivial facts about the words 'true' and 'refers'?—are the reference of Pierre's terms and the truth of his claims just a matter of Pierre, his surroundings, and his usage, or do they also involve translation into my language?

p. 181)—but only in ‘a field where ordinary language is rich and subtle, as it is in the pressingly practical matter of Excuses’ (ibid., p. 182), not necessarily in cases where ‘our interests are more extensive or intellectual than the ordinary . . . fed from the resources of the microscope and its successors’ (ibid., p. 185). Austin’s example is ‘time’ (ibid., p. 182), but I think we can also include Wilson’s case studies from applied science—‘temperature’, ‘hardness’, ‘light intensity’—where direct worldly support, though often elusive and complex, is ultimately identifiable.¹⁶ Here, clearly, our interests are more demanding than ordinary and our usage is fed from experimental resources.

Now consider the word ‘know’. Contemporary epistemologists are inclined to include it in Austin’s second group, to see their work as an effort to discover the features of a concept of knowledge or of a primitive item, knowledge itself. As I understand him, Austin resists this picture, counting ‘know’ in the first group instead and engaging in an ordinary language investigation focused on a particular context of usage: the various functions of the question ‘How do you know?’¹⁷ Viewing the case historically, Robert Pasnau observes that early modern philosophers took the epistemologist’s job to be investigating the reliability of various cognitive avenues—perception, memory, deduction, scientific inference—but that they showed no interest in a broader notion of ‘knowledge’, in this ‘very general word of positive cognitive appraisal’ (Pasnau [2017], p. 2).¹⁸ In this context, we might think of ‘know’ as a free-form honorific we tend to bestow in certain circumstances—another usage open to an ordinary language investigation into our practices and their motivations, but not answerable to a worldly support whose contours we’re out to discover.¹⁹

The point of this digression is that I want to suggest that ‘true’ and ‘refers’ in their various contexts belong with ‘excuse’ and ‘know’, not with ‘time’, ‘temperature’, and ‘hardness’. We’ve seen that in the Second Philosopher’s

16. See Wilson [2006] for these case studies (or my [2007], II.6, for summaries).

17. This is what I take to be going on in Austin [1946]. See [2017a], I.3 and pp. 210–214, for discussion.

18. See Essay #1 for discussion.

19. ‘Belief’ might be different, especially if we restrict attention to an individual cognitive mechanism: visual beliefs, memory beliefs, inferential beliefs. Empirical psychologists study each of these, in varying degrees. Notice also that ‘belief’ doesn’t carry the honorific force of ‘know’; believing is not an accomplishment. Nor do these studies confront questions involving papier-mâché barns.

explanatory contexts, ‘true’ and ‘refer’ are best replaced by the machinery of ‘directivities’ and ‘supports’; what’s left is a rich array of ordinary language uses, beginning with those observed by the disquotationalists. In an Austinian spirit, Wilson identifies another type of example: at times in the development of a science, we may find ourselves with a stretch of discourse whose directivities are undeniably effective, but whose worldly supports we don’t understand; as we grapple with this uncertainty, we might highlight a particular utterance with the hunch ‘whatever else happens, this claim has to be true!’²⁰ The application of ‘true’ here comes to an informed prediction: given the partial information we now have, it seems a good bet that when the supports are fully understood, our successors will choose to retain this sentence—that is, translate it homophonically—in their new, improved theory.²¹ Paraphrasing Pasnau, we might say that ‘true’ in this context is functioning as a general word of positive semantic appraisal. No doubt examples could be multiplied.²²

In some ways, the account of truth and reference sketched here meets straightforward criteria for a disquotationalist account:

‘snow is white’ is true if and only if snow is white

is a trivial fact about the use of ‘true’,²³ not something in need of explanation, and the truth or falsity of Pierre’s ‘la neige est blanche’ rests on facts about Pierre, his environment, and the context and goals of translation, not on Pierre and his environment alone.²⁴ Still, I think it would be a mistake to so classify the position. Obviously, it stands in sharp contrast with any disquotationalism motivated by the conviction that all talk of word-world

20. Wilson [2006], pp. 632–633.

21. This sort of move is central to Essay #2, especially §2.iii. (Incidentally, Wilson [2006], pp. 648–649, recognizes that translational choice is involved in evaluating truth and reference of past theories.)

22. Wilson [202?] describes another, involving multiscalar modeling.

23. Deflationists differ on the sense in which the T- and R-sentences are trivial; e.g., they might be analytic or conceptual or some other alternative. I call them ‘trivial’ in the sense that they don’t require a substantive explanation, hoping to encompass all these variations.

24. Recall footnote 15. Of course, in this case, the contribution of the context of translation is minimal because the one reasonable translation is obvious and unlikely to vary with circumstances (as with ‘feu de signalisation’, above), but the case of Priestley’s ‘dephlogisticated air’ demonstrates that this can’t be taken for granted.

relations is dangerously ‘metaphysical’, and we’ve seen (in connection with Kitcher [2002]) that it includes no attempt to explain the success of human language use solely in terms of T- and R-sentences. But beyond these departures, there’s the more fundamental fact that a disquotational theory of truth and reference remains a *theory* of truth and reference. Here I’m not out to define ‘truth’ or to provide any theoretical account at all; the whole point of classifying ‘true’ and ‘refers’ with ‘excuses’ and ‘know’ rather than with ‘time’ and ‘hardness’ is to deny that they’re things to be theorized about rather than words whose complex usage is to be studied.²⁵ So, if labeling there must be, assuming the more general term ‘deflationism’ can be used without definitional or theoretical implications, my tentative proposal on the Second Philosopher’s behalf is an ordinary language deflationary view of truth and reference.

To fill this in a bit—of course, the only so-called ordinary language philosopher I have in mind here is Austin, but I hesitate to call my proposal ‘Austinian’ because I’m not sure it’s the view he himself offers in his paper ‘Truth’ (Austin [1950]). He begins beautifully:

‘truth’ . . . is an abstract noun . . . We approach it cap and categories in hand: we ask ourselves whether Truth is a substance . . . or a quality . . . or a relation . . . But philosophers should take something more nearly their own size to strain at. What needs discussing rather is the use, or certain uses, of the word ‘true’. (Austin [1950], p. 117)

So far, so excellent, but then he goes on—somewhat uncharacteristically, it seems to me—to isolate the ‘primary’ use, the truth of a statement, and to give this analysis:

A statement is said to be true when the historic state of affairs to which it is correlated by the demonstrative conventions (the one to which it ‘refers’) is of a type with which the sentence used in making it is correlated by the descriptive conventions. (Ibid., p. 122)

25. Here Wilson disagrees. This plays out in his [2006], pp. 632–635, and my [2007], pp. 194–196, and more recently in his [202?] and my reply [202?b].

Perhaps in 1950, when this was published, it was easier than it is today to imagine that spelling out these demonstrative and descriptive conventions—which sound for all the world like substantive theories of reference and truth—would be a fairly straightforward undertaking. However that may be, here Austin appears to depart from what I characterize as ordinary language deflationism—or deflationism of any kind, for that matter.

But this may misread Austin's intent, because he also announces that 'the theory of truth is a series of truisms' (Austin [1950], p. 121), which suggests that he doesn't think the T- and R-sentences stand in need of substantive explanation.²⁶ Furthermore, and most remarkably, he goes on to offer a glimpse of an ordinary language approach to 'true' along the lines of his treatment of 'excuses' just a few years later in 'A plea' (Austin [1956a]). To see how this goes, recall that Austin's topic in [1956a] is 'freedom', about which he observes:

'[F]reedom' is not a name for a characteristic of actions, but the name of a dimension in which actions are assessed. In examining all the ways in which each action may not be 'free', i.e. the cases in which it will not do to say simply 'X did A', we may hope to dispose of the problem of Freedom. Aristotle has often been chidden for talking about excuses or pleas and overlooking 'the real problem': in my own case, it was when I began to see the injustice of this charge that I first became interested in excuses. (Austin [1956a], p. 180)

Austin makes the analogy with 'true' explicit in this very passage: it actually begins 'as 'truth' is not a name for a characteristic of assertions, so 'freedom' is not . . .' The same connection is emphatic even back in 'Truth':

We become obsessed with 'truth' when discussing statements, just as we become obsessed with 'freedom' when discussing conduct. So long as we think that what has always and alone to be decided is whether a certain action was done freely or was not, we get nowhere: but so soon as we turn

26. He does reject something like a redundancy theory, insisting that "is true" is not logically superfluous' (Austin [1950], p. 129), but so does the Second Philosopher.

instead to the numerous other adverbs used in the same connection ('accidentally', 'unwillingly', 'inadvertently', etc.), things become easier, and we come to see that no concluding inference of the form 'Ergo, it was done freely (or not freely)' is required. (Austin [1950], p. 130)

Here Austin foreshadows a central theme²⁷ of 'A plea'—setting aside 'freedom' for an extended treatment of excuses—and introduces the analogy with the case of 'truth'.

Given the analogy, then, what corresponds to excuses in the case of truth; where are we encouraged to direct our efforts instead? The response runs in straightforward parallel to the discussion of 'freedom'; instead of insisting that 'what has always and alone to be decided is whether' an assertion is true or not, we should attend to the

numerous other adjectives which are in the same class as 'true' and 'false', which are concerned, that is, with the relations between the words . . . and the world . . . We say, for example, that a certain statement is exaggerated or vague or bald, a description somewhat rough or misleading or not very good, an account rather general or too concise. In cases like these it is pointless to insist on deciding in simple terms whether the statement is 'true or false'. (Austin [1950], pp. 129–130)

He gives a few typically amusing examples—'is it true or false that . . . the galaxy is the shape of a fried egg?'—and concludes that:

There are various *degrees and dimensions* of success in making statements: the statements fit the facts always more or less loosely, in different ways on different occasions for different intents and purposes. What may score full marks in a general knowledge test may in other circumstances get a gamma. (Austin [1950], p. 130)

27. As two other central themes, I count the nature, scope, and working methods of ordinary language philosophy (pp. 181–189—this is where he draws the distinction between excuses and time) and his insistence on the relevance of science (pp. 189, 203–204).

Just as there are better ways to assess conduct than to focus single-mindedly on ‘freedom’, there are better ways to assess statements than to focus on ‘truth.’²⁸ So, in the end, Austin does present us with an intriguing and challenging direction for an ordinary language investigation of truth, by viewing it as just one among many, often more fine-tuned ways of evaluating our utterances.

I hope this much at least makes room for a kind of ordinary language deflationism that regards the R- and T-sentences as trivialities, requiring no explanation, that declines to define or theorize about truth and reference, that takes our assessments of many claims to depend on the vagaries of translation and interpretation—while, at the same time, emphatically denying that the R- and T-sentences are enough to explain everything that needs explaining about human language use. It’s possible to investigate word-world relations without objectionable metaphysics and without robust theories of reference or truth. The hope is that a view like this could help redirect our energies from an ill-fated focus on such theories to a broader investigation of the menagerie of ways our words—including ‘true’ and ‘refers’—relate to the world.²⁹

28. ‘Like freedom, truth is a bare minimum or an illusory ideal’ (Austin [1950], p. 130).

29. My thanks to Sorin Bangu, Adam Chin, Charles Leitz, David Malament, Chris Mitsch, Jeffrey Schatz, and Evan Sommers for enjoyable and productive discussions of earlier drafts.

The Philosophy of Logic

Let me begin by saying that it was an honor and a privilege to serve a term as president of the ASL. The Association is in many ways the most effective, productive, ambitious professional organization I know of, and its most active members—the people I was fortunate to work with during my time in office—are not only dedicated and selfless champions of logic and logicians around the world, but also imaginative and innovative leaders, always open to new directions and new undertakings. I’m immensely grateful to my fellow officers, and the many editors and committee members, for all their good efforts, and also to various old hands of the Association who helped me so much with their advice and wise counsel.

When the Nominating Committee approached me about standing for the presidency, they made clear that my charge would be to reinforce, or even to some extent to reinstate, the traditional role of philosophy in the Association. Back in the 1930s, when the ASL was founded, key players like Alonzo Church and C. I. Lewis comfortably spanned what we now think of as a stark disciplinary divide between mathematics and philosophy, not to mention those who stood at the origins of what we now think of as computer science. As mathematical logic flourished in subsequent decades, it came to dominate the Association’s official organ, the *Journal of Symbolic Logic*, so the new and growing areas of philosophical logic began to move elsewhere. Meanwhile, more purely philosophical and historical studies were eventually left to find their place in the *Bulletin of Symbolic Logic*, though only when they could fit the *BSL*’s defining profile, as work of interest to and accessible to a wide range of the logic community.

Given this history, it's gratifying that the newly founded *Review of Symbolic Logic* now provides a more comprehensive home for the full range of work in philosophical and non-classical logics, in formal epistemology, in philosophy and methodology of mathematics, in history and philosophy of logic, and in applications of logic to computer science and such fields as linguistics and cognitive science. When the new journal was still in the discussion phase, I was pleasantly surprised to find ASL members even from the remotest parts of mathematics insisting on the importance of philosophy and related fields to the ongoing health and vitality of the Association; more than one expressed the firm conviction that its multi-disciplinary reach is the lifeblood of the organization, what gives it its unique purpose and character. With the *RSL*, that strength and shared purpose is re-affirmed. I'm told that the *Review* is already receiving more good papers, across its whole range of subjects, than could be fit into the initial contractual page count, and that the ASL and Cambridge University Press have recently agreed to boost that number for the 2011 volume.

Participating in these happy developments gave me ample opportunity to reflect on the evolving role in the Association of issues on the non-technical, what might be called the 'prose' side of philosophy. Back in the 30s, questions about the nature of mathematics—its subject matter, its proper methods, its foundations—were closely intertwined with technical logical questions about formalization, proof theory, set-theoretic reduction, and so on. Gradually, as technical logic became mathematical logic, the philosophical questions came to be approached as pure metaphysics and epistemology, without much attention to mathematical detail, or to anything beyond arithmetic and elementary geometry, for that matter. More recently, the philosophical study of mathematics has returned to a careful examination of the practice of mathematics; though this study is often less technical than in the 30s, more broadly methodological, it does come back into contact with mathematical logic itself, now as a full-fledged branch of mathematics and a branch particularly ripe for methodological analysis, given its foundational role. In these and other ways, the philosophy of mathematics and mathematical logic have remained intertwined over the years, if the nature of their interconnection has sometimes shifted along the way.

In the course of these ponderings, I was struck by what still seems to me a surprising anomaly: though the philosophy of mathematics has been an ongoing, well-recognized component of the ASL's disciplinary mix, the 'philosophy of x ' that one might expect to be most prominent in the publications and on the programs of the Association for Symbolic *Logic*—that is, the philosophy of *logic*—has been much less in evidence. What I have in mind here isn't philosophical logic—a technical discipline centered on formal systems of particular interest to philosophers—or even philosophical questions about particular logical topics—the significance of Gödel's theorems, the nature of logical constants, the status of higher-order logics, and so on. What I have in mind, rather, is the sort of fundamental philosophical questions that run parallel to the questions typically asked by philosophers of mathematics: not 'what is the subject matter of mathematics—what makes its truths true?' but 'what is the subject matter of logic—what makes logical truths true?'; not 'how do we come to know the truths of mathematics?' but 'how do we come to know the truths of logic?'¹

Among the membership of the ASL, a passing familiarity with the three great schools in the philosophy of mathematics at the turn of the 20th century—Logicism, Intuitionism, and Formalism—is fairly common coin; there's also a general awareness that Logicism, the idea that mathematics is reducible to logic, soon gave way to Set-Theory-ism, the idea that mathematics is reducible to set theory, a view now often identified with Realism or Platonism about mathematical objects.² But I couldn't help wondering if the various schools of thought on the nature of logical truth are comparably well-known. Perhaps not.

Of course, there are historical and sociological explanations for this, beginning with the way mathematical logic grew out of early explorations in the foundations of mathematics by such seminal figures as Cantor and

1. As we'll see, one way of responding to questions like these is to reject their presuppositions—for example, to deny that logic has a subject matter—but I hope these simple formulations serve to open the discussion and to illustrate the general type of question at issue. Also, though I speak of logical truth, no contrast with logical inference is intended: the two are treated here as more or less interchangeable (the inference from p to q is logically valid iff $p \rightarrow q$ is a logical truth). Finally, let me remark that no analysis of what counts as a logical truth is presupposed: the plan is just to inquire into the status of particular truths that as a matter of fact most people would classify as 'logical'.

2. I personally wouldn't endorse this identification (see [2011b]).

Frege. But given that this is, after all, the Association for Symbolic *Logic*, I thought I might presume upon your patience today, and instead of talking yet again about the philosophy of set theory,³ I might try to lay out a few of the central positions in the philosophy of elementary logic, something like the counterparts to the familiar Logicism, Intuitionism, Formalism, and Realism or Platonism in the philosophy of mathematics.⁴ This taxonomy will naturally reflect my personal predilections, and I reserve the right to give my own views the final word,⁵ but I hope these flaws can be forgiven in a retiring address. In any case, you stand forewarned.

Let's start by getting some markers out onto the table. First we can agree that the logical counterpart to Logicism isn't worth much—to say that logic is reducible to logic in whatever sense doesn't count as news—but we can ask for a counterpart to Realism in the philosophy of mathematics: a view according to which logical truths are objective, mind-independent facts. But what exactly are these logical facts facts about?, what is the subject matter of the science of logic?, what's the counterpart to sets, functions, and Hilbert spaces? If the answers are to run parallel with Platonism in the philosophy of mathematics, the subject matter should be some realm of abstracta, and the truths about these abstracta should be necessary, that is, not just contingent truths about the way this particular world happens to be, but necessary truths about the way any possible world must be. And sure enough, on at least one reading of Frege,⁶ he offers a position of this general stripe.

3. For this topic, see [2011b].

4. Viewing various disparate approaches to the nature of logical truth through this particular organizing lens inevitably involves some selective attention and even some occasional fudging (e.g., see footnote 6), but I hope the exercise is of some interest, nonetheless.

5. But only a brief word. For a more complete presentation, see [2007], Part III. [[Also Essays #8 and #9, [2014b].]]

6. I hope the specialists will forgive me for focusing, for rhetorical purposes, on only one of the various hotly debated approaches to Frege: the resolutely Platonistic line of interpretation championed, e.g., in Burge [1992]. Even on this reading, some fudging is required on necessity: according to Burge, though Frege doesn't employ the full modal notion of necessity, his insistence on the generality of logic, combined with the breadth of his abstract ontology, 'have some of the force and effect of necessary truths that purport to quantify over possible objects or possible worlds' (Burge [2000], p. 370). I include necessity—understood in terms of the intuitive heuristic 'true in all possible worlds'—as part of my characterization of Platonism because it's typically present in versions of the view in the philosophy of mathematics, and because it makes for the clearest contrast with other forms of Realism to be considered later.

Consider a simple inference: if a coin either came up heads or came up tails, and it didn't come up heads, then it must have come up tails.⁷ Frege tells us that a sentence like 'the coin came up heads' expresses a thought:

The thought, in itself imperceptible by the senses, gets clothed in the perceptible garb of a sentence, and thereby we are enabled to grasp it. We say a sentence *expresses* a thought. (Frege [1918], p. 328)

A thought, in turn, is an objective abstract object:

Thoughts are neither things in the external world nor ideas . . . A third realm must be recognized. Anything belonging to this realm has it in common with ideas that it cannot be perceived by the senses, but has it in common with things that it does not need an owner so as to belong to the contents of his consciousness. Thus for example the thought we have expressed in the Pythagorean Theorem is timelessly true, true independently of whether anyone takes it to be true. (Ibid., pp. 336–337)

Frege argues that thoughts pick out other denizens of the third realm—the truth-values, True or False—much as (the senses of) names pick out their referents,⁸ furthermore, words like 'or' or 'not' do similar duty for maps that take truth-values, or pairs of truth-values, to truth-values.⁹ So the story for our little inference would go: the simple sentences pick out truth-values; the logical words pick out truth-value-mappings; the properties of those mappings guarantee that if the premises are true, the conclusion must be true. This is an objective, timeless fact about the relations between various items in the third realm.¹⁰

7. As suggested at the end of footnote 1, I take this to be an uncontroversial case of what's commonly thought of as a logical truth or a logically valid inference. (Relevance logicians reject disjunctive syllogism as a general rule, but this particular case should satisfy any additional requirements of relevance.) By focusing on a simple example, I hope to sidestep debates over matters of classification—what counts as a logical consequence or a logical constant?—and to highlight instead the question of grounding—what makes this claim true, this inference reliable?

8. See, e.g., Frege [1892], pp. 156–159.

9. See, e.g., Frege [1891], p. 147, [1893], p. 217.

10. I've over-simplified the Fregean metaphysics here. It might be more accurate to say that for Frege, the inference is valid because a certain second-level concept—(((Fx or Gx) and not-Fx) implies Gx)—with three unsaturated slots (one for objects, two for first-level concepts) falls under the

Now it's worth asking why we should believe in the abstracta posited to form the subject matter of this account. One consideration that apparently moves Frege here is the fact that we manage to communicate:

A thought does not belong specially to the person who thinks it . . . everyone who grasps it encounters it in the same way, as the same thought. Otherwise two people would never attach the same thought to the same sentence, but each would have his own thought; and if, say, one person put $2 \times 2 = 4$ forward as true whilst another denied it, there would be no contradiction, because what was asserted by one would be different from what was rejected by the other. (Frege [1897], p. 233)

Interesting that Frege seems most concerned that 'there would simply be no common ground to fight on' (ibid., p. 234), but fight or no fight, the fundamental worry is that if thoughts were subjective, I couldn't communicate my thought to you, because your thought would inevitably be different from mine.¹¹

How we manage to communicate in language is obviously a big topic, but for what it's worth, it seems to me unlikely that our best scientific explanation of successful human language use will ultimately take this form. Another nagging worry is the well-known epistemological problem of how we human cognizers come to know about abstracta at all,¹² but this worry is premature: until the third realm earns its scientific stripes, we needn't concern ourselves with the question of access; and if it were to earn those stripes, the particulars of how it did so should form the basis of our epistemological theorizing. In any case, this is our Platonistic option: the subject matter of logic is abstract; its truths are necessary.¹³

third-level concept 'always instantiated, whatever goes in the unsaturated slots'. This isn't a claim about the coin or the properties of coming up heads or coming up tails; indeed, it would seem to belong among those whose 'justification is carried in their own contents' (Burge, [1998], pp. 337–338). In any case, the main point for our purposes is that the laws of logic 'are general laws which concern . . . "what is" . . . in the third realm' (Burge [1992], p. 313). (Burge is alluding to Frege [1918], p. 325: 'it is not a matter of what happens but of what is'.) Thanks to Jeremy Heis for his help on this point.

11. Cf. Burge [1992], p. 305: Frege 'claims, more than once, that the assumption of the relevant entities explains the intersubjective objectivity of science and communication'.

12. I have in mind here the logical counterpart to the famous Benacerraf problem (Benacerraf [1973]).

13. Again, this isn't exactly Frege's own position. See footnote 6.

The second famous school in the philosophy of mathematics is Intuitionism, which initially grew out of the Kantian idea that mathematical objects are constructed in pure intuition. For Kant, our cognition of the world involves two distinct faculties: the sensibility, which shapes the raw ‘matter’ of experience into spatiotemporal form; and the understanding, which brings the sensibility’s empirical intuitions under pure concepts, like the concept of an individual object, or the concept of causal connection.¹⁴ The sensibility’s forms of intuition and the understanding’s pure categories are present a priori—that is, we bring them to experience without the aid of any previous experience—and as a result, we can know a priori that the world we experience must be composed, for example, of spatiotemporal objects standing in causal relations.

Kant held that mathematics is the study of these forms of intuition—space and time—independent of any ‘matter’ of experience. In geometry, for example, we might construct a triangle, a three-sided figure, in our pure intuition, add a line through the apex parallel to the base, and reason from there that the three interior angles must add to two right angles. No amount of pondering the mere concept of a three-sided planar figure will yield this result, but the forms of intuition help shape the path of the construction, and those forms of intuition Kant took to be Euclidean. Alas, by the late 19th century, even before non-Euclidean ideas made their way into physical geometry, mathematical geometry had outgrown this Kantian picture: points at infinity and points marked by complex coordinates outstripped any simple spatial intuition.

The Intuitionists, beginning with Brouwer, saw this as refutation of Kant’s a priori intuition of space, but they hoped to found mathematics instead on our a priori intuition of time. Understood in orthodox Kantian terms, this would mean that mathematics is true of the world, and we can know this a priori, because the world we experience, the only world available to us, is

14. A terminological nicety: Kant does describe the pure category here as ‘cause and effect’ (A80/B106), but until it’s schematized (see below), no spatiotemporality is involved. As our current notion of causal connection would seem to involve both spatial and temporal elements, this usage can be misleading, so from now on I’ll refer to the pure category as ‘ground-consequent’, reserving ‘cause and effect’ for the schematized category.

partly shaped by our temporal form of intuition. This is one portion of Kant's transcendental Idealism.

To call this Idealism 'transcendental' is to distinguish it from the claim that spatiotemporality or causality are products of our particular human psychology, the sort of thing studied by empirical psychologists; for Kant, ordinary empirical science studies the world we experience, and that world has an objective spatiotemporal and causal structure, entirely independent of human psychology. From the ordinary empirical perspective, there's a clear distinction between appearance and reality, for example, between a rainbow and a raindrop:

We would certainly call a rainbow a mere appearance in a sun-shower, but would call this rain the thing in itself, and this is correct . . . in a merely [empirical] sense, as that which in universal experience and all different positions relative to the senses is always determined thus and not otherwise in intuition. (A45/B63)

But when we engage in Kant's critical philosophizing, when we inquire into our a priori knowledge of the world, we gain a new understanding of the empirical object:

If we consider this empirical object in general and . . . ask whether it . . . represents an object in itself, then the question of the relation of the representation to the object is transcendental [as opposed to empirical], and not only these drops are mere appearances, but even their round form, indeed even the space through which they fall are nothing in themselves, but only mere modifications or foundations of our sensible intuition. (A45–46/B63)

Thus Kant combines empirical Realism—from an empirical point of view, the raindrops are objectively real and the rainbow subjectively ideal—with transcendental Idealism—from a transcendental point of view, the raindrops

are partly formed by the structure of our cognition and the real thing as it is in itself is entirely unknown to us.

Now where does logical truth fit into this picture?¹⁵ Here Kant's interest is in discursive knowers, that is, beings who cognize the world using concepts, which essentially means any beings other than God. For discursive cognition to get off the ground, a discursive knower must have some concepts in place independently of experience, *a priori*, and these concepts, the 12 pure categories, correspond exactly to the 12 logical forms that a discursive judgment can take. So, for example, in order to cognize an individual object, a discursive knower needs the pure category of an object-with-properties, and this corresponds in turn to the logical form subject-predicate; to cognize one state of affairs as depending on another, he needs the category ground-consequent,¹⁶ which corresponds to the logical form if-then.

When these pure categories are schematized, that is, brought into accord with our peculiarly human forms of intuition, they become spatiotemporal-object and cause-and-effect, but the forms of sensibility of other discursive knowers needn't be the same as ours; still, the logical forms and pure categories govern the cognition of any discursive knower, whatever their particular form of intuition. Just as the world we humans experience must be spatiotemporal, the world any discursive knower experiences must be structured as the forms of judgment and pure categories dictate: for example, as objects-with-properties standing in ground-consequent relations. Viewed transcendently, the logical laws will be true in such a world because of the structuring the discursive knower brings to it,¹⁷ but viewed empirically, they are simply objective features of the world itself. In this way, much like the case of mathematics, logical truth is empirically real, but transcendently ideal.

So we have here an account of the nature of logical truth that runs parallel to the Kantian account of the nature of mathematics, except that logic is more general: it applies to the world as experienced by a more general class of knowers. Speaking empirically, as ordinary scientists, we say if the

15. Here again I opt for one among many interpretations, this time my own reading of Kant on logic. For a more complete discussion, see [2007], III.2. [[See also [2014b], chapter 1.]]

16. Recall footnote 14.

17. Strictly speaking, logic depends only on the forms of judgment, but a world structured by the forms of judgment *is* a world structured by the pure categories.

coin came up either heads or tails, and it didn't come up tails, it must have come up heads, and that this is true independently of us, because of the way the world is. Speaking transcendently, we say that the whole structure of objects like coins, with properties like landing heads or tails, and so on—all this is imposed by our discursive forms of judgment and pure categories. Viewed this way, logic is the study of our form of cognition, much as the intuitionist's mathematics is the study of our mental constructions. This is one form of Idealism in the philosophy of logic.

Notice, though, that this particular idealistic view of logic carries with it a strong Idealism about the ordinary physical world as well.¹⁸ Might there be a way of regarding logical truth as subjective, as somehow dependent on us, without sweeping the whole world up into the same basket? One straightforward way of doing this would be to regard logic as the study of the ways humans actually infer, as a branch of empirical psychology. Views of this sort were held, for example, by two distinct schools of thought in 19th-century Germany: the so-called 'scientific materialists', who traced our patterns of inference to physiology and from there to chemistry and physics,¹⁹ and the group of neo-Kantians who interpreted (or reinterpreted) Kant's transcendental psychology in terms of ordinary empirical psychology.²⁰ Nowadays I think most philosophers are persuaded by a simple consideration raised by Frege:²¹

[L]ogic is concerned . . . not with the question of how people think, but with the question of how they must think if they are not to miss the truth.
(Frege [1897], p. 250)

The view of logic being rejected here is usually called Psychologism.

18. Insofar as it follows Kant (or Husserl? See van Atten [2004], chapter 6), Intuitionism would also appear to involve a broader Idealism.

19. See Sluga [1980], chapter 1, or Gregory [1977].

20. See, e.g., Anderson [2005] for discussion.

21. Or, as Boole ([1854], II.6, p. 408) put it almost half a century earlier, 'The mathematical laws of reasoning are, properly speaking, the laws of *right* reasoning only'. Despite calling his book *The Laws of Thought*, and despite having introduced it as an investigation of 'the fundamental laws of those operations of the mind by which reasoning is performed' (ibid., p. 1), Boole recognizes in the end that what's at issue isn't how we in fact reason, but how we should reason.

Another way of taking logical truth to be dependent on us, one more popular today, is the idea that logical truth arises from the way we use language. In the words of A. J. Ayer:

[The] validity [of a logical law] depends solely on the definitions of the symbols it contains. (Ayer [1936], p. 78)

In drawing a logical conclusion, 'I am simply calling attention to the implications of a certain linguistic usage,' and thus

there is a sense in which [logical inferences] do give us new knowledge. They call attention to linguistic usages, of which we might otherwise not be conscious. (Ibid., pp. 79–80)

Another positivist, Hans Hahn, puts it this way:

If we remind ourselves of the way in which the designations 'red', 'not red', 'blue', 'red or blue', etc. are applied to objects, we can . . . assert . . . : everything to which both designations 'red or blue' and 'not red' are applied, is also designated as 'blue'—which is usually put more briefly: if a thing is red or blue and not red, then it is blue. (Hahn [1933], p. 156)

This depends only on 'the sense in which the logical words "not" and "or" are used' (ibid.). The same would obviously go for our example of the flipped coin. This might qualify as a sort of linguistic Idealism: as a community, we come together to establish and evolve patterns of shared usage, and along the way, we generate some logical truths.

Faced with this suggestion, it's hard not to worry that explanation has stopped too soon. What's to keep the Fregean from agreeing that logic reflects our linguistic practices, then going on to insist that those practices themselves arise from the properties of abstracta in the third realm?—whereupon the view collapses into Platonism. Or, on the other hand, an observer sympathetic to Psychologism might say instead that those linguistic practices are as they are strictly due to the particulars of human cognitive structure. Ayer explicitly considers a concern somewhere in this neighborhood:

It has . . . been suggested that my treatment of [logical] 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. (Ayer [1946], p. 16)

This thought prompts Ayer to make room for some further explanatory grounding for logical truths:

I allow that the usefulness of [logical] propositions is founded both on the empirical fact that certain symbols are used in the way that they are and *on the empirical fact that the symbols in question are successfully applied to our experience*. (Ibid., p. 17, emphasis mine)

Here Ayer traces the ground of logical truth to empirical facts about language usage *and* to the underlying facts that make such use effective; along these lines, we might call attention to the various physical facts that constrain the evolution of our language. But he steps back from the apparent consequence that logical truth rests on more than language by pointing out that logical statements don't 'describe these facts in the sense in which empirical propositions may describe the facts that verify them' (ibid.).

To sort this out, consider a simple case. Suppose I now declare to you that for the duration of this lecture, I'm going to use the name 'George' to refer to my left hand, and furthermore that I'm also going to use the name 'Edward' for the same purpose. I then assert that 'George is the same as Edward'. I assume we'd agree that this assertion is true, but what makes it true? Nothing much about my left hand really matters here, except for its being an individuated physical object suitable for naming; apart from that, what makes the thing true is those stipulations I just made, despite the fact that 'George is the same as Edward' doesn't describe those stipulations. So the question we've been asking—what makes logical truths true?—isn't the same as the question Ayer is asking—what do logical truths describe? If our question—what makes logical truths true?—is to be answered along Ayer's linguistic lines, it's hard to see that only facts directly dependent on us will be involved, that ordinary physical facts about the world we live in won't also serve to shape

our language. Come to that, we might ask our Psychologist about the role of such ordinary physical facts in the shaping of the psychological patterns that he takes to underlie logic. In both cases, we might begin to doubt that the full story would continue to count as a version of Idealism.

I think the divide between the question I'm asking and the question Ayer is asking can be clarified by looking at Ayer's answer to his question, and at the historical source of that answer. His question is: what do the statements of logic describe? His answer is: 'they do not . . . describe any facts at all' (Ayer [1946], p. 17):

'Either some ants are parasitic or none are' provides no information whatsoever about the behavior of ants, or, indeed, about any matter of fact. And this applies to all [logical] propositions. They none of them provide any information about any matter of fact. In other words, they are entirely devoid of factual content. (Ayer [1936], p. 79)

In a loose way, we might consider this a counterpart to Formalism in the philosophy of mathematics, the kind of Formalism that treats mathematical claims as meaningless, that denies mathematics is in the business of uncovering contentful truths. In both cases, we would meet the challenge of explaining the subject matter of our inquiry—of logic or of mathematics—by denying that it has any subject matter at all.

This idea for the case of logic came to Ayer—and to the logical positivists for whom his book served as a manifesto—from Wittgenstein, more specifically from Wittgenstein's *Tractatus Logico-Philosophicus*.²² This remarkable book can be thought of as beginning from a question that runs parallel to the one that inspired Kant: Kant asks 'what must the world be like that we can know it as we do?'; Wittgenstein asks 'what must the world be like that we can represent it, linguistically, as we do?' Kant thinks part of our knowledge of the world is a priori—geometry, for example; to explain this, as we've seen, he posits the sensibility's forms of intuition and the understanding's

22. Again specialists will note that the *Tractatus* has spawned a bewildering array of interpretations. The one sketched here is a fairly standard 'metaphysical' reading (see Stern [2003] for an overview of the various schools of thought). [[See also [2014b], chapter 3.]]

pure categories. Wittgenstein thinks that we must be able to understand the meanings of our sentences, what it would take to make them true or false, without knowing anything about how the world actually is; on analogy with Kant, he thinks part of our ability to represent the world is *a priori*, just as Kant thinks part of our knowledge of the world is *a priori*.

To explain his quota of *a priori*, Wittgenstein posits that a batch of simple objects exists necessarily, in all possible worlds, that each simple object has a logical form, which determines the ways in which it can and can't combine with other simple objects into atomic facts, and that all the facts of our actual world are fully determined by which of these atomic facts do and don't obtain here. We're able to represent the world, he continues, because our language consists of simple names with the same logical form as the simple objects—that is, with exactly the same possibilities of combination—so we can represent the world as being *thus-and-so* by combining our simple names *thus-and-so*, regardless of whether or not the world really is *thus-and-so*. The actual world doesn't play any role in determining the meanings of our atomic sentences; its only job is to settle whether they're true or false.

Since the atomic facts are entirely independent of each other—all combinations of individual atomic facts are possible—we can describe a complete state of the world by what's essentially a truth-table row, a stipulation of which atomic facts obtain and which don't. The meaning of a complex statement, then, can be understood as a specification of which truth-table rows make it true and which make it false; to put it in Ayer's terms, the complex statement describes the world as being in one of the states that gets the value True and not in any of the states that get the value False. So what's the meaning of a logical truth? It draws no such division in the space of logical possibilities; it tells us nothing about how the world is or isn't. And now we've arrived at Ayer's view.

But for Wittgenstein this is hardly the end of the story. The key here is his important distinction between saying and showing:

A proposition *shows* its sense. A proposition *shows* how things stand *if* it is true. And it *says that* they do so stand. (Wittgenstein [1921], 4.022)

In other words, a statement shows us what it would take for it to be true or for it to be false; this we grasp independently of how the world happens to be, and in particular, independently of whether or not the given statement is in fact true or false. Furthermore, this meaning that is shown to us is not what the statements says, not what Ayer is asking about. When we reach the case of logical truths and logical falsehoods, they say nothing:

Among the possible groups of truth-conditions there are two extreme cases. In one of these cases the proposition is true for all the truth-possibilities of the elementary propositions . . . In the second case the proposition is false for all the truth possibilities . . . Propositions show what they say: tautologies and contradictions show that they say nothing. (Ibid., 4.46–4.461)

By focusing exclusively on what logical claims *say*, Ayer ignores the (at least) equally important matter of what they *show*:

The fact that the propositions of logic are tautologies *shows* the formal—logical—properties of language and the world. (Ibid., 6.12)

What logical truths show us is the features of that logical structure shared by our language and the world, the very shared structure that makes representation possible.

So Wittgenstein's answer to my question, as opposed to Ayer's, is that logic is grounded in the underlying structure of language and the world: the structure the world has by virtue of its being one of the possible ways the necessarily existing objects can combine into atomic facts, the structure language must have if it is to represent how those objects happen to be combined. This is a bold and substantive thesis, one that prompts us to ask whether this structure is shared because language copies the world or because the world copies language. The latter would follow up on our Kantian analogy: the logical structure of our language (the way our simple names can combine, and so on) would partly constitute the world-as-represented, much as Kant's forms and categories partly constitute the world-as-experienced; Wittgenstein's account of linguistic representation would replace

Kant's transcendental psychology. Here we have a new form of Idealism, a linguistic idealism quite different from Ayer's.²³

The other option is that the world comes first, that language copies the world, or more specifically, that language copies the structure that our particular world shares with all possible worlds. This is a kind of Realism: logic reflects an objective structure. Furthermore, logical truths are necessary. Whether or not this adds up to Platonism, as we're using the term here, depends on whether or not the objective structure in question belongs to something abstract. If Wittgenstein's objects were ordinary things, the answer would be no, but they aren't: the existence of an ordinary physical object is a contingent fact, one among the various ways the necessarily existing simple objects can combine. The issue turns on what these simple objects are, but alas Wittgenstein gives us no clue, so best to leave the matter here: on this second reading, Wittgenstein gives us a Realism about logical truth with many of the markers of Platonism.

We'll consider some clearly non-Platonistic forms of Realism in a moment, but first let's look at a more formalistic, more syntactic type of position, the position that in fact provided the historical bridge between Wittgenstein and Ayer. Wittgenstein's idea that logic is somehow empty held great appeal for Rudolf Carnap:

For me personally, Wittgenstein was perhaps the philosopher who, besides Russell and Frege, had the greatest influence on my thinking. The most important insight I gained from his work was the conception that . . . 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. (Carnap [1963], p. 25)

Here Carnap might seem to be missing the way logical truths can show features of the necessary structure of the world, but this aspect of the *Tractarian* view depends on a Wittgensteinian thesis that Carnap rejects:

23. On analogy with 'transcendental idealism', Stenius [1960], p. 220, suggests the term 'transcendental lingualism'.

Wittgenstein . . . overlooked the fact that there is a multiplicity of possible languages. Wittgenstein . . . speaks continually of ‘the’ language. (Carnap [1934], p. 245)

‘Overlooked’ may be a bit mild here: the notion of exactly one logical form that language and the world share is central to Wittgenstein’s answer to his guiding question, ‘how are we able to represent the world?’

In any case, this isn’t Carnap’s question; his goal is quite different:

Philosophy is to be replaced by the logic of science—that is to say, by the logical analysis of the concepts and sentences of the sciences, for *the logic of science is nothing other than the logical syntax of the language of science*. (Carnap [1934], p. xiii)

In pursuit of this goal, Carnap argues that a wide range of languages can be described in his formal syntax, and that the choice between these languages is entirely unconstrained:

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. (Ibid., p. 52)

This is the famous Principle of Tolerance: we may choose one logic over another for pragmatic reasons, but that logic is no more true or false than any other; issues of truth or falsity are only properly raised after a language, with its logic, has been adopted, and they are then resolved by the means specified in the conventions of that chosen language.²⁴ Here Carnap gives us something akin to those stipulations I floated earlier about names for my left hand: a language, for Carnap, has its own formation rules (logical axioms or fundamental assumptions of some scientific theory) and transformation

24. Here again I beg forgiveness from the specialists: this simple reading of Carnap, encouraged by Carnap [1950], adopted by Quine, has been vigorously challenged in recent years. Many now deny that Carnap is offering an account of the ground of logical truth, so these more subtle versions of Carnap are beside the point for our purposes. (See [2007], I.5 for discussion and references.)

rules (rules of inference, or more generally, rules saying which statements in the language count as evidence for which other statements in the language); nothing factual, nothing beyond convenience, constrains our choice of these rules; it's a purely conventional matter.

We can now see how Ayer's position arises from Wittgenstein's logic-is-empty by incorporating Carnap's idea that many languages are possible:

It is perfectly conceivable that we should have employed different linguistic conventions from those which we actually do employ. But whatever these conventions might be, the tautologies in which we recorded them would always be necessary. (Ayer [1936], p. 84)

The difference between Carnap and Ayer is that Carnap has in mind our deciding to adopt a new language, perhaps a formal language, while Ayer is imagining our natural language as the result of many conventional decisions now mostly lost to history. The fact that the actual evolution of natural language appears to be guided by physical and psychological factors other than explicit, arbitrary choices was what led us to our concerns about Ayer's position. Carnap's focus on explicit, even formal conventions promises to sidestep this problem.²⁵

Unfortunately, this Conventionalism about logical truth faces another daunting obstacle, one apparently first described by Lewis Carroll in his story 'What the Tortoise said to Achilles' (Carroll [1895]).²⁶ Quine later reworked the objection in more sober philosophical terms:

The linguistic doctrine of logical truth is sometimes expressed by saying that such truths are true by linguistic convention . . . it is impossible in principle . . . to get even the most elementary part of logic exclusively by the explicit application of conventions stated in advance. The difficulty is the vicious regress, familiar from Lewis Carroll, which I have elaborated

25. In practice, it's not clear this promise would be realized. If we choose these conventions over those on pragmatic grounds, the question inevitably arises: why are these conventions more effective than those? The answer to this question might well return us to the very physical and psychological factors that emerge for Ayer. I set this aside to highlight the new concern in the next paragraph.

26. I give a brief reconstruction in [2007], p. 204.

elsewhere [in Quine [1936]]. Briefly the point is that the logical truths, being infinite in number, must be given by general conventions rather than singly; and logic is needed then to begin with . . . in order to apply the general conventions to individual cases. (Quine [1954], p. 115)

There was no such problem when I stipulated the two names for my left hand, nor would there be if I stipulated the truth of all the axioms of ZFC, assuming logic is functioning in the background, but when we try to stipulate the truth of logic itself, we find that our explicit conventions must be general, and then that these general conventions are without their intended force unless logic is in fact already available to oversee the derivation of particular logical truths from the generalities. So it appears that the suggestion that ‘logic is empty’ either masks a deeper, more complex account of what makes logic true (Ayer) or appeals to a straightforward notion of explicit convention that turns out not to be viable for the case of logic (Carnap).

Let’s pause a moment to take stock. We’ve looked at an array of ideas on the nature of logical truth, all analogous, more or less, to familiar positions in the philosophy of mathematics. Frege presents a Platonistic view: logical truths report the objective features of abstracta in the third realm. Along the lines of Intuitionism or of Idealism more broadly, Kant give us his characteristic combination of transcendental Idealism—logic is true because of the structure of discursive cognition—and empirical Realism—logic is true because of the objective structure of the world. The Psychologists take a more straightforwardly idealistic stance: logical laws are simply the most general patterns of our thinking. Ayer attempts another form of Idealism—logical laws arise from the way we use language—combined with a more formalistic insistence that logical statements are empty. The source of that formalistic idea is Wittgenstein’s *Tractatus*, but there it comes combined with an ambitious account of the ground of logical truth that can be understood either as a form of linguistic Idealism parallel to Kant’s transcendental Idealism or as a form of Realism that leans toward Platonism. A purer analogue to Formalism is Carnap’s Conventionalism: logical truth arises from explicit syntactic conventions.

The last idea I’d like to explore is the one set aside a moment ago, namely, a non-Platonistic form of Realism. The closest analogue in the philosophy

of mathematics to what I have in mind here is a not very popular position usually called empiricism and associated in the first instance with John Stuart Mill:

Three pebbles in two separate parcels, and three pebbles in one parcel, do not make the same impression on our senses; and the assertion that the very same pebbles may by an alteration of place and arrangement be made to produce either the one set of sensations or the other . . . is a truth known to us by early and constant experience—an inductive truth . . . The fundamental truths of [the science of Numbers] all rest on the evidence of sense; they are proved by showing to our eyes and our fingers that any given number of objects, ten balls, for example, may by separation and rearrangement exhibit to our senses all the different sets of numbers the sum of which is equal to ten. (Mill [1843], II.VI.2, pp. 168–169)

Thus arithmetic is a straightforward part of our empirical theory of the world. Mill takes arithmetic to be ‘necessary’, but by this he means ‘no more than certainty’; he denies any ‘claim to the character of necessity in any sense beyond this, as implying an evidence independent of and superior to observation and experience’ (Mill [1843], II.VI.1, pp. 164–165). This is Realism—arithmetic reports objective truths—but not Platonism—its subject matter isn’t abstract, but the ordinary physical world, and its truths are ordinary contingencies.

This view of arithmetic was ridiculed by Frege as ‘Mill’s piles of pebbles and gingersnaps’ (Frege [1884], §27, p. 38):

What a mercy that . . . not everything in the world is nailed down; for if it were, we should not be able to bring off this separation, and $2 + 1$ would not be 3! (Frege [1884], §7, p. 9)

And of course it becomes considerably less plausible as an account of modern infinitary mathematics. But what about elementary logical truth? Though Mill isn’t entirely consistent on these matters, he apparently doesn’t embrace

an analogous position for logic; instead, he treats examples like our inference about the coins as ‘merely apparent’:²⁷

[W]e were not proving the proposition, but only appealing to another mode of wording it. (Mill [1843], II.I.2, p. 104)

Still, hints of a simple inductive Realism can be found in Bolzano:

The only reason why we are so certain that the rules [of syllogistic logic] are valid is because they have been confirmed in thousands of arguments in which we have applied them. (Bolzano [1837], 315.4, p. 354)

And hints of a non-abstract subject matter turn up in Russell:²⁸

[L]ogic is concerned with the real world just as truly as zoology, though with its more abstract and general features. (Russell [1919], p. 169)

In any case, whatever its lineage, the non-Platonic Realist’s notion that logic reflects contingent features of the physical world surely deserves a place in our taxonomy.

Unfortunately, the simple inductivism of the pseudo-Millian version is problematic. To see this, consider how we’re supposed to know that our coin must have come up tails: we know it came up either heads or tails, we know it didn’t come up heads, and we know from experience that disjunctive syllogisms are valid. Two problems arise for this appeal to disjunctive syllogism (DS): how do we know that the various previous cases we’ve observed count as evidence for DS unless we know the logical relation between a generalization and its instances?, and how do we know that DS is relevant to the present case unless we recognize that same relation? For that matter, how would we know that the coin having come up heads would be a violation of

27. See, e.g., Jackson [1941].

28. In fact, this may well not be what Russell has in mind; he might simply be insisting that the world includes the abstracta (like propositional functions) that constitute the subject matter of logic just as robustly as it does the animals that constitute the subject matter of zoology. (My thanks to Peter Hylton for help with this point.) On the other hand, see Klement [2010].

DS unless we knew the law of non-contradiction(LNC)? The general point is that our simple methods of inductive confirmation presuppose many logical relations between statements, so they can't be used to give us knowledge of those logical relations in the first place.

Quine presents a more sophisticated but still empirical picture of logical knowledge, based on his holism:

The totality of our so-called knowledge or beliefs, from the most casual matters of geography and history to the profoundest laws of atomic physics or even of pure mathematics and logic, is a man-made fabric which impinges on experience only along the edges. (Quine [1951], p. 42)

Quine's thought here is that a typical statement has consequences for experience only when taken in combination with a fairly large collection of auxiliaries, so our claims 'face the tribunal of experience not individually but only as a corporate body' (*ibid.*, p. 41). It might seem that a simple perceptual claim would escape this generalization, but Quine insists that:

Even a statement very close to the periphery can be held true in the face of recalcitrant experience by pleading hallucination or by amending certain statements of the kind called logical laws. (*Ibid.*, p. 43)

Here the laws of logic are part of the web like everything else:

Having reevaluated one statement we must reevaluate some others, which may be statements logically connected with the first or may be the statements of logical connections themselves. (*Ibid.*, p. 42)

As an example of a possible revision of logic, he notes that:

Revision even of the . . . law of the excluded middle has been proposed as a means of simplifying quantum mechanics; and what difference is there in principle between such a shift and the shift whereby Kepler superseded Ptolemy, or Einstein Newton, or Darwin Aristotle? (*Ibid.*, p. 43)

Logical truths, then, are known empirically and fallibly along with the rest. We tend to think of logic as fixed and certain only because we're reluctant to revise it, given the widespread changes this would demand, and would do so only as a last resort.

For all its sophistication, this Quinean empiricism faces a difficulty not unlike the one described a moment ago for the pseudo-Millian view: it's fairly easy to imagine how the web of belief might evolve to leave out the law of the excluded middle (LEM)—the Intuitionists do without it, after all, as do we ordinary folk when we judge that Joe isn't quite bald but then again isn't quite not-bald, either—but it isn't easy to imagine how the web would fare without, say, the law of non-contradiction. Without LNC, what would even count as a recalcitrant experience? The problem is that the machinery of web maintenance involves at least some logic, so it's hard to see how that much logic can be counted on a par with the rest of the web.²⁹ Still, the idea that the progress of science could lead us to a different logic just as it once led us to a different geometry, that we could be led to revise our logic on empirical grounds—this idea doesn't depend on a strict or problematic holism.

Just such a picture is presented by Putnam with a focus on the case of quantum mechanics.³⁰ Experimental evidence there confirms a wide range of oddities: electrons seem not to behave as individuated objects with continuous trajectories; they apparently don't enjoy a definite position and a definite momentum, a definite vertical spin and a definite horizontal spin, at the same time; their behaviors can be intertwined in ways that violate the usual laws of causality; and so on. Quine's passing remark about LEM apparently refers to an early proposal of Reichenbach that a three-valued logic be employed:³¹ for example, the claim that a given particle is in a certain position could sometimes take the third truth-value, could sometimes be neither true nor false, but indeterminate. This proposal is now generally regarded as 'a nonstarter',³² but Putnam has a more sophisticated quantum logic in mind.

29. Given that it was Quine who lodged the Lewis Carroll objection so forcefully against Carnap, there's some irony in the fact that both our pseudo-Mill and Quine are subject to a more general form of the same problem: some logic has to be presupposed for their accounts of logical truth to be implemented.

30. See Putnam [1968].

31. See Reichenbach [1944].

32. Gibbins [1987], p. 127. For more on quantum logic, with references, see [2007], III.4, III.6.

To see how this works, let u and d be ‘the electron has vertical spin up’ and ‘the electron has vertical spin down’, and let l and r be the corresponding claims for horizontal spin left and right. Then it’s straightforwardly true that— $(u$ or $d)$ and $(l$ or $r)$ —but it doesn’t follow that one of the four possibilities— $(u$ and $l)$, $(u$ and $r)$, $(d$ and $l)$, $(d$ and $r)$ —must hold, because each of these stipulates a determinate vertical spin and a determinate horizontal spin at the same time, just what quantum indeterminacy denies. The quantum logic advocated by Putnam invalidates this distributive law and thus removes the bad inference. A neat trick.

Putnam’s hope is that this alternative logic will remove the quantum anomalies and at the same time reinstate our ordinary ways of thinking about objects and their properties and relations at the micro-level; in this way, the distributive law would be empirically falsified and a non-standard quantum logic empirically confirmed.³³ Alas, this happy outcome is not achieved. Though the absence of one distributive law blocks the derivation of one quantum conundrum, Peter Gibbins has shown that another distributive law is retained and that it permits a very similar confounding derivation.³⁴ Furthermore, the only bit of common sense attained is that the electron is determinately either spin up or spin down, despite being neither determinately up nor determinately down. But this we knew. Our ordinary ways of thinking demand an objective fact of the matter as to which state the electron is in.

Of course we shouldn’t insist that a viable philosophy of logic must also solve the problems of quantum mechanics—that’s a lot to ask—so the failure of this particular quantum logic needn’t undercut the appeal of a non-Platonistic Realism. We’ve seen that Quine and Putnam, like our pseudo-Mill, take logic to be one part of our ordinary empirical theory of

33. Putnam doesn’t hold that all logical laws are subject to direct empirical disconfirmation: he distinguishes between those whose falsity we can conceive—like LEM or the distributive law—and those whose falsity we can’t conceive—like LNC. This doesn’t mean that LNC couldn’t turn out to be false, but that it ‘cannot be overthrown merely by observations, but only by thinking of a whole body of alternative theory as well’ (Putnam [1995], p. 272). Putnam ([1968], pp. 189–190) also seems to toy with the idea that some changes of logic would be changes of meaning while others would be changes of theory. Of course any such appeal to meaning raises the question we’ve harped on before: what inclines us toward these meanings rather than those?

34. Gibbins [1987], pp. 147–151.

the physical world. It's striking, though, that apart from Russell's 'more abstract and general features', little is said about what aspects of the world are reflected in the logical parts of our theory. This isn't so surprising for Quine, because his holism dictates that no individual claim answers to the world by itself, but only as part of a 'corporate body', and perhaps Putnam is inclined to follow Quine on this. Those of us who reject this holism, who hold that our individual scientific claims are answerable to different parts of the world, would like a better sense of those 'abstract and general features'. I'll close here with a quick review of my own attempt to do this and of the resulting account of logical truth.³⁵

To see how this goes, recall Kant's combination of transcendental Idealism with empirical Realism: viewed transcendently, logical structure is imposed on the world by our discursive modes of thought; viewed empirically, the world simply displays those structures as a matter of objective fact. Leaving the transcendental bit aside for the moment, what's striking here is that Kant is quite specific about what those objective structures are: they're listed in his Table of Judgments and the corresponding Table of Pure Categories. Just how Kant gets from the bare notion of discursive cognition—that is, cognition through concepts—to the conclusion that all such cognition must take these 12 particular forms is a matter much debated by Kant scholars and almost universally admitted not to be one of the great man's finest hours; and of course the Tables rely on accounts of logical form now superseded by Fregean innovations. But let's see what we can salvage here.

In Kant's version, empirical Realism holds that the world is made up of individual objects enjoying various properties and that various states of these objects stand in dependency relations to other states of these objects. These features of the world reflect the forms of judgment ('subject-predicate' and 'if-then') and the corresponding pure categories ('object-with-properties' and 'ground-consequent').³⁶ Fregean updating replaces 'subject-predicate' with 'function-and-arguments' and 'object-with-properties' with

35. See footnote 5.

36. These are only 2 of the 12 forms of judgment and corresponding pure categories, but I propose to ignore the rest. Also, the schematized categories aren't included here because what matters for our purposes is the a priori features of the world of the discursive knower in general, not the human knower in particular.

‘objects-standing-in-relations’. So post-Fregean empirical Realism would hold that the world consists of objects enjoying various interrelations (including the degenerate case of an object enjoying a property) and that various states of these objects (now including more complex relational states) stand in dependency relations, where this structuring now reflects the new post-Fregean form (‘function-and-arguments’) and the corresponding pure category (‘objects-in-relations’).

But what about the transcendental part of Kant’s story? If we try to import it more or less intact into our ordinary empirical context, it becomes a strong version of Psychologism: the logical structure of the world is a reflection of our cognitive mechanisms, not an objective feature of the world itself. But if the world does in fact have this structure, as our updated empirical Realism asserts, then it wouldn’t be surprising that our cognitive mechanisms have evolved to detect it, and in fact, this is just what contemporary developmental psychology suggests.³⁷ What makes this Psychologism unattractive is its second clause, the denial that these psychological structures reflect objective features of the world itself. If we drop this problematic add-on, a first pass at our updated, non-Platonistic Realism would go like this: logical truths are true because the world is made up of objects enjoying various interrelations with dependencies between them, and we tend to believe some of the simpler of these truths because human cognition has been tuned by evolution to detect these very features.

So far so good, but is it true? Is the world so neatly structured? At first blush, it seems the answer is obviously yes: this sort of structuring is so obvious that we hardly notice it, but there it was, for example, in the background of my stipulations about names for my left hand (in the assumption that it’s an individual object, suitable for naming); for that matter, it even comes built into our intuitive notion of a ‘possible world’. Unfortunately, if we’ve learned anything from our brush with quantum mechanics, it’s that the ‘objects’ of the micro-world don’t behave according to these simple patterns, don’t act as

37. Just to be clear, the idea isn’t that these cognitive structures have evolved under pressure from natural selection, therefore they can be counted on to deliver truths. (This is a well-known fallacy: even if our cognitive structures have evolved to be fitness-enhancing, what’s fitness-enhancing to believe may not be true.) Here empirical Realism comes first and then figures in our accounts of cognitive development and evolution.

our most fundamental conceptualizations lead us to expect. If logical truth is in fact grounded in such structuring, then our logic should break down in the micro-world, which, as we've seen, it does. But the mysteries of the quantum world don't change the fact that much; indeed most of the macro-world *is* structured in accord with our ordinary notions: there are individual apples; the presence of this apple tree is a consequence of the planting of that seed. And the fact that much of the macro-world is this way, combined with the fact that we evolved in direct contact with that macro-world, explains why our basic modes of cognition are tuned to these particular structures. So our updated and naturalized Kantian position is that much of the world has these features, that logic applies to a situation insofar as it does have those features, and that our cognitive machinery has evolved to detect those features.

Thus we arrive at a non-Platonistic Realism: logical laws are true in any situation with the right physical structuring; their truth is contingent on the presence of that structuring. Moreover, on the epistemological side, we tend to believe the laws of logic independently of any experience because of our hardwiring, so we know them in a sense *a priori*, and we tend to think of them as necessary;—that is, we tend to build them into our very idea of a 'possible world'³⁸—and all this happens despite the fact that they wouldn't be true if the world were different and in fact don't seem to hold in the actual micro-world. Naturally our language will reflect these fundamental modes of thought, but their grounding goes deeper than that, into the structure of the world we live in: it isn't simply a matter of linguistic convention; we adopt the meanings we do for a reason. Also, though Kant would say that no discursive cognizer can think any other way, it seems best to be more circumspect: because of our hardwiring, it's very difficult for us to think any other way—as demonstrated by our inability to date to come up with a viable interpretation of quantum mechanics—but there's no clear reason to suppose that this is impossible in principle.

38. Descartes espoused a historical precursor to this view: the laws of logic are contingent—God could have made them other than they are—but, having made them as they are, He made us so that we believe them to be not only true but necessarily true. See his [1991], pp. 23–26 (letters to Mersenne in 1630), 235 (letter to Mesland in 1644), 358–359 (letter to Arnauld in 1648), 363 (letter to More in 1649).

Finally, let me note that the physical structuring of the world doesn't seem to me to validate the full range of classical logic; for example, there are tadpoles and there are frogs, but there are also borderline cases; $2 + 2$ being 4 isn't grounded on the fact that nothing red all over is also green all over. This means that LEM and the material conditional aren't part of this more rudimentary empirical logic; rather, they appear as idealizations introduced into that logic for good reasons. From this perspective, it isn't enough for defenders of various deviant logics to point out that a given aspect of classical logic is an idealization; what's needed is an argument that this idealization is ineffective or misleading, either in particular circumstances or in general. Unless this case is made, and a more effective, less misleading alternative offered, we do best to continue to use classical logic, but to do so with care, just as we use idealized scientific theories with care, recognizing that idealizations can lead us astray: if we ignore friction too broadly, we can't explain how we manage to walk; if we apply classical reasoning too stubbornly, a vague predicate can lead us into a sorites puzzle.

Thus ends my survey. We've touched on a broad range of historical and contemporary positions on the ground of logical truth, classified very roughly by their resemblance, in the first instance, to well-known schools of thought in the philosophy of mathematics—Platonism, Intuitionism, Formalism—and in the second, to an implausible position in the philosophy of mathematics—non-Platonistic Realism. After Frege, Kant, the Psychologists, Ayer, the *Tractatus*, and Carnap, this last group included the simple inductivism of our pseudo-Mill, Quine's holism, Putnam's quantum logic, and my own updated and naturalized version of Kant. I hope this leaves you with some sense of the richness of the subject, and perhaps also of its relevance and importance to the intellectual mission of the Association for Symbolic Logic.³⁹

39. My thanks to David Malament, (the late) Ruth Marcus, Patricia Marino, Erich Reck, Kyle Stanford, the audience at the ASL Annual North American Meeting in Berkeley, March 2011, and an anonymous referee for the *BSL* for their comments on earlier drafts. I'm also grateful to Waldemar Rohloff for many helpful discussions of Kant over the years.

A Second Philosophy of Logic

What's hidden in my hand is either an ordinary dime or a foreign coin of a type I've never seen. (I drew it blindfolded from a bin filled with just these two types of objects.) It's not a dime. (I can tell by the feel of it.) Then, obviously, it *must* be a foreign coin! But what makes this so?

It's common to take this query as standing in for more general questions about logic—what makes logical inference reliable? what is the ground of logical truth?—and common, also, to regard these questions as properly philosophical, to be answered by appeal to distinctively philosophical theories of abstracta, possible worlds, concepts, meanings, and the like. What I'd like to do here is step back from this hard-won wisdom and try to address the simple question afresh, without presumptions about what constitutes 'logic' or even 'philosophy'. The thought is to treat inquiries about reliability of the coin inference and others like it as perfectly ordinary questions, in search of perfectly ordinary answers, and to see where this innocent approach may lead.

To clarify what I have in mind here, let me introduce an unassuming inquirer called the Second Philosopher, interested in all aspects of the world and our place in it.¹ She begins her investigations with everyday perceptions, then gradually develops more sophisticated approaches to observation and experimentation that expand her understanding and sometimes serve to correct her initial beliefs; eventually she begins to form and test hypotheses, and to engage in mature theory-formation and confirmation; along the way, she

1. The Second Philosopher is introduced in [2007], and her views on logic detailed in Part III of that book. The discussion here reworks and condenses the presentation there (see also [2014b]). [[The Second Philosopher is also the Plain Inquirer of [2017a].]]

finds the need for, and pursues, first arithmetic and geometry, then analysis and even pure mathematics;² and in all this, she often pauses to reflect on the methods she's using, to assess their effectiveness and improve them as she goes. When I propose to treat the question of the reliability of the coin inference as an ordinary question, I have in mind to examine it from the Second Philosopher's point of view. She holds no prior convictions about the nature of the question; she sees it simply as another of her straightforward questions about the world and her investigations of it.

The first thing she's likely to notice is that neither the reliability of the coin inference nor the truth of the corresponding if-then statement³ depends on any details of the physical composition of the item in her hand or the particular properties that characterize dimes as opposed to other coins. She quickly discerns that what's relevant is entirely independent of all but the most general structural features of the situation: an object with one or the other of two properties that lacks one must have the other. In her characteristic way, she goes on to systematize this observation—for any object a and any properties, P and Q , if Qa -or- Pa and not- Qa , then Pa —and from there to develop a broader theory of forms that yield such highly general forms of truth and reliable inference. In this way, she's led to consider any situation that consists of objects that enjoy or fail to enjoy various properties, that stand and don't stand in various relations; she explores conjunctions and disjunctions of these, and their failures as well; she appreciates that one situation involving these objects and their interrelations can depend on another; and eventually, following Frege, she happens on the notion that a property or relation can hold for at least one object, or even universally—suppose she dubs this sort of thing a 'formal structure'.⁴

Given her understanding of the real-world situations she's out to describe in these very general, formal terms, she sees no reason to suppose that every object has precise boundaries—is this particular loose hair part of the cat or not?—or that every property (or relation) must determinately hold or fail to hold of each object (or objects)—is this growing tadpole now a frog or

2. For more on the Second Philosopher's approach to mathematics, see [2011b]. [[See also Essays ##9–11.]]

3. I won't distinguish between these, except in the vicinity of footnote 5.

4. In [2007] and [2014b], this is called KF-structure, named for Kant and Frege.

not? She appreciates that borderline cases are common and fully determinate properties (or relations) rare. Thinking along these lines, she's led to something like a Kleene or Łukasiewicz three-valued system: for a given object (or objects), a property (or relation) might hold, fail, or be indeterminate; not-(...) obtains if (...) fails and is otherwise indeterminate; (...) -and-(...) obtains if both (...) and (...) obtain, fails if one of them fails, and is otherwise indeterminate; and so on through the obvious clauses for (...) -or-(...), (there is an x , ... x ...) and (for all x , ... x ...). A formal structure of this sort validates many of the familiar inference patterns—for example, the introduction and elimination rules for 'not', 'and', 'or', 'for all', and 'there exists'; the DeMorgan equivalences; and the distributive laws—but the gaps produce failures of the laws of excluded middle and non-contradiction (if p is indeterminate, so are p -or-not- p and not-(p -and-not- p)).⁵ The subtleties of the Second Philosopher's dependency relation undercut many of the familiar equivalences: not-(the rose is red)-or- $2 + 2 = 4$, but $2 + 2$ doesn't equal 4 *because* the rose is red. Fortunately, modus ponens survives: when both (q depends on p) and p obtain, q can't fail or be indeterminate. Suppose the Second Philosopher now codifies these features of her formal structures into a collection of inference patterns; coining a new term, she calls this 'rudimentary logic' (though without any preconceptions about the term 'logic'). She takes herself to have shown that this rudimentary logic is satisfied in any situation with formal structure.

This is a considerable advance, but it remains abstract: what's been shown is that rudimentary logic is reliable, assuming the presence of formal structure. Common sense clearly suggests that our actual world does contain objects with properties, standing in relations, with dependencies, but the Second Philosopher has learned from experience that common sense is fallible, and she routinely subjects its deliverances to careful scrutiny. What she finds in this case is, for example, that the region of space occupied

5. Here, briefly, the distinction between logical truths and valid inferences matters, because the gaps rule out all of the former. Inferences often survive because gaps are ruled out when the premises are taken to obtain.

by what we take to be an ordinary physical object like the coin does differ markedly from its surroundings: it contains a more dense and tightly organized collection of molecules; the atoms in those molecules are of different elements; the contents of that collection are bound together by various forces that tend to keep it moving as a group; other forces make the region relatively impenetrable; and so on. Similarly, she confirms that objects have properties, stand in relations, and that situations involving them exhibit dependencies.

Now it must be admitted that there are those who would disagree, who would question the existence of ordinary objects, beginning with Eddington and his famous two tables:

One of them is familiar to me from my earliest years . . . It has extention; it is comparatively permanent; it is coloured; above all it is *substantial*. By substantial I do not mean merely that it does not collapse when I lean upon it; I mean that it is constituted of 'substance'. (Eddington [1928], p. ix)

Table No. 2 is my scientific table . . . It . . . is mostly emptiness. Sparsely scattered in that emptiness are numerous electric charges rushing about with great speed; but their combined bulk amounts to less than a billionth of the bulk of the table itself. Notwithstanding its strange construction it turns out to be an entirely efficient table. It supports my writing paper as satisfactorily as table No. 1; for when I lay the paper on it the little electric particles with their headlong speed keep on hitting the underside, so that the paper is maintained in shuttlecock fashion at a nearly steady level. If I lean upon this table I shall not go through; or, to be strictly accurate, the chance of my scientific elbow going through my scientific table is so excessively small that it can be neglected in practical life. (Ibid., p. x)

So far, the Second Philosopher need have no quarrel; Eddington can be understood as putting poetically what she would put more prosaically: science has taught us some surprising things about the table, its properties and behaviors.

But this isn't what Eddington believes:

Modern physics has by delicate test and remorseless logic assured me that my second scientific table is the only one which is really there. (Eddington [1928], p. xii)

The Second Philosopher naturally wonders why this should be so, why the so-called ‘scientific table’ isn’t just a more accurate and complete description of the ordinary table.⁶ In fact, it turns out that ‘substance’ in Eddington’s description of table No. 1 is a loaded term:

It [is] the intrinsic nature of substance to occupy space to the exclusion of other substance. (Ibid., p. xii)

There is a vast difference between my scientific table with its substance (if any) thinly scattered in specks in a region mostly empty and the table of everyday conception which we regard as the type of solid reality . . . It makes all the difference in the world whether the paper before me is poised as it were on a swarm of flies . . . or whether it is supported because there is substance below it. (Ibid., pp. xi–xii)

Here Eddington appears to think that being composed of something like continuous matter is essential to table No. 1, that one couldn’t come to realize that its supporting the paper or resisting my elbow arise very differently than I might have at first imagined—that one couldn’t come to realize this, that is, without also coming to realize that there is no such thing as table No. 1. But why should this be so? Why should our initial conceptualization be binding in this way? For that matter, is it even clear that our initial conceptualization includes any account at all of how and why the table supports paper or resists elbows? The Second Philosopher sees no reason to retract her belief in ordinary macro-objects.⁷

6. Some writers reject the ordinary table on the grounds that its boundaries would be inexact. As we’ve seen, the Second Philosopher is happy to accept this sort of ‘worldly vagueness’.

7. Eddington’s two tables may call to mind Sellars’s challenge to reconcile ‘the scientific image’ with ‘the manifest image.’ The manifest image includes much more than Eddington’s table No. 1—‘it is the framework in terms of which, to use an existentialist turn of phrase, man first encountered himself’ (Sellars [1962], p. 6)—but Sellars does come close to our concerns when he denies that ‘manifest objects are identical with systems of imperceptible particles’ (ibid., p. 26). He illustrates with the case of the pink ice cube: ‘the manifest ice cube presents itself to us as something which is pink through and through, as a pink continuum, all the regions of which, however small, are pink’ (ibid.), and of course the scientific ice

So let's grant the Second Philosopher her claim that formal structure as she understands it does turn up in our actual world. This means not only that rudimentary logic applies in such cases, but that it does so regardless of the physical details of the objects' composition, the precise nature of the properties and relations, any particular facts of spatiotemporal location, and so on. This observation might serve as the first step on a path toward the familiar idea, noted earlier, that questions like these are peculiarly philosophical: the thought would be that if the correctness of rudimentary logic doesn't depend on any of the physical details of the situation, if it holds for any objects, any properties and relations, etc., then it must be quite different in character from our ordinary information about the world; indeed, if none of the physical details matters, if these truths hold no matter what the particular contingencies happen to be, then perhaps they're true necessarily, in any possible world at all—and if that's right, then nothing particular to our ordinary, contingent world can be what's making them true.

By a series of steps like these, one might make one's way to the idea that logical truths reflect the facts, not about our world, but about a Platonic world of propositions, or a crystalline structure that our world enjoys necessarily, or an abstract realm of meanings or concepts, or some such distinctively philosophical subject matter. Many such options spring up in the wake of this line of thought, but ordinary facts, ordinary information about our ordinary world has been left behind, and ordinary inquiry along with it—we've entered the realm of philosophy proper.

But suppose our Second Philosopher doesn't set foot on this path. Suppose she simply notices that nothing about the chemical makeup of the coin is relevant, that nothing about where the coin is located is relevant, that only the formal structure matters to the reliability of the rudimentary logic she's isolated. From here she simply continues her inquiries, turning to other pursuits in geology, astronomy, linguistics, and so on. At some point in all this, she encounters cathode rays and black body radiation, begins to

cube isn't at all like this. Here Sellars seems to think, with Eddington, that science isn't in a position to tell us surprising things about what it is for the ice cube to be (look) pink; he seems to agree with Eddington that some apparent features of the manifest ice cube can't be sacrificed without losing the manifest ice cube itself. Indeed, the essential features they cling to are similar: a kind of substantial continuity or homogeneity. The Second Philosopher remains unmoved.

theorize about discrete packets of energy, uses the quantum hypothesis to explain the photo-electric effect, and eventually goes on to the full development of quantum mechanics. And now she's in for some surprises: the objects of the micro-world seem to move from one place to another without following continuous trajectories; a situation with two similar particles *A* and *B* apparently isn't different from a situation with *A* and *B* switched; an object has some position and some momentum, but it can't have a particular position and a particular momentum at the same time; there are dependencies between situations that violate all ordinary thinking about dependencies.⁸ Do the 'objects', 'properties', 'relations', and 'dependencies' of the quantum-mechanical micro-world enjoy the formal structure that underlies rudimentary logic? The Second Philosopher might well wonder, and sure enough, her doubts are soon realized. In a case analogous to but simpler than position and momentum, she finds an electron *a* with vertical spin up or vertical spin down, and horizontal spin right or horizontal spin left—(*Ua* or *Da*) and (*Ra* or *La*)—but for which the four obvious conjunctions—(*Ua* and *Ra*) or (*Ua* and *La*) or (*Da* and *Ra*) or (*Da* and *La*)—all fail. This distributive law of rudimentary logic doesn't obtain!

We're now forced to recognize that those very general features the Second Philosopher isolated in her formal structures actually have some bite. Though this wasn't made explicit, an object in a formal structure was assumed to be an individual, fundamentally distinct from all others; having location, for example, was assumed to involve having a particular (though perhaps imprecise) property, a particular location, not just some location or other. These features were so obvious as to go unremarked until the anomalies of quantum mechanics came along to demonstrate so vividly that they can in fact fail.⁹ Those of us who ventured down that path the Second Philosopher

8. For more on these quantum anomalies, with references, see [2007], III.4.

9. In yet another twist in the tradition of Eddington and Sellars, Ladyman and Ross [2007] begin from this observation—that the micro-world doesn't seem to consist of individual objects—then go on to classify the ordinary table, along with the botanist's giant redwoods and the physical chemist's molecules, as human constructs imposed for 'epistemological book-keeping' (p. 240) on an entirely objectless world. I suspect that this disagreement with the Second Philosopher traces at least in part to differing pictures of how 'naturalistic' metaphysics is to be done. The Second Philosopher's 'metaphysics naturalized' simply pursues ordinary science and ends up agreeing with the folk, the botanist, and the chemist that there are tables, trees, and atoms, that trees are roughly constituted by biological items like cells, cells by chemical items like molecules, molecules by atoms, and so on. She doesn't yet know, and may never know, how to extend this program into the objectless micro-world, but she has good reason to continue

didn't take were tempted to think that her formal structure is to be found in every possible world, but it turns out it isn't present even in every quarter of our own contingent world!¹⁰ Rudimentary logic isn't necessary after all; its correctness is contingent on the very general, but still not universal, features isolated in the Second Philosopher's formal structure.

We've focused so far on the metaphysics—what makes these inferences reliable, these truths true?—but there's also the epistemology—how do we come to know these things? If we followed the philosopher's path and succeeded in dismissing the vicissitudes of the contingent world as irrelevant well before the subsequent shocks dealt the Second Philosopher by quantum mechanics, then we might continue our reasoning along these lines: if logic is necessary, true in all possible worlds, if the details of our contingent world are beside the point, then how could coming to know its truths require us to attend to our experience of this world?¹¹ Again a range of options flourish here, from straightforward theories of a priori knowledge to complex accounts of how logic serves to constitute inquiry and thus can't itself be confirmed. But let's return to the Second Philosopher's more naive inquiries, still well clear of the philosopher's path, and ask how she answers the simple question: how do we come to know that rudimentary logical inference is reliable?

trying, and even if she fails, she doesn't see that this alone should undermine our belief in the objects of our ordinary world. In contrast, the 'naturalized metaphysics' of Ladyman and Ross is the work of 'naturalistic philosophical under-labourers' (ibid., p. 242), designed to show 'how two or more specific scientific hypotheses, at least one of which is drawn from fundamental physics, jointly explain more than the sum of what is explained by the two hypotheses taken separately' (ibid., p. 37)—and it's this project that delivers the surprising result that ordinary objects are constructed by us. From their perspective, the Second Philosopher's 'metaphysics naturalized' is just more science: the botanist and the physical chemist make no contribution to ontology; metaphysics begins only when their hypotheses are unified with fundamental physics. From the Second Philosopher's perspective, there's no reason to suppose that ordinary objects are human projections or to insist that assessments of what there is must involve unification with fundamental physics. Indeed, from her perspective, given our current state of understanding, quantum mechanics is perhaps the last place we should look for ontological guidance!

10. This incidentally removes another sort of skeptical challenge to the Second Philosopher's belief in ordinary macro-objects, namely, the charge that an inquiry starting with objects with properties, etc., will inevitably uncover objects with properties.
11. An inference from necessary to a priori is less automatic in our post-Kripkean age, when many philosophers recognize a posteriori necessities, but logical truth seems a poor candidate for this sort of thing. In any case, what I'm tracing here are tempting paths, not conclusive arguments.

In general, the Second Philosopher's epistemological investigations take the form of asking how human beings—as described in biology, physiology, psychology, linguistics, and so on—come to have reliable beliefs about the world—as described in physics, chemistry, botany, astronomy, and so on.¹² Work in psychology, cognitive science, and the like is primary here, but the Second Philosopher's focus is somewhat broader: not only does she study how people come to form beliefs about the world—she also takes it upon herself to ask how these beliefs align with what her other inquiries have told her about how the world actually is, and to assess which types of belief-forming processes, in which circumstances, are reliable. Though her epistemology is naturalized—that is, it takes place roughly within science—it's also normative.

In the case of rudimentary logic, the Second Philosopher's focus is on formal structure: her other studies of the world have revealed the existence of many objects, with properties, standing in relations, with dependencies, and she now asks how we come to be aware of these worldly features. Here she recapitulates the work of an impressive research community in contemporary cognitive science.¹³ The modern study of our perception of individual objects reaches back at least to the 1930s, when Piaget used experiments based on manual search behavior¹⁴ to argue that a child reaches the adult conception of a permanent, external object by a series of stages ending at about age two. Conflicting but inconclusive indications from visual tracking suggested that even younger children might have the object concept, but it wasn't until the 1980s that a new experimental paradigm emerged for testing this possibility: habituation and preferential looking. In such an experiment, the infants are shown the same event over and over until they lose interest, as indicated by their decreased looking time (habituation); they're then shown one or another of two test displays, one that makes sense on the adult understanding of an object, the other inconsistent with the adult understanding;

12. This is reminiscent of many of Quine's descriptions of his 'epistemology naturalized', but Quine also tends to fall back on more traditional philosophical formulations, asking how we manage to infer our theory of the external world from sensory data (see [2007], I.6, for more).

13. I can give only the smallest sampling of this work here. For more, with references, see [2007], III.5.

14. E.g., does the child lift a cloth to find a desirable object she's seen hidden there?

if an infant is thinking like the adult, the inconsistent display should draw a longer gaze (preferential looking).

So, for example, suppose a metal screen is attached to a long hinge that extends from left to right on a stage; the screen can lie flat toward the viewer on the stage surface, and it can pivot through a 180° arc to lie flat away from the viewer. The infant is habituated to seeing the screen move through this range of motion. Then the screen is positioned toward the infant, a box is placed behind it, and the screen is rotated backward. The consistent display shows the screen stopping when it comes to rest on the now hidden box; the inconsistent display shows it moving as before and coming to rest on the stage surface away from the viewer. If the infant thinks the box continues to exist even when it's hidden by the screen, and that the space it occupies can't be penetrated by the screen, then the inconsistent display should draw the longest gaze. (Notice that the inconsistent display is exactly the one the infant has been habituated to, so its very inconsistency would be sufficiently novel to overcome the habituation.) In early uses of the new paradigm, this is exactly what was observed in infants around five months of age.

Obviously, this is only the beginning of the story. For example, does the infant understand the box as an individual object, as a unit, or just as an obstacle to the screen? Experiments of similar design soon indicated that infants as young as four months perceive a unit when presented with a bounded and connected batch of stuff that moves together. Now imagine a display with two panels separated by a small space. An object appears from stage left, travels behind one screen, after which an object emerges from behind the second screen and vanishes stage right. One group of four-month-olds is habituated to seeing an object appear in the gap between the screens, as if it moved continuously throughout; another group is habituated to seeing an object disappear behind the first screen and an object emerge from behind the second screen without anything appearing in the gap. The test displays are then without panels, showing either one or two objects. The result was that the infants habituated with the apparently continuous motion looked longer at the two-object test display than the infants habituated with the scene where no object was seen in the gap. It seems an object is regarded as an individual if its motion is continuous.

Of course there's much more to this work than can be summarized here, but the current leading hypothesis is that these very young infants conceptualize individual units in these terms: they don't think that such a unit can be in two places at once or that separate units can occupy the same space, and they expect them to travel in continuous trajectories. In the words of Elizabeth Spelke, a pioneer in this field, the infant's objects are 'complete, connected, solid bodies that persist over occlusion and maintain their identity through time' (Spelke [2000], p. 1233):

Putting together the findings from studies of perception of object boundaries and studies of perception of object identity, young infants appear to organize visual arrays into bodies that move cohesively (preserving their internal connectedness and their external boundaries), that move together with other objects if and only if the objects come into contact, and that move on paths that are connected over space and time. Cohesion, contact, and continuity are highly reliable properties of inanimate, material objects: objects are more likely to move on paths that are connected than they are to move at constant speeds, for example; and they are more likely to maintain their connectedness over motion than they are to maintain a rigid shape. Infants' perception appears to accord with the most reliable constraints on objects. (Spelke et al. [1995], pp. 319–320)

Partly because so much of this research depends on experiments conducted with habituation/preferential-looking and closely related designs, partly for other reasons,¹⁵ these conclusions can't be taken as irrevocably established, but then the fallibility of ongoing science is an occupational hazard for the Second Philosopher. Let's take the early emergence of a modest human ability to detect (some of) the world's individual objects as a tentative datum.

15. E.g., Hatfield [2009b] argues that the findings of Spelke and her collaborators establish only that young infants perceive 'bounded trackable volumes', not 'individual material objects'. Of course, Spelke (e.g., in Spelke et al. [1995], cited by Hatfield) does allow that the infant's object concept continues to develop in early childhood, so there is room here for clarification of levels or degrees of object perception. A question like this would prompt the Second Philosopher to get down to sorting things out, but I'm not so idealized an inquirer and leave these further investigations to others.

As for properties and relations, the infants' sensitivity to these plays a role in the habituation/preferential-looking studies mentioned earlier: habituating to green objects then preferentially looking at red ones must involve noticing those colors, likewise the spatial relations of objects and the screens. What's surprising is that object properties aren't initially used to individuate them. Ten-month-old infants watched as a toy duck emerged from the left side of a single screen, followed by a ball emerging from the right side of the screen; one of the two test displays then showed the duck and the ball, the other just the duck—and no significant difference between their reactions was found! The same experiment run on slightly older one-year-olds delivered what an adult would have expected: the test display with the duck alone drew greater attention.

On reflection these results aren't so bewildering. While infants begin with simple but highly reliable spatiotemporal constraints on object identity (as Spelke notes), property distinctions require more judicious application: a red ball can turn blue and still be the same object; a human can change clothes and still be the same person. Some experience and learning must be needed for the child to realize that ducks don't generally turn into balls, and considerably more to reach the full adult concept:

We are inclined to judge that a car persists when its transmission is replaced, but would be less inclined to judge that a dog persists if its central nervous system were replaced . . . Because we know that dogs but not cars have behavioral and mental capacities supported by certain internal structures, we consider certain transformations of dogs to be more radical than other, superficially similar transformations of cars. (Spelke et al. [1995], pp. 302–303)

With this in mind, it's less surprising that the beginnings of the child's identification of objects by their properties comes a couple of months later than their identification by the more straightforward spatiotemporal means, and perhaps even that this new development apparently coincides with the acquisition of their first words—property nouns like 'ball' and 'duck'!

So as not to belabor this fascinating developmental work, let me just note that similar studies have shown that young infants detect conjunctions

and disjunctions of object properties, the failure of properties or relations,¹⁶ simple billiard-ball-style causal dependencies, and so on. It's also notable that many of these abilities found in young infants are also present, for example, in primates and birds. This suggests an evolutionary origin, and clearly the advantages conferred by the ability to track objects spatiotemporally, to perceive their properties and relations, to notice dependencies, would have been as useful on the savanna as they are in modern life.

All this leaves the Second Philosopher with two well-supported hypotheses: the ability to detect (at least some of) the formal structure present in the world comes to humans at a very early age, perhaps largely due to our evolutionary inheritance; whether by genetic endowment, normal maturation, or early experience, the primitive cognitive mechanisms underlying this ability are as they are primarily because humans (and their ancestors) interact almost exclusively with aspects of the world that display this formal structure. From here it's a short step to the suggestion that the presence of these primitive cognitive mechanisms, all tuned to formal structure, is what makes the simpler inferences of rudimentary logic strike us as so obvious. Assuming that the Second Philosopher has this right—that the formal structure is often present, that we are configured to detect it, and that this accounts for our rudimentary logical beliefs—then a sufficiently externalist epistemologist might count this as a case of *a priori* knowledge. An epistemologist of more internalist leanings might hold that the sort of a posteriori inquiry undertaken here would be required to support actual knowledge of rudimentary logic. The Second Philosopher isn't confident that this disagreement has a determinate solution, isn't confident that the debate is backed by anything more substantial than the various handy uses of 'know', so she's content to offer a fuller version of the story sketched here, and to leave the decision about 'knowledge' to others.¹⁷

16. [[An update: in [2007], p. 259, my argument, such as it was, that detecting negations is part of our basic cognitive machinery rested on developmental experiments with infants showing that they can classify objects together as sharing some property—the suggestion being that they classified the rejected objects as lacking that property. The trouble is that classifying some objects as 'dogs' doesn't imply grouping the other objects together as 'non-dogs'. Since then, Gonzalez [n.d.] has identified evidence from an even more primitive system in intermediate-level vision: given a brief glimpse of an array of colored dots, subjects were able to locate the center of gravity of a heterogeneous group of non-blue dots only slightly less accurately than they located the center of gravity of the blue dots.]]

17. [[See Essays #1 and 6 for more on 'know'.]]

Notice, incidentally, that if this is right, if the Second Philosopher's formal structure is so deeply involved in our most fundamental cognitive mechanisms, this explains why it's so difficult for us to come up with a viable interpretation of quantum mechanics, where formal structure goes awry. But this observation raises another question: if formal structure and hence rudimentary logic are missing in the micro-world, and if these are so fundamental to our thought and reasoning, how do we manage to carry out our study of quantum mechanics? Some suggest that we should adopt a special logic for quantum mechanics,¹⁸ but the question posed here is how we manage to do quantum mechanics now, apparently using our ordinary logic. I think the answer is fairly simple: what we actually have in quantum mechanics isn't a theory of particles with properties, in relations, with dependencies, but a mathematical model, an abstract Hilbert space with state vectors.¹⁹ This bit of mathematics displays all the necessary formal structure—it consists of objects with properties, in relations, with (logical) dependencies—so our familiar logic is entirely reliable there.²⁰ The deep problem for the interpretation of quantum mechanics is to explain how and why the mathematical model works so well, to figure out what worldly features it's tracking, but in the mathematics itself, our natural ways of thinking and reasoning are on impeccable footing.

Now for all its advertised virtues—reliability in a wide variety of worldly settings, harmony with our most fundamental cognitive mechanisms—rudimentary logic is in fact a rather unwieldy instrument in actual use. We've seen, for example, that the presence of indeterminacies eliminates the law of excluded middle, the principle of non-contradiction, and indeed all logical truths. An inference rule as central as *reductio ad absurdum* can be seen to fail: that $(q\text{-and-not-}q)$ follows from p tells us only that p is either false or indeterminate. And the substantive requirements on dependency relations undercut most of our usual manipulations with the conditional. Though he's

18. See the discussion of deviant logics below.

19. As noted above (footnote 9), Ladyman and Ross [2007] argue on grounds similar to the Second Philosopher's that the micro-world doesn't consist of objects. Given her account of how our cognition and our logic work, the Second Philosopher would predict that these authors should encounter some difficulty when it comes to describing the subject matter of quantum mechanics, and in fact, what they say on that score is consistent with the line taken here: 'it is possible that dividing a domain up into objects is the only way we can think about it' (p. 155); 'we can only represent [the non-objectual structures of the micro-world] in terms of mathematical relationships' (ibid., p. 299).

20. In fact, the mathematical world is in some ways more amenable to our logical ways (see footnote 21 below).

speaking of a full Kleene system, with a truth-functional conditional, I think Feferman's assessment applies to rudimentary logic as well: 'nothing like sustained ordinary reasoning can be carried on' (Feferman [1984], p. 95).

Under the circumstances, a stronger, more flexible logic is obviously to be desired. The Second Philosopher has seen this sort of thing many times: she has a theoretical description of a given range of situations, but that description is awkward or unworkable in various ways. To take one example, she can give a complete molecular description of water flowing in a pipe, but alas all practical calculation is impossible. In hope of making progress, she introduces a deliberate falsification—treating the water as a continuous substance—that allows her to use the stronger and more flexible mathematics of continuum mechanics. She has reason to think this might work, because there should be a size-scale with volumes large enough to include enough molecules to have relatively stable temperature, energy, density, etc., but not so large as to include wide local variations in properties like these. This line of thought suggests that her deliberate falsification might be both powerful enough to deliver concrete solutions and benign enough to do so without introducing distortions that would undercut its effectiveness for real engineering decisions. She tests it out, and happily it does work! This is what we call an 'idealization', indeed a successful idealization for many purposes. (It would obviously be unacceptably distorting if we were interested in explaining the water's behavior under electrolysis.) In similar ways, we ignore friction when its effects are small enough to be swamped by the phenomenon we're out to describe; we treat slightly irregular objects as perfectly geometrical when this does no harm; and so on.

With the technique of idealization in mind, the Second Philosopher looks for ways to simplify and streamline her theoretical account of formal structure, that is, her rudimentary logic, in ways that make it more flexible, more workable, and to do so without seriously undermining its reliability. To this end, she makes two key idealizations, introduces two falsifying assumptions—that there is no indeterminacy, that any particular combination of objects and properties or relations either holds or fails; and that dependencies behave as material conditionals²¹—and at a stroke, she

21. Though these idealizations involve falsification in her description of the physical world, they are satisfied in the world of classical mathematics: excluded middle holds and the dependencies are logical. For more on the ontology of mathematics, see [2011b].

transforms her crude rudimentary theory into our modern classical logic. There can be no doubt that full classical logic is an extraordinarily sophisticated and powerful instrument; the only open question is whether or not the required idealizations are benign. And as in the other examples, this judgment can be expected to vary from case to case.

This is where some of the so-called ‘deviant logics’ come in. Proponents of one or another of the various logics of vagueness, for example, may insist that indeterminacy is a real phenomenon,²² may condemn ‘the lamentable tendency . . . to pretend that language is precise’ (Burgess [1990], p. 434). On the first point, the Second Philosopher agrees—indeterminacy is real—but she views the classical logician’s pretending otherwise as no different in principle from the engineer’s pretending that water is a continuous fluid; what determines the acceptability of either pretense isn’t the obvious fact that it’s a pretense, but whether or not it is beneficial and benign in the situation at hand. Most logics of vagueness begin from a picture not unlike the Second Philosopher’s, in which, for example, a property can hold of an object, fail to hold, or be indeterminate for that object; there’s also the problem of higher-order indeterminacy, that is, of borderline cases between holding and being indeterminate, between being indeterminate and failing. So far, I think it’s fair to say that there is no smooth and perspicuous logic of vagueness, no such logic that escapes Feferman’s critique. It is, of course, true that classical logic can lead us astray in contexts with indeterminacy—this is the point of the sorites paradox—but at least for now the Second Philosopher’s advice is simply to apply classical logic with care,²³ as one should any idealization, rather than switch to a less viable logic.²⁴

22. There is serious disagreement between various writers over the source of the indeterminacy: is it purely linguistic or does the world itself include borderline cases and fuzzy objects? Here the Second Philosopher sides with the latter, but this shouldn’t affect the brief discussion here.

23. Sorensen [2012] credits this approach to H. G. Wells: ‘Every species is vague, every term goes cloudy at its edges, and so in my way of thinking, relentless logic is only another name for a stupidity—for a sort of intellectual pigheadedness. If you push a philosophical or metaphysical enquiry through a series of valid syllogisms—never committing any generally recognized fallacy—you nevertheless leave behind you at each step a certain rubbing and marginal loss of objective truth and you get deflections that are difficult to trace, at each phase in the process’ (Wells [1908], p. 11).

24. Williamson [1994] also advocates retaining classical logic, but his reason is quite different: because there is no real vagueness, because apparent borderline cases really just illustrate our ignorance of where the true borderline lies. This strikes many, including me, as obviously false.

Advocates of various conditional logics protest the Second Philosopher's other bold idealization: replacement of real dependencies with the simple material conditional. There are many proposals for a more substantial conditional, far too many to consider here (even if my slender expertise allowed it), but perhaps the conditional of relevance logic can be used as one representative example. The motivation here speaks directly to the falsification in question: the antecedent of a conditional should be 'relevant' to the consequent.²⁵ To return to our earlier example, the redness of this rose isn't relevant to the fact that $2 + 2 = 4$, despite the truth of the corresponding material conditional (if the rose is red, then $2 + 2 = 4$). Of course, as before, the Second Philosopher fully appreciates that the material conditional is a falsification, that the rose inference is an anomaly, but the pertinent questions are whether or not the falsification is beneficial and benign, and whether or not the relevance logician has something better to offer. Again I think that for now we do best to employ our classical logic with care.

So we see that some deviant logics depart from the Second Philosopher's classical logic by rejecting her idealizations,²⁶ and that our assessment then depends on the extent to which the falsifications introduced are beneficial and benign, as well as the systematic merits of the proposed alternative. But not all deviant logics fit this profile; some concern not just the idealizations of classical logic, but the fundamentals of rudimentary logic itself. Examples include intuitionistic logic—which rejects double negation elimination—quantum logic—which rejects the distributive laws—and dialetheism—which holds that there are true contradictions. Given the connection of rudimentary logic with the Second Philosopher's formal structure, the challenge for each of these is to understand what the world is like without

25. Relevance logicians are particularly unhappy with what they call 'explosion', the classical oddity that anything follows from a contradiction. For related reasons, full relevance logic rejects even some rudimentary logical inferences not involving the conditional, like disjunctive syllogism, but I leave this aside here. (For a bit more, see [2007], p. 292, footnote 24.)

26. Some other deviant logics respond to idealizations of language rather than the worldly features of rudimentary logic; e.g., free logicians counsel us to reject the falsifying assumption that all naming expressions refer. Here, too, our assessment depends on the effectiveness of the idealization and the viability of the alternative. In practical terms, leaving aside the various technical studies in the theory of free logics, I'm not sure using a free logic is readily distinguishable from being careful about the use of an existential quantifier introduction in the context of classical logic. In any case, our concern here is with worldly idealizations, not linguistic ones.

this formal structure, what the world is like that this alternative would be its logic.²⁷

Of the three, intuitionistic logic comes equipped with the most developed metaphysical picture, but it's suited to describing the world of constructive mathematics, not the physical world.²⁸ Quantum logic at first set out to characterize the non-formal-structure of the micro-world, but in practice it has not succeeded in doing so;²⁹ the problem of interpreting quantum mechanics remains open. And dialetheism faces perhaps the highest odds: as far as I know, its defenders have focused for the most part on the narrower goal of locating a compelling example of a true contradiction in the world, perhaps so far without conspicuous success.³⁰ The Second Philosopher tentatively concludes that rudimentary logic currently has no viable rivals as the logic of the world, and that classical logic likewise stands above its rivals as an appropriate idealization of rudimentary logic for everyday use.

In sum, then, the Second Philosopher's answer, an ordinary answer to the question of why that coin must be foreign, is that the coin and its properties display formal structure and the inference in question is reliable in such situations. This answer doesn't deliver on the usual philosophical expectations: the reliability of the inference is contingent, our knowledge of it is only minimally *a priori* at best. The account itself results from plain empirical inquiry, which may lead some to insist that it isn't philosophy at all. Perhaps not. Then again, if the original question—why is this inference reliable?—counts as philosophical—and it's not clear how else to classify it—then the answer, too, would seem to have some claim to that honorific. But the Second Philosopher doesn't care much about labels. After all, even 'Second Philosophy' and 'Second Philosopher' aren't her terms but mine, used to describe her and her behavior. In any case, philosophy or not, I hope the Second Philosopher's investigations do tell us something about the nature of that inference about the coin.³¹

27. Our interest here is in the logic of the world, not the logic that best models something else, as, e.g., paraconsistent logic (a variety of relevance logic) might serve to model belief systems (see [2007], pp. 293–296).

28. See the discussion of Creator Worlds in [2007], pp. 231–233, 296, [2014b], II.

29. See [2007], pp. 276–279, 296.

30. See [2007], pp. 296–297.

31. My thanks to Patricia Marino for helpful comments on an earlier draft and to Penelope Rush for editorial improvements.

PART IV

MATHEMATICS

Psychology and the a priori Sciences

The ‘a priori sciences’¹ to be considered here are logic and arithmetic;² the ‘psychology’ includes experimental, especially developmental psychology, neurophysiology, and vision science. My goal is to examine the role these empirical theories can play in the philosophies of those sciences, or more precisely, the role I think they should play. Most of the psychological studies referred to here will be familiar to readers of this volume,³ though perhaps not the use to which I hope to put them.

1. LOGIC

Common sense tells us that much of the world comes packaged into middle-sized objects—stones, coins, snails, apples, trees, the bodies of cats, apes, human beings—and not without reason; these items are what we see and touch, encounter and engage with, in everyday life. Of course common sense doesn’t always hold up under scrutiny, but meticulous science confirms that each of these is a rough collection of molecules, held together by various forces, resisting penetration due to other forces, moving as a bounded unit on a continuous spatiotemporal path.⁴ Scientifically refined common sense

1. I use the term ‘a priori science’ as the customary label, not to endorse the view that these disciplines are in fact a priori in some sense or other.

2. With a nod toward set theory in footnote 51.

3. [[This essay was published in Sorin Bangu’s collection on the role of psychology and cognitive science in the philosophies of logic and mathematics.]]

4. Some philosophers question this simple view on the grounds that the commonsense table is intuitively ‘solid’, while the scientific object is largely empty space, so the two cannot be the same. Even assuming that common sense does picture things as continuous matter (over and above being impenetrable), it seems more natural to say science has taught us that the objects of common sense are different than we

also reveals that these objects have properties and stand in relations: stones come in a variety of sizes and shapes, apples in various colors, domestic cats are generally smaller than adult humans.⁵ This simple structuring validates a certain amount of rudimentary logic: if the apple is either red or green, and it's not red, then it must be green. This might all seem so obvious, so unavoidable, as to be true no matter what, true in 'all possible worlds', but in fact it breaks down at the quantum level: the particles don't behave as bounded units on continuous paths; the sense in which they enjoy properties (like position and momentum) is problematic; and some simple logical laws (e.g., the distributive law) appear to fail.⁶ The inferences of this rudimentary logic are reliable as long as the requisite structure is in place, but not otherwise.

Psychology comes into this story as an investigation of how we come to those early commonsense beliefs about objects and their features. The groundbreaking developmental work of the 1980s and 90s⁷ showed that young infants track cohesive, bounded, solid⁸ individuals, despite occlusions, using spatiotemporal criteria like contiguity, common fate, and continuous motion; though they're aware of objects' other properties, they typically don't use regularities of shape, color, texture, or motion, or features or kinds to determine object boundaries or identity.⁹ Studies of neonates and non-human animals suggest a distant evolutionary origin.

first imagined, not that they don't exist. (For a bit on more such thinking from Eddington, Sellars, and Ladyman and Ross, see Essay #8, pp. 219–220.) Other philosophers go farther, rejecting everything in science and common sense on radical skeptical grounds, but this challenge, too, I set aside for present purposes (for more, see [2017a]).

5. Dependencies between one situation and another are also important—the coin is on the floor because the cat shoved it off the table—as are universal properties, but I leave these aside for simplicity in this quick sketch of [2007], III.4 (also Essay #8).
6. In another skeptical move, it's sometimes suggested that science can't serve to ratify the objects of common sense, because any science that begins with those objects will inevitably end up ratifying them. In fact, a science (ours) that begins with them has ended up without them in the quantum world.
7. For a summary with references, see [2007], pp. 245–258. Carey [2009] (chapters 2 and 3) is a much-discussed survey and philosophical elaboration by one of the leading researchers in the area. This work swept away the earlier seminal theories of Piaget (featured in [1990], pp. 54–55), according to which the ability to represent objects comes later in development. Carey [2009], pp. 46–55, gives a fascinating re-analysis of Piaget's evidence.
8. That is, impenetrable (a feature of both Eddington's tables in footnote 4).
9. Animate/inanimate, human/non-human appear to be exceptions. See Carey [2009], pp. 263–284, especially pp. 276–277, for more on this point.

There's some disagreement over the precise interpretation of these experimental results. The contents of 'the object concept' vary slightly from writer to writer;¹⁰ disagreements arise over whether the abilities catalogued are purely perceptual or somehow conceptual;¹¹ and so on.¹² Fortunately, these niceties needn't trouble us here because any of these options will be enough to serve as the building blocks, the 'objects with properties', in the rudimentary logical structure described above.

But 'object with properties' isn't all there is to that rudimentary structure: a stone has a size and a shape, an apple can be red or green, a coin can fail to be a quarter.¹³ More developmental work of the 1980s and 90s shows that young infants classify cats and dogs so as to exclude birds, and even cats so as to exclude superficially similar dogs. They're also sensitive to correlations of features: infants¹⁴ aware of three possibilities for each of the features A, B, C, D, and E (that is, A1, A2, A3, and so on), habituated to items with correlations between these features—for example, items with $(A1 \wedge B1 \wedge C1) \wedge (D1 \vee D2) \wedge (E1 \vee E2)$ and items with $(A2 \wedge B2 \wedge C2) \wedge (D1 \vee D2) \wedge (E1 \vee E2)$ —find a new correlated combination (like A2-B2-C2-D1-E2) familiar but an uncorrelated combination (like A1-B2-C1-D1-E2) just as novel as one that's entirely new (like A3-B3-C3-D3-E3)! In addition to these conjunctions, infants

10. In one of the more dramatic examples, Burge [2010] departs from many psychologists (from Piaget on) in holding that the 'constitutive conditions' for representing bodies as such don't include tracking through occlusion: 'a capacity to perceptually track a body as a three-dimensionally bounded and cohesive volume shape while it remains in view . . . suffices' (p. 460). In contrast, Hatfield [2009b] requires tracking through occlusion, but holds that the developmental evidence doesn't conclusively show infants are representing objects as 'individual material objects (not as mere local collections of properties) . . . that . . . occupy . . . distinct . . . space-time worms . . . throughout their existence' (p. 241).

11. E.g., Burge [2010], pp. 438–450, in disagreement with Spelke [1988].

12. It could be that some of these disagreements run deeper than the sort of thing scouted in footnote 10. Sticking with Burge and Hatfield as our examples, notice that Burge takes the goal of the project to be determining what's 'constitutive' of objecthood—'Our question concerns necessary minimal constitutive conditions for having the capacity to attribute the kind *body* in perception' (Burge [2010], p. 465)—where this presupposes a fact of the matter to be discovered (perhaps by rational intuition, perhaps with a hint of essentialism in the appeal to 'natural kinds'). In contrast, Hatfield ([2009b], p. 241, claims only that 'we as adult perceivers typically see [things] as individual objects' with the features listed in footnote 10, and that the developmental evidence doesn't establish that infants do this, too. He describes this situation by referring to 'the adult concept', but there's no indication of an underlying Burge-like metaphysics; he could just be using the phrase to highlight the possibility of a significant cognitive shift.

13. See footnote 16 of Essay #8.

14. This pattern and the next emerge in 10-month-olds, still pre-linguistic.

also appear sensitive to disjunctions—habituated to cats or horses, they find a dog novel—and to relations—for example, ‘above’, ‘below’, ‘between’. Results like these strongly suggest that we humans are sensitive to rudimentary logical structures from an early age.¹⁵

Still, as is well known, it’s entirely possible to respond to a feature of the world without representing it: the frog’s visual system might allow it to detect (then catch and eat) flies without representing them as flies. On this point, I’m less confident than Tyler Burge that ‘representation’ is a psychological natural kind,¹⁶ and even more doubtful that its contours can be discovered, as he suggests, by uniquely philosophical means.¹⁷ In contrast, Gary Hatfield [1988] undertakes a more modest task, firmly grounded in contemporary vision science.¹⁸ An ongoing debate pits those who believe that the visual system employs symbolic representations in an internal symbolic ‘language’—that rules are encoded and applied, hypotheses formed and tested (in the tradition of Helmholtz)—and those who insist that the visual system is not representational, that it’s simply tuned to directly register the rich and complex information available in the ever-changing array of ambient light (Gibson and his followers). Both sides acknowledge that processing takes place between the retina and the visual experience. The debate between them hinges on the question: is this processing purely physiological or does it break down into psychologically significant components, and in particular, into components with a characteristically representational role?¹⁹ Hatfield threads the needle between the two schools of thought, arguing that there are representational components, but that they needn’t involve a symbolic system.

To see how this goes, consider using a slide rule to multiply n times m : locate n on the A scale; slide 1 on the B scale beneath n on the A scale; find the

15. For a summary, with references, see [2007], pp. 258–262.

16. See Burge ([2010], p. 291): ‘Psychological explanations have a distinct explanatory paradigm. Psychology . . . discovers its own kinds. One of them is the kind representation.’

17. See Burge ([2010], p. xviii): ‘philosophy has . . . a set of methodological and conceptual tools that position it uniquely to make important contributions to understanding the world . . . Many of its topics remain of broadest human concern. Where, constitutively, representational mind begins is such a topic.’

18. Obviously this contrast (elaborated in footnotes 16 and 17) is reminiscent of the one in footnotes 10 and 12.

19. See Ullman [1980], p. 374; Hatfield [1988], §1.

number on the A scale that's above m on the B scale. The procedure works because the scales are laid out logarithmically and $n \propto m = \ln^{-1}(\ln(n) + \ln(m))$. That same equation could be programmed into a digital computer and multiplication carried out in that way, in which case the logarithmic algorithm itself would be encoded, represented, in the computer's program, but this isn't true for the slide rule: there, the algorithm is effectively followed, but it isn't literally represented.²⁰ The lengths on the slide rule represent numbers because of what the device is designed to do (multiply, among other things) and how it was designed to do it (relying on, but not representing, the properties of logarithms). The computer is also designed to multiply (among other things), but it's designed to do so quite differently, by applying explicit rules in an internal symbolic system. So, what a device does or doesn't represent depends on how it does what it's designed to do.

If we now replace the slide rule designer or the computer programmer with the evolutionary pressures on our species,²¹ the representational status of some element of the visual system can be assessed in the same way: it depends on the function of the visual system in the evolved human organism, the function of that element within the visual system, and the method it uses to perform that function.²² Hatfield ([1988], pp. 63–65) gives the example of seeing a circle at a slant rather than an ellipse. This function could be achieved by registering the retinal ellipse, registering slant information from shading, and computationally combining these two; or it might be achieved by a single registration of shading across the retinal ellipse. Obviously it's an empirical matter which of these algorithms is actually implemented; it can be investigated by psychological experiments with carefully timed disruptions or by physiological investigation of the neuroanatomy. In these ways, we could determine whether or not, say, the projective retinal shape by itself is

20. As Burge ([2010], p. 504) points out, an odometer's computation of the distance traveled depends on the circumference of the tires (it records a tick for each rotation), but the circumference is nowhere represented. Hatfield ([1988], p. 75) makes a similar point about a 'tension adder': n and m are represented by small weights placed on a pan and their sum registered by a pointer on the front of the device, but no algorithm is encoded.

21. In practice, determining what aspects of the visual system are adaptations and which are spandrels is a very difficult undertaking. See, e.g., Warren [2012], Anderson [2015].

22. Burge ([2010], chapter 8) soundly rejects accounts of representation based in biological function, for reasons I don't fully understand and won't attempt to explicate.

represented. But either way, it's not at all obvious that this sort of representation would involve a symbolic system.²³

So, to return to our theme, we know that infants respond to conjunctions, disjunctions, negations, and so on, but do they actually represent them as such? As Burge notes ([2010], p. 406), the fact that we infer in accord with a logical rule doesn't imply that the rule is somehow encoded in our psychology, presumably in some language of thought. Our concern here, though, isn't with inference, but with simple logical structuring, and (following Hatfield) the representation needn't be symbolic. The question is whether the infant represents the stone as small and round, the apple as red or green, the coin as not-a-quarter. Assuming that sensory sensitivity to these worldly features is adaptive,²⁴ a Hatfield-style answer to this question hinges on how that sensory sensitivity is achieved: does the scientific story of that ability break down into psychologically significant parts, into representational components, like the separate representations of projective shape and of shading information in the multi-stage algorithm for seeing the circle at a slant? For our case: given that the infant can represent stones, smallness, and roundness, is her representation of a small, round stone related to other representations in a way that merits describing it as a conjunctive representation? This needn't involve encoding in some language of thought any more than the circle-at-a-slant case does, but it is a straightforward empirical question for experimental psychology and neuroscience.

If a definitive answer to this question is known, it isn't known to me, but the study of visual working memory offers a hint of how a small part of it might go. Evidence suggests that we're able to store information about a limited number of objects (around four) and their features over short periods of

23. Does this mean that the frog is representing flies? Opponents of biological function views suggest that evolutionary considerations aren't enough to show that the frog is representing flies as opposed to moving black dots, or even flies as opposed to nothing at all, given that a frog's detector will occasionally go off on its own. Regarding the first point as subject to further investigation, granting it now only for the sake of argument, Hatfield contends that nevertheless biological function 'can serve as the basis for ascribing to states of the frog's visual system the content *target fly/moving dot*, or some such coarse-grained content', that '[a]mong the functions of the frog's visual system is to represent small moving things as being there when they are, and not to represent them as being there when they aren't' (Hatfield [1991], pp. 122–123). In other words, the biological function account has room to regard the frog's detector as having misfired when it goes off on its own.

24. Recall footnote 21.

time. This raises the question how several features of one object are bound together: what distinguishes a scene with a vertical red bar and a horizontal green bar from one with a vertical green bar and a horizontal red one? One proposal is synchronized neural firing: a particular neuron fires repeatedly to encode a single feature; when the repeated firings of the neurons for two separate features are synchronized, they form a unit, ‘cell assembly’.²⁵ In Hatfield’s terms, the initial firings represent red, green, horizontal, and vertical bars; when the ‘vertical’ and ‘red’ neurons fire in unison, the resulting assembly represents a vertical red bar. The position I’m proposing, on pure speculation, requires that this isn’t exceptional, that rudimentary logical structuring is widely represented, one way or another.

If all this is granted, what role is psychology playing in this philosophy of logic? The ground of logical truth, what makes it true (where it is true), is the objective logical structuring in the world, so there’s no trace of psychologism. Psychology’s role, then, might be thought to be epistemological. For example, a sufficiently externalist epistemologist, one who thinks the evolutionary pressures responsible for our logical cognition produce a reliable process,²⁶ might conclude that we know (at least some of) the world’s logical structure a priori. I prefer to leave the policing of ‘know’²⁷ and ‘a priori’ to the specialists and to say only this much: we come to believe what we do about the logical structures in the world on the basis of primitive cognitive mechanisms, many of which we share with other animals, but our evidence for the correctness of those beliefs comes from common sense and its subsequent (partial) ratification by scientific means.

So far, this is a fairly slight philosophical impact for psychology, but I think there’s an important moral concerning our philosophical preconceptions about logical truth. Because our logical beliefs rest on such primitive cognitive mechanisms, it’s hard for us to see how things could be otherwise, how a world failing to instantiate those rudimentary logical forms is even

25. See Vogel, Woodman, and Luck [2001] for discussion and references. Also Olson and Jiang [2002].

26. This needn’t be a fallacious argument of the form ‘this evolved, therefore it’s reliable’. Instead, it might run roughly along the lines traced here: first science establishes that much of the world is logically structured; then psychology defends an evolutionary story of how we come to detect and represent that structure.

27. [[See Essays ##1 and 6 for more on ‘know’.]]

possible. When quantum mechanics shows us not only that a world can fail to do this, but that our very own micro-world so fails, often the result is that we find quantum mechanics deeply problematic, not that we take logic itself to be contingent.²⁸ It seems to me that the psychology here is showing us *why* we're so easily inclined to believe that logical truth is necessary, a priori, certain—a stubborn preconception that vastly distorts our theorizing about it.²⁹ It's hard to imagine a more valuable lesson for the philosophy of logic!

2. ARITHMETIC

Obviously any patch of the world with logical structuring into objects with properties, standing in relations, will also have number properties: so many objects, so many with this particular feature, so many standing in this relation to this particular individual, and so on. When it comes to our cognitive access to those number properties, though, it's well known that the first four or so have special status: infants' expectations about how many objects will appear behind a screen after individual objects have been added or removed are accurate up to three; adults can hold three to four objects in working memory³⁰ and track three to four objects through complex motions, but these abilities break down quickly for higher quantities. Non-human animals share these abilities and limitations, indicating another primitive cognitive system.³¹

28. I suspect many of us have heard our fellow philosophers assert with great confidence that quantum mechanics must be false, on a priori grounds.

29. See [2014b], chapter 6, for a comparison of this conclusion with the late Wittgenstein's take on logic.

30. Feigenson [2011] describes how visual working memory can encode more than four slots' worth of information by 'chunking', as when we remember a phone number by dividing it into three blocks of digits, or one of her infant subjects remembers two cats and two cars, but not four individual cats. (See also Carey [2009], pp. 149–150.) The 'chunk' is often referred to as a 'set', as exhibiting the higher ranks that differentiate sets from mere aggregates. I once appealed to analogous considerations (e.g., in [1990], p. 165), but for what it's worth, I'm no longer convinced anything essentially 'higher-order' is involved in such cases. Seeing two cats and two cars could just be a particular way of seeing the cats and the cars, not a way of seeing something else (a set of cats, a set of cars).

31. See [2007], pp. 319–326, for more on the story in this and the following three paragraphs, with references. See Carey ([2009], chapter 8) for expert elaboration.

The mechanism underlying these abilities—the object-tracking or parallel individuation system—apparently includes both the so-called ‘object files’ of mid-level vision,³² which follow objects spatiotemporally and encode features as they go,³³ and visual working memory, which keeps visual information accessible over short periods.³⁴ The two are closely intertwined, with some evidence of complementary emphases, on tracking over motion and retention of object properties, respectively.³⁵ Though the infant expectation experiments are often described in arithmetic terms— $1 + 1 = 2$, $3 - 2 = 1$ —it’s widely agreed that these representations are not truly numerical: not ‘3’ but the simply logical ‘a thing, another thing, and yet another thing’, most likely the opening of three successive object files.³⁶ In cases of ‘subitizing’—immediate recognition (without counting) of up to three or four objects—perhaps visual working memory is engaged, but again, most likely through the opening of three distinct information slots³⁷ rather than an explicit numerical representation.

Yet another primitive system we share with other animals is sensitive to approximate quantities: it can distinguish one dot from three more easily than two dots from three (the ‘distance effect’), and it can distinguish two dots from three, but not eight dots from nine (the ‘magnitude effect’).³⁸ The mechanism for this is so far unknown (at least to me), but neurological studies on monkeys suggest a two-step process that begins with a group of neurons that encode locations of objects, ignoring other features, then feeds into an array of neurons whose responses are bell-shaped curves, each peaking at a certain number.³⁹ This model would explain the distance effect—the ranges of firing

32. Kahnemann et al. [1992]. See [2007], pp. 255–257, 319–320, for a brief discussion with references.

33. ‘It’s a bird, it’s a plane, it’s Superman.’

34. See, e.g., the references in footnote 25.

35. See Hollingworth and Rasmussen [2010]. The two are often lumped together without comment, or even identified.

36. Burge points out that ‘[t]here need be no use of conjunction or negation in the perceptual representation (as in: this is a body and this is a body and this is not that)’ (Burge [2010], p. 486). He’s right: it’s unlikely that anything like this is encoded in a language of thought. But in Hatfield’s terms, the opening of three successive object files could represent the corresponding logical feature of the scene.

37. See Chesney and Haladjian [2011] for evidence that subitizing and object tracking rely on a shared visual mechanism.

38. See, e.g., Carey [2009], pp. 118–137; Dehaene [2011].

39. See Dehaene [2011], pp. 247–254; Nieder [2011].

for ‘1-neurons’ and ‘3-neurons’ overlap less than those for ‘2-neurons’ and ‘3-neurons’—and the magnitude effect—the bell curves for large numbers are broader. In any case, this is clearly a more quantitative system than the object tracker, but it can’t truly be said to represent cardinality. Burge ([2010], p. 482) suggests a return to the ancient notion of ‘pure magnitude’, neither continuous nor discrete, but nevertheless stands in ratios. However that may be, what matters for our purposes is that features of the world’s logical structure are being represented, albeit only approximately.⁴⁰

So far, we’re in step with the non-human animals, still far short of human arithmetic. The leading theory is that what sets us apart is the child’s ability to combine the proto-numerical fruits of the object-tracking system with the approximate system via her command of the counting sequence.⁴¹ ‘One, two, three . . .’ is first learned as a verbal nonsense scheme—like ‘eeny, meeny, miney, moe . . .’—and the act of reciting it while pointing to each of a group of objects in turn is just play, of no numerical significance. Young children do realize that use of the word ‘one’ correlates with the presence of a single object, with a single opened object file, or a single item in visual working memory, but the sense of larger number words emerges only gradually, in two-and-a-half- to three-and-a-half-year-olds: first ‘two’ is associated with the presence of an object and another; a few months later ‘three’, and maybe even ‘four’, gain meaning from the object-tracking system.⁴² This far the non-human animals can follow, but what happens next is uniquely human: apparently the child notices that an extra object in the scene corresponds to the next number in the counting sequence, and suddenly the true meaning of counting becomes clear: the last number recited in the procedure is the number of objects in the scene.

40. Oddly enough, on small numbers, where the two systems overlap, the infant’s object tracking system appears to override the approximate system. E.g., they prefer a box where three treats have been placed to a box where one or two treats have been placed, but when the numbers are two and five, beyond the object tracker’s capacity, they perform at chance—despite the fact that the ratio is big enough for the approximate system to detect easily. See Carey [2009], pp. 84–85, 139–141, 153–155.

41. Here again, Piaget was in disagreement. See Dehaene ([2011], pp. 30–36) for an amusing account of how the empirical results were misinterpreted.

42. Some hold that the object-tracking system isn’t involved, that the underlying mechanism here is the approximate number system (see, e.g., Piazza [2011]), which is most precise for small numbers. Dehaene ([2011], pp. 256–259), who once entrusted small numbers to the approximate number system (what he calls ‘the number sense’), explains what changed his mind.

It's sometimes assumed that this is the end of the story of how humans come to a full understanding of arithmetic, but it isn't, for at least two reasons. First, consider a child who knows how to count, and knows there are 'just as many' of these as those when the same number word results from counting these as those. That is, she knows that if she counts n children and n cookies, she'll be able to give each child exactly one cookie with no cookies left over. Richard Heck makes the case that a child can know all this without having the notion of a one-to-one correspondence, which is, after all, 'very sophisticated' (Heck ([2000], p. 170). Of course when she counts, she forms what we understand to be a one-to-one correspondence, but she needn't understand it as such; she's just implementing the counting procedure.⁴³ So this is one respect in which the child still hasn't grasped a notion some consider essential to the concept of 'cardinal number'.

Another tempting assumption is that a child who understands that one more object corresponds to the next number word must also understand that there's no largest number.⁴⁴ The only empirical study touching on this question that I know of, Harnett and Gelman [1998], actually aims to show that it's relatively easy for children to learn that the number sequence has no end—easy compared to learning fractions!—so its design includes more coaching than would be ideal for present purposes. Still, children in kindergarten and first grade⁴⁵ did quite poorly on questions like 'is there a biggest number of all numbers?' and 'is there a last number?' They did somewhat better but still far from perfectly on leading questions like 'if we count and count and count, will we ever get to the end of the numbers?' and 'can we always add one more, or is there a number so big we'd have to stop?', despite having been primed with exercises in counting larger and larger numbers.⁴⁶

43. Heck also notes that the child can understand 'just as many' without understanding counting: there are just as many cookies as children if she can make sure everyone has exactly one cookie with none left over. He then shows how the Peano Axioms can be derived with 'just as many' in place of Frege's 'one-to-one correspondence'.

44. Some of the line of thought in the remainder of this section is also explored in [2014a].

45. Averaging just under six and seven years old, respectively.

46. One group of subjects in one of the studies was questioned about the largest number, etc., before the counting exercises. Their performance was even worse than the group who did the counting exercises first.

Explaining their answers, the six-year-olds might suggest that we have to stop counting ‘because you need to eat breakfast and dinner’ or ‘because we need sleep’ or that we couldn’t then start up again where we left off because ‘you forget where you stopped’. There’s even a hint at mortality: if we try to add one more after counting to a very big number, ‘I guess you’ll be old, very old’. Though answers like these were classified as ‘unacceptable’, there is a straightforward sense in which the children have it right: there *are* practical limitations on how far we’re inclined to count, and even physical limits on how far we can count.⁴⁷ The young children aren’t wrong exactly, they’re just failing to grasp the spirit of the question. What’s being asked is whether there’s any limit to how far we can count, *in principle*.

In contrast to the kindergarteners and first-graders, the second-graders⁴⁸ in this study generally answered the questions as they were intended: there is no largest number, period. Closer analysis of the experimental results led Harnett and Gelman to the observation that the children in a position ‘to benefit from a conversation that offers cues’ (Harnett and Gelman [1998], p. 361) were those who could count beyond 100:

Once children master the sequence from 1 to 20 and the list of decade words, they have most but not all of the vocabulary they need to apply the recursive procedures by which larger and larger numbers are generated. As they count beyond 100, they come to learn that not only the digits, but also the decade terms, are recycled over and over. [Younger c]hildren are still at work memorizing the teens and decade terms and are less able to appreciate that the count sequences is systematic. (Ibid., p. 361)⁴⁹

This suggests, as the psychologist Paul Bloom proposes, that

47. Russell once remarked that running through an infinite decimal expansion is ‘*medically impossible*’ (Russell [1935–1936], p. 143).

48. Averaging just under eight years old.

49. Though Harnett and Gelman speak of ‘recursive procedures by which larger and larger numbers are generated’, obviously they’re talking about linguistic procedures that generate numerical expressions. (Understanding that adding one results in a larger number was another predictor for successful response to the cues.)

the generative nature of human numerical cognition develops only as a result of children acquiring the linguistic counting system of their culture. Many, but not all, human groups have invented a way of using language to talk about number, through use of a recursive symbolic grammar. (Bloom [2000], p. 236)

This would mean that children's belief in the infinity of the numbers derives from their belief in the infinity of numerical expressions, not vice versa:

[I]t is not that somehow children know that there is an infinity of numbers and infer that you can always produce a larger number word. Instead, they learn that one can always produce a larger number word and infer that there must therefore be an infinity of numbers. (Ibid., p. 238)

In this way, our question—how do we come to believe there's no largest number?—is pushed back one step to: how do we come to believe that there's no largest numerical expression?

Harnett and Gelman's studies show that it's quite easy for children to come to this view once they've appreciated the systematic intricacies of generating numerical expressions. What's striking is that they don't seem bothered by concerns about the practical or physical limitations on, for example, the length of those numerical expressions or the breath needed to utter them or the need to stop for lunch—all that apparently matters is coming to grasp the recursive character of the rules of formation. Why is the intended 'in principle' reading of the question more natural here, when it's posed for numerical expressions, than it was when posed for the numbers themselves? To engage once again in rank speculation, I suggest that this traces to the recursive element of the innate linguistic faculty, whatever it is in our genetically endowed cognitive machinery that underlies our ability to understand and produce indefinitely varied and complex linguistic items:

All approaches agree that a core property of [the linguistic faculty] is recursion . . . [The linguistic faculty] takes a finite set of elements and yields a potentially infinite array of discreet expressions. This capacity . . . yields a

discrete infinity (a property that also characterizes the natural numbers).
(Hauser, Chomsky, and Fitch [2002], p. 1571)

The suggestion is that this linguistic capacity is what produces our intuitive grasp of the ‘in principle’ question.

Assuming this sketch of the psychology is roughly right—a big assumption, subject to empirical test—the consequences for the philosophy of arithmetic are fundamental. Simple arithmetical claims like $2 + 2 = 4$ and $12 < 191$ are ordinary facts about worldly logical structures (where they’re present), but the subject matter of mathematical arithmetic—the standard model, what we now think of as an omega-sequence—doesn’t depend on any contingent features of the actual world, which may or may not be finite. Insofar as arithmetic is ‘about’ anything, it’s about an intuitive picture of a recursive sequence of potentially infinite extent, an intuitive picture we humans share thanks to the evolved linguistic faculty common to our species.

Now we all tend to believe that the structure of the standard model of arithmetic, that simple omega-sequence, is coherent, unique, and determinate. But if it’s really just a matter of an intuitive picture, what reason do we have to believe these things? As Wittgenstein once asked, ‘what if the picture began to flicker in the far distance?’ (*RFM*, V.10). Our innate cognitive structuring may well give rise to these firm convictions, but if the story told here is correct, our capacity for mathematical arithmetic could be a mere spandrel, generated by the way as we evolved toward language, and even if it is an adaptation in itself, that’s no guarantee of reliability.⁵⁰ Under the circumstances, we reflective beings should want more support for our faith in the cogency of an omega-sequence than just our brute inclination to believe it. I think there are facts we can appeal to, but they’re hardly conclusive: our biological similarity as humans is reason to think your intuitive picture is more or less the same as mine; the apparent coherence of the picture, plus long experience of the species with mathematical arithmetic, provides some evidence for its consistency; the lack of any important independent statement comparable to the Continuum Hypothesis suggests it may be fully determinate.⁵¹ But our

50. See the fallacy described in footnote 26.

51. There’s an analogous question for set theory, where the relevant intuitive picture—the iterative hierarchy—seems to rest on three elements: recursion (presumably based in the same cognitive faculty

sense that arithmetic is more secure than that may be an illusion—another valuable lesson from psychology!

3. CONCLUSIONS

Though psychologists sometimes take their work to support a brand of anti-realism about mathematics—Stanislas Dehaene’s influential *Number Sense*, for example, bears the subtitle *How the Mind Creates Mathematics*—in fact their skepticism doesn’t extend to the contingent logical/numerical structure I’ve been attributing to the world or our cognitive access to it:

[A]rithmetic . . . draw[s] upon a store of fundamental knowledge accumulated over millions of years of evolution *in a physical world which, at the scale we live it, is . . . numerically structured*. (Dehaene in Dehaene and Brannon [2011], p. 187, emphasis added)

This type of straightforward realism breaks down, I’ve suggested, with the potential infinite, the standard model of arithmetic, where attention to the psychological facts reveals that our cognitive architecture does, in a sense, ‘create’ the subject matter under investigation. In addition to this positive semantic or metaphysical conclusion, empirical work in psychology also uncovers the less than firm underpinings of some of our firmest philosophical preconceptions: that logic is necessary, that arithmetic is obviously

as the standard model of arithmetic); the combinatorial notion of an arbitrary subset, not beholden to any rule, definition, or construction (perhaps related to Heck’s ‘very sophisticated’ one-to-one correspondence?); and Cantor’s gusty bet on the completed infinite (see [1988], I.5). This picture isn’t definitive of the field in the way the standard model is for arithmetic: it wasn’t present when set theory was founded by Cantor and others, and it could be altered or replaced in the future (e.g., by the multiverse conception (though for now I’m skeptical about that (see [2017b], III))), much as it itself once replaced the picture of a set as the extension of a concept). In any case, given the added vagaries of the two additional elements, any case for cogency is correspondingly weaker: determinacy is undercut by independent statements like the CH, and our biological similarity gives less support for uniqueness. Perhaps the apparent coherence of the conception delivers some evidence of consistency, but considerably less than in the case of arithmetic. Still, this would be a form of so-called ‘intrinsic’ support distinct from the merely instrumental role described in [2011b].

cogent (coherent, unique, determinate). This valuable therapeutic helps free the philosophies of these subjects from traditional baggage and set them on a more vital course. In these ways, psychological inquiry stands to play a central, highly beneficial role in our philosophizing about the a priori disciplines.⁵²

52. Thanks to Gary Hatfield, Ethan Galebach, Reto Gubelmann, and Sorin Bangu for helpful comments on earlier drafts.

Do Numbers Exist?

We deal with ordinary counting numbers from our earliest years, but even mathematicians find it difficult to say what they are, or even whether they exist. Given that simple arithmetic tells us, for example, that there's a number between three and five, it's tempting to conclude that numbers do exist, that they are objects—much as stones and cats and planets are objects—on the straightforward grounds that arithmetic is true. We then note that numbers don't appear to be located anywhere in space, that they don't begin or end in time, that we don't pet them or trip over them or observe them in the night sky, which makes them different from cats and stones and planets. Unlike 'concrete' or 'physical' objects like these, numbers are 'abstract'.

I don't exactly disbelieve this nice story, but I do think it misses some of what's most interesting and intriguing about numbers and arithmetic. Let me explain.¹

We first encounter numbers in perfectly ordinary circumstances of everyday life: we find three apples on the table, five fingers on each hand, nine players on the baseball diamond. It's natural to describe these experiences as encounters with numbers, but perhaps more accurately, what we encounter are apples, fingers, and baseball players—not numbers exactly, but numbering. In the baseball case, for someone sadly unfamiliar with the game, this numbering might involve explicit counting; in other cases, the result will be obvious at a glance. But what is it that's obvious here or detected by a counting routine?

1. For a more detailed version of what follows, see [2014a]. [[See also Essay #9.]]

Now you might want to say that this ordinary visual experience or counting procedure has brought us into contact with the abstract object three or five or nine. This raises a question of what kind of contact this could be, given the lack of petting or tripping or observing,² but what troubles me more is that these ordinary situations don't seem on their face to involve anything other than the apples, fingers, and baseball players; what we've detected in each case appears to be simply a feature of that portion of the world. To get pedantic about it, there are, on the table, molecules held together in certain arrangements; some of these conglomerations of molecules are apples (a certain phase in a life cycle from seed, to tree, to fruit, to seed again); there is such a conglomeration here (pointing), another here (pointing again), yet another here (pointing yet again), and after these, there are no more such conglomerations on the table.³ This is a simple fact about the stuff on the table, a fact that's truly described by saying three apples are there. Likewise for the fingers and the baseball players. Much as some of the apples are red, others green; some of the fingers longer than others; one of the baseball players is a pitcher, another a center fielder—the stuff on the table is three apples. In each case, some portion of the world is made of objects, and those objects have various properties and stand in various relations.

Understood in this way, 'number' is a straightforward feature of a worldly situation, as real as any other property. But concrete objects like apples aren't the only things we number: if you're lucky, a genie might grant you three wishes. I'll leave it to others to decide whether wishes are objects and if so of what kind; all I want to claim is this: in whatever sense we speak of a wish, another wish, and yet another wish, in that sense, it seems to me, we're entitled to speak of three wishes. What matters is that our subject, real or imagined, comes separated into distinct items with properties and relations; as long as that structure is present, the items, real or not, can be numbered. For simplicity, let's stick to the concrete.

2. Some argue that our confidence that we do know about three and five and nine shows that nothing like tripping or bumping is required. That may be right, but it leaves a lingering curiosity about how that knowledge is obtained, if not by this kind of contact.
3. Notice that the stuff on the table also consists of many more than three molecules of various sorts. That structure is equally real and that larger number property equally present. But the fact that various worldly structures can overlap or cross-cut each other doesn't show that any structure at all can be attributed willy-nilly or that the structures that are there aren't real.

So far, the number talk we've considered has involved the detection of number properties: the apples on the table have *the property of being three*, much as the individual apples have *the property of being red*. But claims like 'there are three apples on the table' are only the entering wedge of our engagement with numbers. The argument for the existence of numbers sketched at the beginning is built on arithmetic, on claims like ' $2 + 2 = 4$ ' or 'there are three odd numbers between two and eight'. As it happens, the understanding in terms of number properties can be extended to cases like these: to say there are three odd numbers between two and eight is to say that there are three odd-number-properties between the-number-property-two and the-number-property-eight.⁴ To say that $2 + 2 = 4$ is to say that if you have a batch of stuff with twoness and another (non-overlapping) batch of stuff with twoness, then you have a big batch of stuff with fourness. We can understand all this talk about numbers as talk about properties of ordinary things, without appeal to any special abstract objects.

Of course, there's far more to arithmetic than simple existence claims like 'there's a number between three and five' and simple identities like ' $2 + 2 = 4$ '. Our more sophisticated arithmetical beliefs include, for example, generalities like the commutativity of addition: adding n to m gives the same answer as adding m to n , no matter what numbers n and m might be. This too can be understood as a claim about the behavior of number properties: the number property of the result of combining a batch of stuff with the number property n with a non-overlapping batch with property m is the same as it would be if they were combined in the reverse order. More ambitiously, we believe that for any number there's a next bigger number. We picture the numbers marching off into the distance without end, describing this with the suggestive notation of the dot-dot-dot: 1, 2, 3, 4, . . . Can this too be understood in terms of our down-to-earth number properties?

Let me sneak up on this question by asking another first: how do we come to believe that the numbers go on forever?⁵ As a matter of fact, this conviction comes rather late in childhood: youngsters of kindergarten age, despite

4. 'Odd-number-property' and 'between' for number properties can be characterized in obvious ways.

5. The following discussion of the psychology of the dot-dot-dot attempts an accessible version of the corresponding passage in Essay #9.

a full grasp of counting and of elementary arithmetic, are often stumped by questions like ‘is there a biggest of all numbers?’ or ‘is there a last number?’, and even by leading questions like ‘if we count and count and count, will we ever get to the end of the numbers?’ and ‘can we always add one more, or is there a number so big we’d have to stop?’ Asked to explain their answers, they suggest that we have to stop counting ‘cause you need to eat breakfast and dinner’ or ‘because we need sleep’, and we couldn’t start up again where we left off because ‘you forget where you stopped’, or that an attempt to add one more after counting to a very big number might fail because ‘I guess you’ll be old, very old’.

The psychologists who conducted these experiments naturally classified responses like these as ‘unacceptable’, but if we look at the question without preconceptions, there’s a real sense in which these young children are right: there *are* limits to how far any of us is willing to count, and more to the point, there are physical limits on how high any of us—or even the human race, assuming it dies out eventually—*can* count. For that matter, as far as physicists have been able to determine, there may well be a limit on the number of particles in the universe, a limit on the number properties realized anywhere in the world. What makes the young children’s answer ‘unacceptable’ is that this isn’t what we’re thinking of when we ask if there’s a largest number; we aren’t thinking of the amount of breath we have or the fate of the human race or the size of the physical universe. Where the young children have gone wrong is that they’ve failed to enter into the spirit of the question.

So what *are* we thinking of when we so confidently declare that there’s no largest number? Why are we so dismissive of the child’s concern that we might drop dead before we get around to counting the next number or the physicist’s possibility that the universe might be finite? My suggestion is that the underlying line of thought here goes something like this: even if there are only finitely many number properties actually present in our world, this is just happenstance—in *principle* there could be arbitrarily large arrangements of things; even if I can count only finitely far in my lifetime, this is just an accidental impediment—in *principle*, I could always generate a new number by adding one. What the kindergarteners are missing is that the question isn’t about what can actually be done or what actually exists, but about what holds *in principle*.

It turns out that second-graders do understand the question this way: they answer without hesitation that there is no largest number, end of story. So, what's changed between kindergarten and second grade that turned this into the utterly obvious answer? The relevant difference, the researchers found, was the older children's fluency with the various ways of generating numerical words and phrases, what philosophers call numerical 'expressions'. After the single digits (0–9), the teen words (11, 12, 14 . . .), and the decade terms (20, 30, 40 . . .) have all been memorized, the older children come to recognize the immensely useful shortcut that the digits repeat (21, 22 . . . 31, 32 . . .), and that, after 100, the decade terms too begin to repeat (110, 120 . . .). While the younger children are still at an early stage in this process, still memorizing the teen and decade terms, the older children have reached the point of seeing how these linguistic building blocks can be repeated, and from there, it's apparently an easy step to the conviction that they can be repeated *indefinitely*. So the puzzle becomes: what makes this step so easy?

Now even younger children know that if you've counted out three things and you add one more thing, then when you count them all again, you'll get four, the next number word in the counting sequence—they know, in other words, that the sequence of number properties marches in lock-step with the sequence of numerical expressions—but they apparently regard neither sequence in the *in principle* sense. The slightly older children, after more experience with the sequence of numerical expressions, see both in the *in principle* sense. Psychologists hypothesize that it's the conviction that the sequence of numerical expressions continues indefinitely (regardless of limitations of breath) that produces the conviction that arbitrarily large numerical properties are possible—in *principle*. We think there's no largest number property because we think there's no largest numerical expression.

This train of thought leaves us with one last question: when we become fluent in generating larger and larger numerical expressions, why is it so natural to move to the *in principle* understanding, to think that the sequence can be continued indefinitely? Why does any worry about the limits of breath or lifetime spontaneously fall away? I want to suggest that the answer traces to a simple fact about human language—a fact that separates it from the communication systems of even the most intelligent non-human animals—namely, that it's generative: at its core, there are a few basic elements and a set of rules

that produce new expressions from old (think of ‘your father’s friend, your father’s father’s friend, your father’s father’s father’s friend . . .’ or ‘the cat who saw the rat who ate the cheese that fell from the shelf . . .’). The current theory is that humans come equipped at birth with a generative learning mechanism that enables the young of our species to learn the full range of their native language from a small array of samples.⁶ The encoded rules of this system are apparently fully general, with no special proviso for accidental limitations, despite the perfectly ordinary limitations we obviously encounter in practice. The suggestion, then, is that the unlimited generative rules of this inborn mechanism, on vivid display in the recurring patterns of ever-larger numerical expressions, are the unconscious source that makes it so natural to think the sequence continues indefinitely.

Assuming this is right—a big if, subject to empirical test—then the dot-dot-dot of arithmetic, our intuitive picture of the endless series of numbers, is grounded in an innate cognitive mechanism that’s part of our genetic endowment as humans. Claims like ‘every number has a successor’ are descriptions, not of the physical world, not of some independent realm of abstract objects, but of this shared human picture implicit in the psychological mechanisms that underlie our capacity for language. This means that arithmetic as a whole in fact serves two masters: on the one hand, it answers to ordinary number properties of ordinary things, properties we often simply see or discover by counting; on the other, it involves the dot-dot-dot, our idealized picture of the numbers going on forever, a picture entirely independent of how many ordinary physical things there actually are in nature. Nothing guarantees that the two will mesh successfully, that the sophisticated mathematical treatment of the shared picture might not imply, for example, false elementary identities like ‘ $3 = 5$ ’ or ‘ $2 + 2 = 5$ ’. Still, the apparent coherence of the picture and our long practical and mathematical experience with arithmetic provide some evidence that we aren’t in for any such rude surprises down the road.

Which brings us back, at last, to the question addressed by the argument we opened with: do numbers exist?⁷ On the view I’ve been sketching,

6. For a readable account of this view of language, see Pinker [1994].

7. The remainder of this essay attempts an accessible version of a metaphysical account like that of [2011b], but for numbers instead of sets.

simple claims like ‘there’s a number between three and five’ and ‘ $2 + 2 = 4$ ’ are straightforward truths about the number properties exemplified in the world. If we take ‘numbers’ to be these number properties, then they’re on a par with other properties—like color or flexibility or being an infielder—and the question of their existence is subsumed under the larger question of the existence of properties, much discussed by metaphysicians. I leave that debate to the experts. For our purposes, what matters is that when we take the apples on the table to have the number property three, the only objects involved are the apples, not some abstract object three.

Looking back the argument itself, then, notice that it’s simple claims like these that figure there as premises: the reader is expected to agree without hesitation that of course there’s a number between three and five, which is supposed to imply the existence of the abstract object four. The trouble with this, we now see, is that the truth of ‘there’s a number between three and five’ guarantees the existence only of a number property, not of an abstract object. To reach beyond this, we have to call on idealized arithmetic—the full force of the dot-dot-dot—for claims like ‘every number has a successor’ or ‘there is no largest number’. If these more sophisticated claims are true, what makes them true isn’t garden-variety properties like the number of fingers on my right hand; instead, they’re true in our intuitive picture of the numbers as objects marching off endlessly into the distance, objects just like stones and cats and planets except that they aren’t spatial or temporal, can’t be petted, tripped over, or observed. Our opening argument was intended to show that there are abstract objects, but it works only if it’s based on sophisticated claims of idealized arithmetic, because it’s only the truth of claims like this that actually require such things.⁸ Would the argument be so compelling if the premise were ‘the numbers go on forever’?

Perhaps not, but what really matters is this: are these stronger premises true?, does the argument establish the existence of abstract objects? Some philosophers would say that it does, that idealized arithmetic is true and thus that there are numbers, which are abstract objects. Other philosophers would

8. There might be metaphysical reasons to regard these ‘numbers’ as properties rather than objects, but if so, those properties would still be detached from worldly exemplars in the same way, still quite different from ordinary number properties. So I ignore this nicety.

say that it doesn't, that idealized arithmetic is an extremely useful theory, but that—like other idealizations—it's not literally true. Though it might jeopardize my membership in the philosophers' union, I'm not convinced that there's a fact of the matter about the existence or not of abstract 'numbers' in this sense. I take the simple-minded position that ordinary science is where our inquiry into what's true and what exists begins. It tells us about those stones and cats and planets, and many other things detected by tests far finer and theorizing far more subtle. Along the way, we uncover simple number properties and the truths about them. Eventually we take the step to the sophisticated arithmetic of the dot-dot-dot, a step guided not by the world but by our shared intuitive picture of numbers marching off endlessly into the distance. Up to that point, our understanding of 'true' and 'exist' has involved only concrete situations, known by ordinary scientific means, so the question before us now is whether the new mathematical enterprise of idealized arithmetic is just more of the same (where 'true' and 'exist' apply as before) or something else entirely (where truth and existence perhaps aren't the point).

In many ways, our pursuit of mathematical arithmetic *is* like our previous scientific investigations: it speaks of objects with properties, standing in relations; it uses the same logic; it values more general claims over less general ones (e.g., 'for all numbers n and m , $n + m = m + n$ over $2 + 3 = 3 + 2$ '); and so on. Still, its objects are abstract rather than concrete, and its methods too are quite different: one uses observation, experimentation, theory formation, and testing; the other uses axioms, definitions, and proofs. If you weigh the similarities more heavily than the differences, you're inclined to think that idealized arithmetic is well confirmed, most likely true, and as a result, that our argument shows there probably are abstract objects in addition to concrete ones. If, on the other hand, you weigh the differences more heavily, you might think truth and existence aren't really what's at issue here; you might think we're using our shared intuitive picture of the endless number sequence to guide the development of a mathematical theory that serves to organize and systematize the many simple truths about number properties—a mathematical theory that's been wonderfully effective in that role, as well as a source of the fascinating, purely mathematical elaborations of advanced number theory. On this second view, the only truths are the elementary

ones; the rest is a theoretical apparatus, a story, in which those elementary truths have been embedded. The question, then, is which considerations are in fact more weighty, which of these understandings of arithmetic is correct. Is idealized arithmetic a well-confirmed theory of an abstract subject matter or a wonderfully effective perspective on ordinary number properties that's not really in the business of describing a domain of objects?

Most philosophers would insist that there's a right and a wrong answer here—though they differ sharply over which is which—but this is where my union membership comes under suspicion: I'm not so sure. Suppose we agree that the elementary claims like ' $2 + 2 = 4$ ' are straightforward truths about the world and that the mathematical theory of the dot-dot-dot is guided by our intuition of the number sequence and serves to systematize the elementary claims. One side is inclined to conclude from this that the theory is true and numbers exist, the other that the theory is a wonderful thing but not one where truth and existence are relevant. It seems to me that they're just describing the same facts, the facts they agree on, in different ways.

Consider for comparison a case from physical chemistry. When water is cooled very quickly, it forms a solid without the usual crystalline structure of ordinary ice, but an amorphous structure more like that of glass. Some chemists speak of this as a kind of ice; they call it 'amorphous ice'. Others describe it as a distinct way that water can solidify; they call it an 'ice-like solid'. So who's right? It seems to me that there's no fact of the matter here; once we agree on the underlying facts about how water solidifies, those facts can be described either way. What I'm proposing is the analogous stance on numbers: once we agree on the underlying facts—about ordinary number properties, about the cognitive basis of our picture of the dot-dot-dot, about the role of mathematical arithmetic in organizing and systematizing the elementary truths about number properties—once we agree to all that, we can describe the situation either way, with 'true' and 'exist' or without.

But even if you disagree with this last, potentially heretical move, I commend to your attention those underlying facts!

Enhanced If-thenism

Perhaps the most fundamental question in the philosophy of mathematics is simply this: what are we doing when we engage in mathematical thinking? For centuries of recorded history, the developing answers to this question were straightforward. Mathematics began with the ancients in purely practical problems—we were measuring lengths and volumes, exchanging money, farming and building, studying the stars, constructing calendars—and gradually expanded into a complex study of the underlying structure of the physical world. This way of understanding the subject continued through the profound developments of the 17th century—Descartes, Galileo (with his ‘great book of Nature’, written in the language of mathematics), Newton, Leibniz—and into the 18th—with Euler foremost among many. But this near-identification of mathematics with natural science was gradually overthrown in the course of the 19th century, with increasing interest in more esoteric mathematical structures (abstract algebras, pathological functions, transfinite numbers); with the study of non-Euclidean alternatives that detached mathematical geometry from an exclusive focus on physical space; and with the rise of kinetic theory, which converted the smooth differential equations of the 18th century into abstract approximations to physical phenomena actually discrete in the small. The bold new pure mathematics was free to pursue whatever seemed of mathematical interest, whether or not the results were of use in natural science. The application of mathematics then became a matter of selecting an effective model from among the wide range of abstract structures now being studied.¹

1. This case is developed at greater length in [2008] and [2011b], chapter 1. See also [202?b].

While this new practice of pure mathematics produced an explosion of exciting new directions, the fundamental question—what are we doing when we engage in mathematical thinking?—was transformed into a far more difficult problem in the process. When mathematics is no longer a matter of investigating the underlying structure of the physical world, what is it? Two starkly different answers have occurred to reflective mathematicians. One of these, realism, echoes Plato in taking mathematics as the study of an independent realm of non-spatiotemporal, acausal abstracta.² Satisfying as this answer might be from the mathematician’s perspective, philosophers have been troubled by questions about the nature and status of these abstracta and our epistemic access to them. The other response, if-thenism (or ‘deductivism’),³ holds that contemporary pure mathematics is simply a matter of determining what conclusions follow from which assumptions. The various forms of realism have been widely discussed and debated in the philosophical literature, but if-thenism somewhat less so. My goal here is to explore its prospects—and ultimately to propose a version that seems to me an attractive possibility.

1. SIMPLE IF-THENISM

In the philosophical literature, if-thenism is often set in the context of the famous correspondence of 1899 to 1903 between Gottlob Frege and David Hilbert.⁴ This fascinating exchange touched on many fundamental issues—the nature of axioms and definitions, and the possibility of consistency proofs among them. Along the way, both isolated versions of if-thenism, ‘although’, as Michael Resnik notes, ‘neither adopted it’ (Resnik [1980], p. 106)—Frege because he stuck to the realistic notion that axioms must be true, Hilbert because he felt the pull of finitism.

2. I have in mind here a version what’s called ‘Robust Realism’ in [2011b]. (As often happens, the canonical nod toward Plato here masks the complexity of his actual views.)

3. Putnam [1967a] calls it ‘if-thenism’; Resnik ([1980], chapter 3), Shapiro ([2000], pp. 148–157), and Linnebo ([2017], pp. 48–55) call it ‘deductivism’. I follow Putnam’s usage because it seems to me a more vivid way of capturing the core idea that the mathematician is interested only in what follows from what. (See footnote 8 for more on this terminology.)

4. E.g., see Resnik [1980], chapter 3, and Shapiro [2000], pp. 148–157.

Around the same time, Bertrand Russell did adopt a version of the view

The typical proposition of mathematics is of the form $\varphi(x, y, z, \dots)$ implies $\psi(x, y, z, \dots)$, whatever values $x, y, z, \dots$ may have. (Russell [1903], p. 6)

as did Hilary Putnam, with a nod to Russell, some decades later:

Russell advocated a view of mathematics which he somewhat misleadingly expressed by the formula that mathematics consists of ‘if-then’ assertions. What he meant was not, of course, that all well[-]formed formulas in mathematics have a horseshoe as the main connective! but that mathematicians are in the business of showing that *if* there is any structure which satisfies such-and-such axioms (e.g. the axioms of group theory), *then* that structure satisfies such-and-such further statements (some theorems of group theory or other). (Putnam [1967a], p. 20)

As it happens, neither Russell nor Putnam held to if-thenism for long: Russell soon turned to the logicism of *Principia Mathematica* and Putnam to a realism inspired by the indispensability arguments. Their reasons for defecting are addressed below, along with a host of other standard objections, but first a few clarifications of the position under consideration.

Because of the early association with thinking that led Hilbert to his formalism, and perhaps because logical implication itself is in some sense ‘formal’, there’s been a tendency to characterize if-thenism as involving the claim that mathematical statements are uninterpreted or ‘meaningless’:

A *deductivist* . . . insists that . . . the practice of mathematics consists of determining logical consequences of otherwise uninterpreted axioms. The mathematician is free to regard the axioms (and the theorems) of mathematics as meaningless. (Shapiro [2000], p. 149)

This further component seems to me both inessential and ill-advised: there’s no obstacle to speaking of logical relations between meaningful statements, and to deny meaning to ‘ $2 + 2 = 4$ ’ or the Fundamental Theorem of Calculus is to employ a notion of meaning at considerable remove from the way the

term is ordinarily used in everyday life or linguistics.⁵ Notice also that the stipulation of a universally quantified hypothetical is unnecessarily restrictive: it's natural to see a simple theorem of group theory as saying anything that's a group satisfies the conclusion, but the general goal of determining what follows from what is present in other contexts as well, for example, in axiomatic set theory, where we want to claim simply that the axioms logically imply the theorem.⁶ Finally, some care must be taken in isolating the relevant axioms: only the 'axioms' of group theory (that is, the definition of 'group') are at play in a theorem of elementary group theory like 'every group has a unique identity', but some set-theoretic background is implicit, for example, in the classification of finite simple groups. In the first case, the if-thenist takes herself to have established one of those universally quantified theorems of logic—if x is a group, then x has a unique identity—in the second case, the logical implication is more complex—if ZFC (or perhaps something less) and if x is a finite simple group, then x is cyclic, alternating, Lie or sporadic.^{7,8}

5. Putnam appears to agree: the claim that the statements of infinitary mathematics 'are *meaningless combinations of signs* . . . does extreme violence to our intuitions. The view taken here . . . seems far closer to common sense' (Putnam [1967a], p. 35).

6. The literature on if-thenism is curiously wedded to the more restrictive formulation, insisting on versions like the above-quoted 'if there is any structure that satisfies such-and-such axioms . . . then that structure satisfies such-and-such further statements' (Putnam [1967a], p. 20) rather than the simpler 'such-and-such axioms imply such-and-such theorems' (ibid., p. 30). The latter seems to me a more natural rendition of the basic intuition that we're just trying to figure out 'what follows from what'. At one point, Putnam does raise the question of 'what it means to suppose that' structures exist, but he leaves this hanging and eventually opts for the idea that they are at least possible (ibid., p. 33). This way of thinking soon led to his modalism (Putnam [1967b]; see also his [1979], p. xiii), later elaborated in Hellman's [1989] modal structuralism. 'Structuralism' does seem a more apt label. [[See footnote 61.]]

7. I think this is more or less Putnam's second suggestion (Putnam [1967a], p. 22), minus the reference to structures (see previous footnote). He also confronts the objection that we only mean to be considering standard models of our 'ifs', which again strikes me as contrary to the spirit of if-thenism: e.g., ordinary first-order ZFC proves that all models of second-order ZFC of the same height are isomorphic and standard; some observers then worry that this proof could be taking place in a non-standard model of first-order ZFC, so those isomorphic models of second-order ZFC are really just non-standard in precisely the same way. But this purported problem arises only if we imagine there's some ambient, pre-theoretic semantics for the first-order ZFC in which we're carrying out our proofs, which no proper if-thenist would do. Something like this may be what Putnam means when he says that we only need a 'relational notion of non-standardness' (ibid., p. 24).

8. One last note on terminology: though I use the term 'if-thenism' and speak of 'the antecedent of the if-then' and the like, I don't mean to assume that the theories whose consequences are being explored are finitely axiomatizable (as should be clear from the reference to ZFC). Those so inclined might regard such cases as employing a purely disquotational notion of truth ('if the axioms of ZFC are true, then so-and-so is also true'), but the simpler route is just to regard the use of 'if-then' terminology here as a case of what mathematicians like to call 'abuse of notation'. I trust the meaning is clear.

As Resnik notes:

Deductivism is a powerful and appealing philosophy of mathematics [that] avoid[s] commitments to abstract entities, and . . . simplifies . . . epistemology. (Resnik [1980], p. 136)

Of course it simplifies the epistemology of mathematics only by reducing it to the epistemology of logic, but there's no point addressing that challenge before we consider how if-thenism fares against some fairly immediate objections. All philosophies of mathematics face serious, at least apparent, obstacles, and if-thenism is certainly no exception!

2.. SOME COMMON OBJECTIONS TO IF-THENISM

Let's begin with three traditional objections to if-thenism, only the third of which strikes me as serious enough to demand some substantial rethinking of the view.⁹ The first of these objections is quite simple: if mathematics is just a matter of figuring out what conclusions follow logically from which assumptions, what was going on before the advent of modern axiomatic methods? If Newton and Euler, Fermat and Gauss isolated no explicit postulational or definitional setting, if their proofs weren't always up to our standards of logical rigor, are we forced to say that they weren't, after all, mathematicians?! Surely this would be an unacceptable consequence, a sad and sudden end for if-thenism!

Fortunately for the if-thenist, this objection can be turned away with a dash of entirely appropriate humility. Though philosophies of mathematics often claim, it seems to me unwisely, to account for all of past, present, and future mathematics, if-thenism as characterized here only aims to describe the mathematics of the 20th and early 21st centuries, the practice as it arose out of the turn to pure mathematics over the course of the 19th. In earlier periods, especially the great flowering of analysis in the 18th century, the notion of 'pure' mathematics hadn't yet arisen; as noted earlier, mathematics

9. More objections are treated along the way in later sections.

was seen as the investigation of the underlying mathematical structure of the physical world. The calculus of Newton and Leibniz was riddled with foundational mysteries, and its powerful extensions in the hands of Euler and many others only exacerbated these problems, but the grounding of the resulting mathematics in empirical application pushed aside these concerns. As Lacroix famously remarked, ‘such subtleties as the Greeks worried about we no longer need.’¹⁰ The historian Morris Kline describes the situation this way:

The physical meaning of the mathematics guided the mathematical steps and often supplied partial arguments to fill in nonmathematical steps. The reasoning was in essence no different from a proof of a theorem of geometry, wherein some facts entirely obvious in the figure are used even though no axiom or theorem supports them. (Kline [1972], p. 617)

But when pure mathematics was freed of any necessary link to the world, this guidance was gone and the character of the subject changed. If-thenism (as I understand it here) only aspires to describe this new way of doing mathematics.

This response to the first objection is to remind ourselves of how mathematics was peeled off the world over the course of the 19th century; the second objection concerns what happens when parts of the new pure mathematics are subsequently applied in physical science. Intuitively, it seems that to apply ‘ $2 + 2 = 4$ ’ or the Mean Value Theorem, we need ‘ $2 + 2 = 4$ ’ itself, not ‘the Peano Axioms imply that $2 + 2 = 4$ ’, we need the Mean Value Theorem itself, not ‘ZFC implies the Mean Value Theorem.’ To do their job in the service of true descriptions, don’t the consequents themselves have to be unconditionally true?

When Putnam considers this question during his if-thenist phase, he focuses first on the use of simple arithmetical identities like our ‘ $2 + 2 = 4$ ’ and concludes that it’s essentially a valid inference of first-order logic in disguise.¹¹ My if-thenist

10. Quoted in Kline [1972], p. 618.

11. Putnam [1967a], pp. 27–29.

agrees.¹² Though Russell was apparently deterred from his if-thenism by simple mixed statements like ‘there are nine planets’, Putnam is happy to read these as purely logical claims as well.¹³ But this approach doesn’t go far. Putnam’s broader if-thenist solution to the problem of how mathematics is applied involves the claim

that mathematics is only used to derive statements in nominalistic [non-mathematical] language from statements in nominalistic language.
(Putnam [1979], p. xiii)

Soon, Putnam’s embrace of indispensability considerations, derived from W. V. O. Quine, led him to reject this assumption, and if-thenism along with it.

The indispensability phenomena are well-known: mathematics plays an ineliminable role in our best-confirmed scientific theories; indeed, in many cases, these theories can only be stated in mathematical terms. While these observations are entirely accurate, they don’t seem to me to have the consequences for mathematical truth and ontology that have been claimed for them: confirmation from experience doesn’t rise uniformly through all aspects of a scientific theory, and much of the mathematization of science occurs in contexts that are explicitly or implicitly idealized or approximate.¹⁴ In recent years, the argument has been recast in terms of explanation instead of confirmation, but it’s not clear to me that this changes the situation materially.¹⁵

12. She includes here identities involving sums, products, and exponents (see below, around footnote 22; also Essay #9). Once again the nature of logical truth is postponed to §4.

13. If ‘there are nine planets’ is understood to mean ‘if *M* is a model of *Principia Mathematica*, then the set of planets in *M* is a member of the 9 (the set of nine-element sets, defined in purely logical terms) of *M*’, as Russell seemed to think it must, then ‘we seem to be in trouble’ (Putnam [1967a], p. 30): if there is no model of *Principia*, the number of planets could be any number at all! Of course this isn’t the only way to understand it—there’s the familiar purely logical formulation—as Putnam goes on to point out.

14. This case against the indispensability arguments is presented in [1997], II.6 and II.7, and revisited in [2007], pp. 94–96, 314–317.

15. See Marcus [2014] for a recent entry in this debate, with a helpful overview of the literature. Note, in passing, that one of the central examples of a purported mathematical explanation is the famous life cycle of the cicada: these come in prime numbers to ‘minimize the intersection of cicada life-cycles with those of both predators and other species of cicadas’ (Marcus [2014], p. 352). The ‘mathematics’ in this case is ‘13 and 17 are prime’, which counts as a straightforward logical truth on the if-thenism being explored here (see footnotes 12, 22), and thus isn’t subject to if-then treatment in the first place. This response would seem to be open to Putnam as well.

The root of this difference of opinion on the viability of indispensability arguments lies, it seems to me, in two fundamentally contrasting visions of how mathematics works in scientific applications. To illustrate, consider Putnam's dramatic appeal to the law of gravitation:

To sketch the [indispensability] argument in a nutshell: . . . one wants to say that the Law of Universal Gravitation makes an objective statement about bodies—not just about sense data or meter readings. What is the statement? It is just that bodies behave in such a way that the quotient of two numbers *associated* with the bodies is equal to a third number *associated* with the bodies. But how can such a statement have any objective content at all if numbers and 'associations' (i.e., functions) are alike mere fictions? It is like trying to maintain that God does not exist and angels do not exist while maintaining at the very same time that it is an objective fact that God has put an angel in charge of each star and the angels in charge of each of a pair of binary stars were always created at the same time! If talk of numbers and 'associations' between masses, etc. and numbers is 'theology' (in the pejorative sense), then the Law of Universal Gravitation is likewise theology. (Putnam [1975], pp. 74–75)

The rhetoric here is undeniably persuasive, but let's pause a moment to ponder the case we've been offered.

Consider two bodies, say the earth and the sun. The Law of Universal Gravitation says:

$$F = G \frac{m_1 m_2}{r^2},$$

where m_1 is a number associated with the earth (its mass), m_2 is a number associated with the sun (its mass), r is the distance between them, G is a constant, and F is the gravitational force between them.

But now consider: what is r exactly? what is 'the distance between the earth and the sun'? It won't do to say we measure between the front surface of each, because the surface of the earth is quite bumpy in the large and ill-defined in the small, and the surface of the sun is even worse. In fact, as we learn

in elementary physics, a nice theorem going back to Newton shows that the gravitational force exerted by a spherical body acts as if all the mass were concentrated at its center, so in applying the Law of Universal Gravitation to the earth and the sun, we actually associate m_1 and m_2 with the points at their centers and take r to be the distance between those two points. But now we notice that ‘the center of the earth’ and ‘the center of the sun’ are also indefinite, and in fact, the proof of the nice theorem involves regarding the ‘body’ in question as a perfect sphere, breaking that sphere up into a collection of infinitesimally narrow ‘shells’ (hence the name ‘the shell theorem’), then summing over these by integration.¹⁶ As this point, the earth and the sun have dropped cleanly out of the picture, replaced by mathematical abstracta. What this ‘application’ of the Law of Universal Gravitation actually describes is an abstract mathematical situation involving two perfect spheres equipped with continuous ‘mass-density’ functions and so on.¹⁷

This suggests a picture of applications quite different from Putnam’s. He views the Law of Universal Gravitation, as applied to the earth and the sun, as a single claim that involves physical objects and properties—the earth, the sun, their masses—mathematical abstracta—real numbers—and functions between the two—from masses to numbers, from spatial separations to numbers. The question then arises of how this instance of the law, as a whole, could be true if the mathematical abstracta it apparently refers to don’t exist. In contrast, I’ve taken the application to involve a completely mathematical model—the earth and sun replaced by perfect spheres, masses by functions from those mathematical objects to real numbers, etc.—and the applied mathematician’s claim to be that this abstract model resembles¹⁸ the worldly

16. Newton no doubt understood this process differently than we do today. There was considerable confusion, e.g., about ‘physical infinitesimals’ and related notions, until the physical and the mathematical sorted themselves out with the rise of pure mathematics.

17. Application of the Law of Gravitation to more complicated objects wouldn’t involve perfect spheres, of course, and the calculus used would be more subtle than the shell theorem. But the point remains: the complicated physical objects would be replaced by pure geometric figures with precise boundaries, and the mathematics applied to these abstract surrogates.

18. ‘Resembles’ here is intended only as a very broad catch-all for the sorts of relations applied mathematicians cite to explain why the idealizations in their mathematical modeling are benign and beneficial. E.g., water can be regarded as a continuous fluid because there’s an intermediate point between a volume too small to have a stable temperature and a volume too large to have a uniform temperature; perfect billiard ball models of ideal gases work as well as they do because molecules have fairly stable effective radii; the earth can be replaced by a perfect sphere for reasons like those in the following footnote. It’s not clear to me that any general account of ‘resemblance’ could cover the variety

situation well enough to be used for purposes of describing the gravitational force, that the simplifications and idealizations involved are beneficial and benign.¹⁹ When we apply mathematics to the world, we sometimes know in detail how the abstract model relates to the physical situation (as in fluid dynamics), sometimes we know this only roughly (as in general relativity), and sometimes we have no clear idea at all (as in quantum mechanics). What's actually claimed about the world varies as this knowledge varies, but the general form is that the model resembles the physical situation well enough to be used for certain purposes.²⁰

On this view of the application of mathematics, nothing at all follows about the existence of abstracta: just as we might describe the familial dynamics surrounding a powerful but aging person by comparison with 'King Lear', we can describe a given worldly situation by comparison with an abstract model. Pure mathematics doesn't need to deliver truths or existing entities in order to play this role any more than Shakespeare does.²¹ All it needs to do is describe a generous range of interesting structures and leave it to the physicist to determine which best serves which descriptive purpose.

This response to the second objection reveals that even our Simple If-thenism isn't entirely simple. Not all statements of mathematics are regarded as disguised if-thens: elementary arithmetic claims, like $2 + 2 = 4$, are actually

of such explanations in particular cases. (See [2008], pp. 29–35, or [2011b], pp. 21–30, for a related discussion.)

19. Defending this claim involves pointing out that, e.g., the variations on the earth's surface are negligible compared to the distance to the sun, so simplifying its shape to a physical sphere and replacing that physical sphere with a mathematical sphere shouldn't distort the situation unduly. Of course the ultimate defense is its empirical success.
20. This is the intended upshot of the evolution described so quickly in the introductory paragraph above and at greater length in [2008] and [2011b], chapter 1. See also [2011b], IV.2, for discussion of a related concern raised by Liston.
21. Given this straightforward observation, it might be surprising that Leng [2010] expends so much effort explaining how her fictionalism about mathematical objects is compatible with applications. The explanation is that she's thinking of applications as Putnam does, of the Law of Universal Gravitation, for example, as involving both real physical objects and fictional abstracta. To make sense of this, she appeals to Walton's theory of fictions with props, which allows her to import the actual earth and the sun as props into a fiction about real numbers and functions. That this instance of the law is a legitimate move in the fiction-with-props is an indirect way of saying something literally true about the physical situation. My suggestion is that the earth and the sun are actually replaced by abstracta in a purely fictional story, which is then compared with the purely physical situation. [[For more on how a Putnam-like view of applications functions in Leng's fictionalism and Hellman's modal-structuralism, see [202?a].]]

logical truths, and the same goes for identities of arithmetic involving addition, multiplication, exponentiation, and their Boolean combinations.²² (Exactly what this means for the application of elementary arithmetic to the world has to wait for the treatment of logic in §4.) Though a common preference for generality produces many philosophies of mathematics that treat all parts, all branches, as essentially the same, there are unavoidable differences, and Simple If-thenism recognizes at least one of these. Furthermore, this important caveat considerably weakens some of the intuitive resistance to if-thenism: the if-thenist's opponent is robbed of the rhetorically powerful—'but surely $2 + 2 = 4$ is unconditionally true!'—leaving the somewhat less potent protest that the existence of a successor for every number shouldn't be conditioned on the Peano Axioms or the Mean Value Theorem on some ambient set theory.

On this view of applications, where pure mathematics describes an array of abstract models that may or may not end up being applied, a new question arises: obviously not every such description succeeds in specifying a coherent mathematical structure; in the brave new age of unfettered pure mathematics, free from the constraints of physical reality, how do we determine which structures are coherent, legitimate? The answer that emerged in the early 20th century, the answer that is now standard orthodoxy, holds that set theory is the final adjudicator: the legitimate structures are those modeled in V (the universe of sets).²³ To maintain the new freedom of mathematical inquiry, set theory should be as generous as it can be, and this provides one of the fundamental methodological commitments of set-theoretic practice.²⁴ For over a century, V has served this purpose well, first as codified by Zermelo, now with extensions including large cardinals (LCs). Of course, there's no guarantee that this success will continue indefinitely—this standard would no doubt fall if mathematically interesting and/or physically useful structures were located outside of V ²⁵—but for now, this is how mainstream pure mathematics is practiced.

22. See [2007], pp. 318–319; [2014a]; Essays ##9 and 10.

23. See [2011b], pp. 30–32. For a more nuanced account of set theory's foundational role, see [2017b], [2019].

24. This admonition to 'maximize' is a central theme of [1997].

25. Some argue that this has already happened, e.g., that the category of all groups or the category of all categories ought to be legitimate but isn't supported by set theory. This raises the question

This brings us to the third, and I think most serious, common objection to if-thenism: if pure mathematics is just a matter of determining what follows from what, any consistent theory should be as good as any other in the ‘if’ part.²⁶ But this isn’t how mathematics is actually done: ‘ZFC + LCs’ plays the fundamental role just sketched; the Peano Axioms, the Group Axioms, the axioms for any number of particular branches of mathematics, are all preferred over a vast store of alternative possibilities. Resnik puts the concern this way:

The deductivist position on the choice of axioms conflicts with actual mathematical practice. According to the deductivist, it would be perfectly legitimate for mathematicians to make up axiom sets through some random method and then proceed to investigate their logical properties. But mathematics does not proceed in this way . . . For example, we would not develop a set theory with the negation of the pair-set axiom, although it is possible to set up a set theory in which the negation of the pair-set axiom is consistent with the other axioms. (Resnik [1980], p. 132)

This objection seems to me entirely cogent: mathematics isn’t just a study of what follows from what; it also involves judicious selection of preferred concepts and assumptions to play the ‘if’ role in its if-thens. Resnik’s example involves an axiom—why should the set theorist adopt the Axiom of Pairing?—but many of the if-thenist’s conditionals hinge on preferred concepts or definitions—‘if x is a group, then x has a unique identity’—and the corresponding question—why should we study groups (rather than some odd but consistent variant)?—is just as important, and just as troublesome to the Simple If-thenist. Let’s now consider how the view might be enhanced to take these facts this into account.

of what exactly is being asked of set theory for the business of ‘founding’ (see [2017b], [2019]), but leaving these subtleties aside, Ernst has shown that no workable account of such ‘unlimited categories’ is possible—e.g., there can be no category of all graphs in any such foundational scheme (see Ernst [2015])—so this isn’t a strike against set-theoretic foundations in particular. Let me also note that even if set theory were removed from its foundational role, it would presumably continue as a branch of mathematics in its own right (though its proper methods might well shift in that circumstance).

26. The paraconsistent logician would admit even inconsistent theories!

3. ENHANCED IF-THENISM

Resnik believes that what differentiates the preferred assumptions from the merely consistent ones is that they are true:

Once one can distinguish between consistent axiomatic systems, one can make room for unconditional truth in mathematics . . . the mathematicians themselves . . . say: ‘Sure, set theory with the negation of the pairing axiom is consistent, but it is not true.’ (Resnik [1980], p. 133)

Appealing to truth wouldn’t immediately solve the problem of concept-formation, but let’s leave that aside. Resnik imagines the if-thenist responding to this challenge by locating the source of mathematical preferences elsewhere: as a matter of ‘the aesthetics of mathematics rather than its truth’ (ibid.).²⁷

This is certainly a style of response one hears, especially informally. All too often, it’s invoked as a way of dismissing the question of what supports these preferences by removing it from the foundations or philosophy of mathematics and relegating it to some other inquiry. That other inquiry might be systematic aesthetics (see below), but the dismissal often goes further, belittling the preferences as ‘merely aesthetic’, understanding this to entail ‘purely subjective’, and shifting the relevant questions to psychology or sociology. This might do for the mathematician, whose main goal is to deflect distracting philosophical challenges, but a philosopher or methodologist out to understand the practice of mathematics can hardly be satisfied: the processes of concept-formation and axiom choice are among the most far-reaching and consequential in all of mathematics; to legislate them out of mathematics entirely is just an unpersuasive redrawing of boundaries.

Still, mathematicians do sometimes make more thoughtful use of aesthetic language to describe what they do. For example, von Neumann writes that the contemporary pure mathematician

27. Following Quine and Putnam, Resnik eventually rejects this if-thenist response by appeal to the (purported) inseparability of mathematics from natural science (a line of thought I hope to have debunked in §2 and the references cited there).

has a wide variety of fields to which he may turn, and he enjoys very considerable freedom in what he does with them . . . I think that it is correct to say that his criteria of selection, and also those of success, are mainly aesthetical . . . mathematical ideas originate in empirics, although the genealogy is often long and obscure. But, once they are so conceived, the subject begins to live a peculiar life of its own and is better compared to a creative one, governed by almost exclusively aesthetical motivations, than to anything else, and, in particular, to an empirical science. (von Neumann [1947], pp. 2062–2063)

Similar remarks can be found even among natural scientists. For example, consider this from Dirac:

It is more important to have beauty in one's equations than to have them fit experiment . . . It seems that if one is working from the point of view of getting beauty in one's equations, and if one has really sound insight, one is on a sure line of progress. (Dirac [1963], p. 241)²⁸

Recently, some observers have made positive efforts to understand how considerations of beauty might function and with what level of rationality (see Montano [2014]; Kennedy and Väänänen [2015]).

I haven't followed this path myself, partly out of ignorance of aesthetic theory, but also partly for methodological reasons: it seems to me best to start from the mathematics itself, to identify and evaluate the factors behind our preferences directly, in concrete cases, before considering any theoretical superstructure. All too often in the history of the subject, direct philosophical

28. The context of this quotation tells of Schrödinger's early worries about his famous equation, when it seemed to conflict with experiment. Eventually, when electron spin was discovered, that conflict was shown to be merely apparent. In the omitted portion of the above quotation, Dirac remarks that 'if Schrödinger had been more confident of his work, he could have published it some months earlier, and he could have published a more accurate equation' (Dirac [1963], p. 241). According to Dirac, Schrödinger shared his view on the centrality of mathematical beauty: 'Schrödinger and I both had a very strong appreciation of mathematical beauty, and this appreciation of mathematical beauty dominated all our work. It was a sort of act of faith with us that any equations which describe fundamental laws of Nature must have great mathematical beauty in them' (as quoted in Olive [1998], p. 89). See also: 'The research worker, in his efforts to express the fundamental laws of Nature in mathematical form, should strive mainly for mathematical beauty' (as quoted in Pais [1998], p. 36).

study of mathematics has been sidetracked by analogies—mathematics is like natural science, mathematics is like a game, mathematics is like fiction—when it could well turn out that an accurate picture of mathematics would see it as like nothing but its unique self.²⁹ In any case, the mathematical virtues that emerge in this sort of ground-up investigation might turn out to be ‘aesthetic’ in some important sense, and a general aesthetic theory might ultimately illuminate how this works, but at least to begin with, it seems to be best to allow the mathematics to speak for itself.

So, on what grounds do we prefer some axioms, some concepts to others?³⁰ We can all agree, for example, that the concept of ‘group’ is preferable to many consistent alternatives.³¹ It’s commonplace to express this by saying that ‘group’ is a ‘fruitful’ concept, but we can in fact do better than that: we can specify what goals abstract algebra was out to serve (‘to unify various seemingly diverse and dissimilar mathematical domains’ (Kline [1972], p. 1157)), and note that ‘group’ does this job significantly better than the alternatives. To take an example that actually played a role in the historical development, an early switch from requiring cancellation laws to requiring inverses was motivated by the desire to bring important, newly discovered infinite groups into play. So the rationality of preferring ‘group’ to its near neighbors is a simple case of means-ends reasoning: if you want a concept to do this mathematical job, go for ‘group’.

Similarly, if you want your set theory to found classical mathematics³²—in the weak sense, identified earlier, of providing an all-inclusive arena in which

29. See Smith [2014] for some congenial skepticism that an aesthetic sense is what’s at work here. His case study concerns concept-formation (the fractional derivative).

30. I hope I can be forgiven for passing quite briefly over the important considerations in the next three paragraphs. They are treated at length in [1997] and [2011b].

31. For a bit more on the development of the group concept, with references, see [2007], IV.3.

32. Linnebo ([2017], pp. 52–55)—whose account of if-thenism incidentally follows Putnam in emphasizing structures over logical implication—lodges this objection to the view: if-thenism is ‘legitimate’ for ‘structural’ axioms, like the group axioms, but not for ‘foundational’ axioms, like ZFC. The argument seems to be that set theory can’t guarantee that a given theory is ‘in good standing’ (i.e., consistent) unless the axioms of set theory have some special status. But why can’t we simply derive the consistency of the theory from the axioms of ZFC in ordinary if-then style? (This isn’t a guarantee, of course, but we know better than to ask for more than relative consistency.) The discussion in the text indicates that for some axioms to be ‘structural’ (really definitions) and some ‘foundational’ (playing the role set theory plays) is just a matter of the mathematical jobs we devise them to do (see also [2007], pp. 352–355).

all coherent structures are realized³³—then you want it to be as generous as possible, so as to avoid any unnecessary limits on the reach of pure mathematics. Faced with a choice between Gödel’s Axiom of Constructibility ($V = L$) and Large Cardinal Axioms,³⁴ it’s enough to note that $V = L$ rules out various structures that Large Cardinals admit, but that Large Cardinals preserve all the structures of $V = L$, now as the theory of the inner model L . Assuming sufficient evidence for consistency, since inconsistency scotches all benefits, there’s every reason to prefer $ZFC + LCs$ to $ZFC + V = L$.

This line of thought takes our preferences for the ‘if’ part of the if-thens to be based in concrete mathematical facts and goals (not mere psychology or sociology) and explains the propriety, the rationality, of those preferences without appeal to truth (as the mathematical realist would have it). What’s wrong with Simple If-thenism is that it leaves out this whole realm of important mathematical activity, so the enhancement I propose is to add it back in: mathematics is a matter of figuring out what follows from what, where the concepts and axioms in the ‘if’ part are chosen with an eye to facilitating important mathematical goals.³⁵

Happily, this Enhanced If-thenism also has the resources to turn aside yet another of the familiar objections to Simple If-thenism. The objection in question appears most forcefully in Quine’s discussion of ‘truth by convention’ (Quine [1936]). He’s considering the Huntington axioms for Euclidean geometry, which involve as primitives only ‘sphere’ and ‘includes’; the rest are defined in terms of these. Now suppose that φ is some statement of Euclidean geometry expressed in these terms. Quine entertains the position that φ

33. As opposed to, say, a foundation that purports to provide the fundamental concepts and methods for all of mathematics. Again, see [2017b], [2019] for more.

34. $ZFC + V = L$ implies that there are no measurable cardinals, so we can’t have both.

35. In [2011b], p. 99, this is called ‘sophisticated if-thenism’ and it bears obvious kinship to the Arealism proposed there for the special case of set theory. Of course there may well be other ways of ‘enhancing’ if-thenism, that is, other ways of accounting for our preferences that don’t involve truth. For the sake of concreteness, the focus here is on this particular type of enhancement: means-ends assessment in pursuit of mathematical advantage.

insofar as it is conceived as a mathematical truth, is to be construed as an ellipsis for ‘if [the conjunction of the Huntington axioms] then ϕ ’, [which is] a truth of logic. (Quine [1936], p. 82)

This is Simple If-thenism. Quine goes on to object:

The body of all such hypothetical statements, describable as the ‘theory of deduction of non-mathematical geometry’, is of course a part of logic; but the same is true of any ‘theory of deduction of sociology’, ‘theory of deduction of Greek mythology’, etc., which we might construct in parallel fashion with the aid of any set of postulates suited to sociology or to Greek mythology. (Ibid., p. 83)

Quine concludes:

The point of view toward geometry which is under consideration thus reduces merely to an exclusion of geometry from mathematics, a relegation of geometry to the status of sociology or Greek mythology; the labeling of the ‘theory of deduction of non-mathematical geometry’ as ‘mathematical geometry’ is a verbal *tour de force* which is equally applicable to the case of sociology or Greek mythology . . . we are not interested in renaming. (Ibid.)

Of course, the idea that there is something properly called ‘non-mathematical geometry’ with the same status as sociology, say, isn’t really a defect—we do now believe that physical geometry is empirical—but to put both physical geometry and sociology on a par with Greek mythology is more than merely troublesome. More to the point, in an effort to avoid questions of truth and existence, the Simple If-thenist turns mathematics into logic, but then seems unable to block the same move for everything else.³⁶ Surely, we don’t want to remove all or even large chunks of empirical science from the realm of truth and existence!³⁷

36. Following Quine, we engage here in the fantasy that axiomatizing a natural science is a straightforward affair, which of course it isn’t.

37. Well, some do, but not I.

Now consider this problem from the point of view of our Enhanced If-thenist: she reduces mathematics not just to what follows from what, not just to logic, but to logic supplemented with a mathematically rational process of selecting concepts and assumptions for the ‘if’ part of the if-thens. The counterpart for sociology or biology or physics would be to say that all its claims are really logical if-thens, and there’s a rational, properly sociological, biological, or physical inquiry into what’s appropriate in the ‘if’ clause. This boils down to the fairly mundane observation that a science can be regimented as a set of assumptions and their logical consequences,³⁸ which isn’t particularly surprising; most of the effort in fact goes into isolating and defending those ‘assumptions’ for the ‘if’ part—formulating and confirming the theory in question—and the ‘logical consequences’ part ordinarily goes without saying. Given the central role of proof in contemporary mathematics, there’s some point in this case to highlighting that aspect—proving theorems is, after all, a large portion of what mathematicians do—but the point remains that the same general format seems to work quite broadly. If it follows that truth and existence are irrelevant to mathematics, doesn’t the same follow for all that rest?

The answer lies in the ‘rational, properly sociological, biological, or physical inquiry’ that grounds preferences for the respective sociological, biological, or physical ‘if’ clauses: this is just a convoluted way of describing our empirical investigation into the sociological, biological, or physical facts. In these cases, we devise methods to uncover those facts—beginning with ordinary observation, extending to controlled experimentation, theory formation and testing, and so on—and we constantly reflect back on our methods to assess their reliability and to improve them. Our standard for assessing those methods is how well they track the truth about the existing worldly situations in question, be they sociological, biological, subatomic, or cosmological. What makes mathematics different from the rest isn’t that it can be formed into if-then statements—pretty much anything can³⁹—but the standards of assessment for what goes into the ‘if’ clauses—these standards are different. Our decisions about the mathematical ‘if’ clauses involve the

38. Recall footnote 36.

39. Again, see footnote 36.

local and global mathematical goals of the practice—as group theory aims to draw together disparate structures in illuminating ways, as set theory aims to found classical mathematics—but they don’t involve the existence of abstracta or the truth of claims about them. So when mathematical if-thenism is enhanced, this doesn’t yield the same ontological or epistemic conclusions as it does for sociology or biology or physics. Mathematics is different from the natural sciences.⁴⁰

Interestingly enough, the Quine of ‘Truth by convention’ seems to reach this same conclusion—that mathematics is different—by a different route. Though he’s rejected Simple If-thenism out of hand, he seriously entertains a variant: regard the conjunction of the Huntington axioms as true by convention; this, combined with the logical conventions,⁴¹ yields our theorem φ as true by convention as well. Then comes the familiar objection: this can also be done for ‘the so-called empirical sciences’ (Quine [1936], p. 100). Indeed,

in thus elevating the statement from putative to conventional truth, we still retain the right to falsify the statement tomorrow if those events should be observed which would have occasioned its repudiation while it was still putative: for conventions are commonly revised when new observations show the revision to be convenient. (Ibid., pp. 101–102)

A few years later, in the classic essays ‘Two dogmas’ (Quine [1951]) and ‘Carnap and logical truth’ (Quine [1954]), considerations like these lead Quine to question the very distinction between conventional and ‘putative’, but here, back in 1936, he takes a different tack:

There is the apparent contrast between logico-mathematical truths and others that the former are a priori, the latter a posteriori . . . Viewed behavioristically and without reference to a metaphysical system, this contrast retains reality as a contrast between more and less firmly accepted

40. And from Greek mythology: settling the ‘if’ part in that case is essentially forming the myth; whatever constraints guided that process, they weren’t the search for mathematical advantage.

41. Of course, the punch line of Quine [1936], which comes later in the paper, is that logic can’t be treated as conventional, for Lewis Carroll-like reasons.

statements . . . There are statements which we choose to surrender last, if at all . . . and among these there are some which we will not surrender at all, so basic are they to our whole conceptual scheme. Among the latter are to be counted the so-called truths of logic and mathematics, regardless of what further we might have to say of their status in the course of a subsequent sophisticated philosophy. (Quine [1936], p. 102)

This difference serves as motivation for the move to a conventionalist variant of if-thenism for the case of mathematics:

Now since these statements are destined to be maintained independently of our observations of the world, we may as well make use here of our technique of conventional truth assignment and thereby forestall awkward metaphysical questions as to our a priori insight into necessary truths. On the other hand this purpose would not motivate extension of the truth-assignment process into the realm of erstwhile contingent statements. On such grounds, then, logic and mathematics may be held to be conventional while other fields are not; it may be held that it is philosophically important to circumscribe the logical and mathematical primitives by conventions of truth assignment but that it is idle elaboration to carry the process further. (Ibid., pp. 102–103)

And what does Quine himself think of this idea?

Such a characterization of logic and mathematics is perhaps neither empty nor uninteresting nor false. (Ibid.)

Was there ever a more delightfully hedged assessment?! Of course, the denouement is that conventionalism cannot be sustained for logic, but it appears that the reduction of mathematics to conventions plus logic remains a live option for Quine at this point.⁴²

42. To be honest, I don't see why Quine couldn't have given the same style of response on behalf of the Simple If-thenist (in the passages quoted a few pages back from p. 83 of Quine [1936]), leaving that view live as well.

Though this early Quine and the Enhanced If-thenist may agree that this type of objection can be turned away with some version of the idea that mathematics is different from science, I hope it's clear that those versions and the respective routes to them are not the same. Quine is motivated by a philosophical puzzle: mathematics is *a priori*; how are we to account for it, metaphysically and epistemologically? His conventionalist version of Simple If-thenism presents itself as a way of 'forestalling' these difficult questions. In contrast, the Enhanced If-thenist is simply out to describe mathematics as she sees it practiced: proving theorems from assumptions is perhaps the most conspicuous activity involved, but looking more closely she sees the importance of the selection of those ifs. It's only her examination and assessment of the methods there—her developed understanding of what's at stake, what guides and constrains these practices—that reveals how fundamentally they differ from theory formation and testing in the empirical sciences. So she and Quine agree that mathematics is different from science, but for very different reasons. As a result, they turn away the objection that natural science, too, is open to 'if-then' treatment in very different ways. He thinks there's simply no motivation to extend that treatment to the natural sciences. She thinks we do so extend it when we codify an empirical science into an explicit theory, which is often something quite worth doing, but for her, that process itself—ultimately an empirical investigation—reveals what separates mathematics from the rest: it's in the methods of 'if' isolation, which float free of ontology and truth in mathematics but not in empirical science.

Assuming, then, that our Enhanced If-thenist has plausibly launched her attempted reduction of mathematics to logic and the isolation of ifs, there remains the question left open at the end of §1: what is logical truth and how do we know it?

4. THE STATUS OF LOGIC

Accounting for logical truth is a difficult problem for anyone, but Resnik suggests that it's especially difficult for the if-thenist:

Part of the usual justification for accepting a rule of deductive inference is that following it invariably leads from premises that are true . . . to conclusions which also are true. But since deductivism provides for no concept of mathematical truth beyond logical truth, it can be sure only of the correctness of the logical rules in their nonmathematical applications. Because such applications presumably are restricted to finite physical structures . . . doubts about the use of logic in application to abstract infinite structures—even those only hypothetically existing—remain unanswered. (Resnik [1980], p. 136)

This is clearly an existential threat to if-thenism of any variety: what justification is there for the logical transition in the ‘if-then’? Let’s begin by thinking about the logic appropriate in direct application to the physical world (Resnik’s ‘finite physical structures’).

So, to begin, what is the logic of the world?⁴³ Our Enhanced If-thenist⁴⁴ looks around herself and discerns what certainly appears to be a world made up largely of individual objects—stones, trees, stars—with various properties—the cat is black, the house is more than 100 years old—standing in various relations—Mount Everest is taller than Mount Baldy, the book is closer to the table than to the floor—and some situations that depend on others—this flower is larger than that one because I fed a well-balanced plant food to the first plant and not the second. Scientific examination of these various commonsense beliefs largely ratifies them.⁴⁵ The result is a confirmed belief that the world exhibits a wide range of instances of what I’ve called KF-structure:⁴⁶ a domain of individuals, with properties, standing in relations, where some of these situations amount to conjunctions, disjunctions, or negations of others, some hold universally, and some depend on others. There’s every

43. The next three paragraphs attempt to condense the position set out in Part III of [2007]. Essay #8 is a more leisurely summary (with some additions). [2014b] explores the position in compare-and-contrast with Wittgenstein’s views, both early and late.

44. From here on, I often leave the ‘Enhanced’ implicit, trusting to the possessive ‘our’ and the capital ‘I’.

45. For the existence of these individual objects, I have in mind here investigations that show, e.g., that the atomic structure inside the area we take to be occupied by the stone is indeed quite different from what’s found outside, that various forces conspire to hold these structures together as they move, to make them resist penetration, etc. Of course facts like these don’t keep some from denying the existence of ordinary objects (see Essay #8, pp. 219–220) for some discussion of this point).

46. For Kant-Frege, from whom the notion derives.

reason to suppose that these properties and relations are sometimes indeterminate—there are tadpoles and there are frogs, but between these there's a vague transitional phase; there's a fuzzy boundary between sloughing hairs that are still part of the cat and those that aren't—and that this indeterminacy rises into compound situations in the familiar three-valued way: for example, $\text{not}-(\dots)$ obtains if $(\dots)$ fails, fails if $(\dots)$ obtains, and is indeterminate if $(\dots)$ is.⁴⁷ Since dependencies presuppose some connection or other between two situations, they can't be treated compositionally.

This much logical structure validates a rudimentary logic that includes many classical inferences involving 'not', 'and', 'or', and the quantifiers—for example, double negation elimination, the DeMorgan laws, the distributive laws, universal instantiation—as well as modus ponens; but there are no simple logical truths (e.g., $(\dots)$ or $\text{not}-(\dots)$ is indeterminate when $(\dots)$ is) and some familiar inferences, like modus tollens, are invalid (e.g., if p grounds q and q fails, all that follows is that p can't hold—it might be indeterminate). The power and simplicity of full classical logic is achieved by idealizing away both the indeterminacies, assuming all properties and relations are determinate, and ignoring the subtleties that separate dependencies from a material conditional. These idealizations are justified in a given application in just the way any scientific idealization is justified: if they provide significant benefits and don't distort the situation in ways relevant to the matter at hand. So it's not enough for a supporter of a deviant logic that rejects the law of excluded middle or installs a different style of conditional to point out that classical logic has falsified these things; that's what an idealization does. What's needed is a persuasive argument that the relevant falsification is damaging, perhaps in a particular kind of case, and that the proposed alternative logic is more effective than classical logic applied with sensitivity and care.⁴⁸ It isn't clear that such a case has ever been successfully made.

47. For the rest, $(\dots)$ and $(__)$ obtains if both do, fails if either does, and is otherwise indeterminate; $(\dots)$ or $(__)$ obtains if one does, fails if both do, and is otherwise indeterminate; $(\text{every } x)(\dots x \dots)$ obtains if $(\dots a \dots)$ does for every individual in the KF-structure, fails if there is an individual there for whom $(\dots a \dots)$ fails, and is otherwise indeterminate. These clauses mirror the so-called Strong Kleene or Łukasiewicz connectives.

48. Williamson [1994] also advocates the use of classical logic despite the fact of vagueness, but he does so by embracing 'epistemicism': for borderline Joe, there's a fact of the matter about whether or not he's bald; we're just unable, in principle, to know which it is. My alternative suggestion is that we apply classical logic in vague situations with some care and sensitivity, just as we exercise care and sensitivity

Given that the laws of rudimentary logic hold in any KF-structure, those laws are valid in any worldly situation with that structure. So, whenever the if-thenist has good evidence for the presence of KF-structuring, she has good evidence that the laws of rudimentary logic are straightforwardly truth-preserving—and she has such evidence in many, many cases. (One conspicuous exception is the quantum world, where the apparent dependencies, properties, and even objects don't behave in familiar KF-fashion, and there she finds rudimentary logic is unreliable.) In addition, given the elementary arithmetical identities among the truths of logic (as in §2), we can now describe them more precisely as appearing among the literal validities of rudimentary logic, as literal truths about much of the world. This fills in the picture of applications from §2: much of applied mathematics enlists mathematical structures as abstract models that resemble phenomena in important ways, but elementary arithmetic straightforwardly describes the world's KF-structures. As for full classical logic, even in KF-structured cases, it needs to be applied with caution, like any idealized theory.

So, returning to Resnik, his worry begins from the assumption that 'logic' takes us from truths to truths when applied to finite physical structures; if 'logic' is understood as rudimentary logic and 'finite physical structures' as KF-structures, what we've seen so far is that our If-thenist agrees. The question is what justifies the use of logic in the mathematical if-thens, in connection with 'abstract infinite structures—even those only hypothetically existing' (Resnik [1980], p. 136). We've seen how the Enhanced If-thenist offers an account of what properly guides our preferences for the 'if' clauses, locating the grounds for those decisions in means/ends rationality, and thus favoring ifs formulated in terms of concepts like 'group' and set-theoretic axioms like Choice and Large Cardinals. In her pursuit of such mainstream contemporary mathematics, this If-thenist also prefers classical logic to its various alternatives. What justification does she have for this preference?

The answer, it seems to me, is that 'hypothetically existing . . . abstract infinite structures' presented in the favored 'if' clauses have been imagined as straightforwardly KF. In fact, they are more precise than most real-world

in applying any idealized theory: neglect friction carelessly and you predict that humans can't walk; neglect vagueness carelessly and you generate a sorites paradox.

KF-structures, because they admit no vagueness and the material conditional is entirely appropriate. This means that, as hypothesized, they are subject not only to rudimentary logic but to full classical logic. So the appropriateness of applying classical logic in our thinking about them rests on the very same considerations that supported the relevant ‘if’ clause in the first place: structures like this are effective ways of realizing our mathematical goals, of gaining mathematical advantage. This doesn’t in any way rule out that other mathematical goals and other important mathematics could be facilitated, for example, by using a constructivist logic; there are clearly areas of mathematics that benefit from that alternative choice. Just as she embraces a number of different important options for her ‘if’ clauses, our If-thenist sees the merits of using different logics for different purposes, but the underlying criterion of selection is always the same: means/ends reasoning in pursuit of mathematical advantage.⁴⁹

5. FROM LOGIC TO ARITHMETIC

We’ve seen how the Enhanced If-thenist locates elementary arithmetical identities among the literal validities of rudimentary logic, not among the mathematical claims taken to be consequents of disguised if-thens. But it’s well-known that this logical treatment of arithmetic can’t be extended to its quantified claims: that no number is smaller than zero, that addition is commutative, that no number has more than one successor.⁵⁰ Many of these can be thought of as straightforward generalizations from the observed facts, but eventually we want a more systematic treatment. Presumably the world’s KF-structures are all finite and indeed bounded by some large finite size—the number of particles in the observable universe is thought to be smaller than 10^{80} —so the ordinary facts, individual and general, about the world’s number properties are also bounded. The trouble is that any attempt at a

49. Obviously, this line of response may not be open to other types of Enhanced If-thenist—nothing much can be said without knowing how their distinctive ‘enhancements’ would go (see footnote 35)—but a first thought would be that the preference for a given logic would be based on the same considerations as the preference for certain ‘ifs’.

50. See Essays #9 and 10 for more on the following approach to arithmetic.

comprehensive theory limited to just this literal part of arithmetic would be too cumbersome to contemplate.

What we do instead is abstract away from the bounded finiteness of actual KF-structures. We figure that ‘in principle’, there could be arbitrarily large KF-structures, arbitrarily large numbers;⁵¹ we then concoct an abstract mathematical model of that structure, with the ‘...’ in ‘1, 2, 3, ...’ as codified in the informal Peano Axioms (PA). In this way, elementary arithmetic—a portion of rudimentary logic, grounded in the world, literally true—is embedded in mathematical arithmetic, a branch of abstract mathematics. Of course, for the If-thenist, there is no actual model, no existing omega sequence; there’s just the description, given by the informal Peano Axioms, of what such a structure would be like, installed in the ‘if’ of the arithmetic if-thens.

In this way, our If-thenist’s informal PA delivers as thens such rudimentary truths as ‘ $2 + 2 = 4$ ’ and ‘addition is commutative’. Idealized as it is, it also delivers various welcome falsehoods, like ‘every number has a successor’. There’s no a priori guarantee that the thens of these if-thens won’t also deliver unwelcome falsehoods, for example, that it won’t come into conflict with the rudimentary facts of elementary arithmetic. As with other cases of applied mathematical modeling—for example, when the facts about actual liquids are embedded in an abstract mathematical theory of continuous substance—the viability of embedding literal numerical truths in the abstract mathematical theory of the potential infinite must meet the test of experience; the model must demonstrate both its safety and its effectiveness. Like fluid dynamics, mathematical arithmetic has passed this test brilliantly. This is what justifies the Enhanced If-thenist’s fixing on informal PA for the ‘if’ parts of her abstract systematization of worldly numerical facts.⁵²

51. Developmental research strongly suggests that we first come to believe that ‘in principle’ we can generate arbitrarily large numerical expressions, a conviction grounded in the recursive nature of the language-learning device, and infer from this that ‘in principle’ there could be arbitrarily large KF-structures (see [2014a], Essays ##9 and 10).

52. [2014a] (also, briefly, Essay #9, footnote 51) highlights some differences between the considerations that support PA and those that support, e.g., ZFC and its extensions. Though [2014a] and Essay #10 leave open the possibility of attributing truth to arithmetic beyond the elementary, this isn’t obligatory, so I ignore it here to focus on a stricter form of if-thenism.

6. THE STATUS OF META-MATHEMATICS

With this understanding of arithmetic, we can face one final line of objection to if-thenism. It begins with the simple observation that it would be pointless to pursue an inconsistent theory. This might create some difficulty for a Simple If-thenist, but the Enhanced If-thenist naturally prefers consistent ‘if’ clauses, and the ‘pointlessness’ worry itself gives her good reason for this preference. The trouble is that the consistency of a given ‘if’, viewed meta-mathematically, is apparently itself a mathematical claim. Again Resnik gives this worry clear voice:

If [the deductivist] is to believe that . . . some mathematical theories are non-trivial, he must commit himself to their unconditional consistency and thereby to some unconditional mathematical truths. (Resnik [1980], p. 119)

Concerns of this sort are exacerbated as meta-mathematics proceeds, with objections based on Gödel’s incompleteness theorems, on the independence the Continuum Hypothesis (CH) from ZFC, and so on: if the Enhanced If-thenist doesn’t believe the relevant meta-mathematical theorems are unconditionally true, why should she give up on an arithmetical proof of the Gödel sentence of first-order arithmetic or on settling CH in ZFC?

To evaluate this objection, we need to assess more carefully what our If-thenist actually believes about consistency. The question arises once she’s installed informal PA in those ‘if’ clauses—generating a practice of logically deriving ‘then’s’, in natural language, with varying degrees of rigor and completeness. The consistency concern is just this: will this practice, properly executed, produce proofs with conflicting results; will it lead to the dead-end of a contradiction? Part of her case for the safety and effectiveness of this abstract systematization, noted at the end of the previous section, is her conviction that the answer is no. Why does she believe this? Perhaps at least partly because she has confidence that she and her fellows share a clear intuitive picture of a potentially infinite sequence that underlies informal PA,⁵³

53. [2014a] and Essays ##9 and 10 take this picture to be grounded in the language-learning device (see footnote 51).

that this picture is KF and thus amenable to classical logic, and thus that no harm should come of tracing ever more complicated consequences. She might also draw on the community's long experience with informal Peano arithmetic, on her own firsthand familiarity with this and related systems, on the apparent absence of familiar sources of inconsistency, and the like. So far, though, this is an ordinary, somewhat tentative belief about the features of a particular human practice, not Resnik's properly mathematical belief.

The mathematical consistency claim doesn't appear until she begins to see, with Hilbert, the possibility of investigating this informal consistency more systematically, by mathematizing the situation. This is another piece of applied mathematics: there's a worldly phenomenon of interest—this time the activity of human mathematicians deriving consequences from informal PA—which we undertake to study by concocting a mathematical model (for our If-thenist, a description of a model to fit in the 'if' clauses of a body of if-thens). As so often happens when we prepare to apply mathematics, the worldly phenomenon is first simplified and regimented—the ordinary language of this proving activity is replaced by certain strings of symbols ('formulas'), proofs by certain strings of strings of symbols ('proofs'), and so on—but these are still physical objects (as Hilbert emphasized). If fact, they're straightforward KF-structures, just the sort of thing whose number properties have been embedded in the abstract arithmetic governed by informal PA. This time, though, instead of attending to their number properties, we focus on their structure as strings of symbols and strings of strings of symbols and set out to develop an informal theory of regimented informal proofs.

Just as in the case of number properties, the finiteness of the physical world means that these actual proof-theoretic properties form an unwieldy domain, and once again our systematization abstracts away from the physical or psychological limits to the lengths and complexities of the strings that appear in the practice. This time, we figure that PA's not leading to a contradiction isn't a matter of our limited interest or life span or any other such contingency; we think that PA is consistent 'in principle'. So we countenance indefinitely long formulas and proofs, embedding the ordinary facts in an abstract model that includes the '...' of the potential infinite. And, just as before, our If-thenist relies on a codified description of this model, on some

natural list of axioms for her informal Proof Theory (PT), and in this way generates an applied mathematical theory that serves as an idealized but still effective and appropriate abstract model for the human practice of proving theorems from informal PA. This PT runs parallel to and independent of PA: just as PA speaks of numbers and their successors and includes mathematical induction, PT speaks of strings of symbols, strings of strings of symbols, and their concatenations, and includes its own inductions on formulas and theorems. In this model, informal PA, expressed in natural language, is replaced by a regimented formalization, PA^* , and informal provability from PA by $PA^* \vdash$, which asserts the existence of certain strings of strings. With informal PT in the ‘if’ slot, our If-thenist generates if-thens like: if PT, then $PA^* \vdash \forall x \exists y (y = Sx)$.⁵⁴

At this point, the Enhanced If-thenist’s informal consistency claim—that no ordinary proof from informal PA will lead to a contradiction—is transformed into the metamathematical claim: if PT, then $PA^* \not\vdash 1 = 0$. The claim in the ‘then’ clause of this if-then is a mathematical claim, a claim in the mathematically idealized but still informal PT, perhaps the very claim that Resnik accuses the if-thenist of believing unconditionally. But does she? She’s fairly confident of the informal consistency of informal PA, a non-mathematical claim, and her experience with informal PT leads her to believe that it’s a reliable piece of applied mathematics, a mathematical model whose idealizations are both effective and benign. So the question is: does our If-thenist’s confidence in the informal consistency of PA commit her to similar confidence in the unconditional truth of the mathematical consistency claim?

The answer appears to be no. We’ve seen (in §2) that a mathematical model can do its job without any requirement of truth or existence, so simply modeling the informal consistency claim via the formal consistency claim isn’t enough by itself to show that she must believe the latter. In fact, what our If-thenist actually believes is this: if PT is a good enough model to reflect this particular part of the actual practice of informally proving from PA, then

54. In a small abuse of notation, I implicitly assume that in contexts like ‘ $PA^* \vdash$ ’, the item on the right-hand-side of the turnstile has also been converted from ordinary mathematical language into a formal string of symbols, parallel to the move from PA to PA^* .

the mathematical consistency statement will appear as the ‘then’ in an if-then with PT as the ‘if’. But as for $PA^* \not\vdash 1 = 0$ itself, it’s still appearing only as the ‘then’ of an if-then.

So far so good for the Enhanced If-thenist, but Gödel’s second incompleteness theorem appears to threaten from the wings.⁵⁵ We normally take it to show that we shouldn’t bother trying to prove $PA^* \not\vdash 1 = 0$ in PT—more or less what Hilbert had hoped for, a proof-theoretic proof of consistency—but if we don’t unconditionally believe the theorem, why should we draw this moral?

In Gödel’s hands, meta-mathematics matured beyond a simple study of (strings of) strings of symbols to the fully arithmetized undertaking we know today. Combining informal PT with informal PA,⁵⁶ we define an encoding of strings as numbers; this allows us to generate strings like $\text{Fla}(x)$, for codes of strings that are formulas of PA^* , and $\text{Pbl}(x)$, for codes of formulas for which there are proofs in PA^* . We now have three levels of consistency claims: the worldly ‘there is no informal proof of $1 = 0$ from informal PA’; the formalized claim in PT that $PA^* \not\vdash 1 = 0$; and the equivalent arithmetized claim in PT + PA that $\neg \text{Pbl}(\#(1 = 0))$. The Enhanced If-thenist’s understanding of Gödel’s second theorem for PA^* is then: If PT + PA, then $PA^* \not\vdash 1 = 0$ implies $PA^* \not\vdash \neg \text{Pbl}(\#(1 = 0))$.

What is our If-thenist to make of this? What morals should she draw from this theorem about the abstract model PT + PA? To begin with, PA and PT idealize real-world number properties and real-world strings in parallel ways, with the same characterization of the crucial ‘...’, so combining the two in a straightforward way shouldn’t change the Enhanced If-thenist’s assessment of the strengths and weaknesses of the idealizations involved. This means that, despite the fiction that numbers and formulas and proofs can be indefinitely large, she has in PT + PA what looks like an effective and benign model of actual number properties, statements, and proofs. Since it’s a good model,

55. There are anti-if-thenist arguments based on the first incompleteness theorem, but I don’t think they raise any additional issues. (Some of these assume that the if-thenist is out to define, e.g., ‘ σ is true in arithmetic’ as ‘PA implies σ ’, but our If-thenist isn’t defining ‘mathematical truth’; she’s eschewing it.)

56. Notice that this combination, which I’ll refer to as ‘PT + PA’ below, isn’t just the joint list of their individual axioms; it also includes, e.g., the functions from the domain of PT to the domain of PA that are involved in Gödel numbering.

the fact that some claim holds in that model is a fairly good indicator that the corresponding claim holds in the world. So the question is: for what worldly situation is ‘ $PA^* \not\vdash 1 = 0$ implies $PA^* \not\vdash \neg Pbl(\#(1 = 0))$ ’ an abstract surrogate?

‘ $PA^* \not\vdash 1 = 0$ ’ is easy: this is the mathematical consistency claim, the model’s surrogate for the informal consistency of informal PA. ‘ $PA^* \not\vdash \neg Pbl(\#(1 = 0))$ ’ takes just a little more unpacking: in $PT + PA$, ‘ $\neg Pbl(\#(1 = 0))$ ’ is a coded equivalent of $PA^* \not\vdash 1 = 0$,⁵⁷ so $PA^* \not\vdash \neg Pbl(\#(1 = 0))$ is the model’s arithmetized surrogate for ‘PT doesn’t informally prove $PA^* \not\vdash 1 = 0$ ’.⁵⁸ Putting the two parts together, the worldly counterpart to ‘ $PA^* \not\vdash 1 = 0$ implies $PA^* \not\vdash \neg Pbl(\#(1 = 0))$ ’ is ‘if PA is informally consistent, then informal PT doesn’t informally prove $PA^* \not\vdash 1 = 0$ ’. Insofar as we trust the model $PT + PA$, we should believe this worldly counterpart; since we also believe that PA is informally consistent, we shouldn’t bother trying to informally prove $PA^* \not\vdash 1 = 0$ from PT. And this is exactly the moral we had hoped to draw.⁵⁹ The fact that PT doesn’t capture the formal counterpart ($PA^* \not\vdash 1 = 0$) to something we believe unconditionally about informal proving from PA (that it won’t lead to $1 = 0$) is a disappointment: we can trust what the model tells us, but we can’t count on it to tell us everything. This is a shortcoming of the model, but not a challenge to the viability of Enhanced If-thenism.⁶⁰

So in the end, none of the standard objections seems to me to be effective against Enhanced If-thenism. It recognizes the centrality of proof in contemporary pure mathematics as well as the subtle mathematical considerations that go into a judicious choice of concepts and axioms. It supports the use of classical logic in most cases and preserves the unconditional truth of elementary arithmetic. What it doesn’t do is raise extraneous metaphysical or epistemological questions about abstracta or attempt to explain away serious

57. In $PT + PA$, it’s easy to see that $Pbl(\#(1 = 0))$ iff there is an n in ω that codes a formal proof of $1 = 0$ from PA^* iff $PA^* \vdash 1 = 0$.

58. A complete version of this argument would require an explicit formulation of PT and $PA + PT$ and a proof that $PA^* \vdash \neg Pbl(\#(1 = 0))$ iff $PT^* \vdash (PA^* \not\vdash 1 = 0)$. Thanks to Will Stafford for verifying that this can be done.

59. Related considerations show that we shouldn’t expect to informally prove CH or not-CH from informal ZFC; insofar as we trust our formal theory of proving from ZFC, these worldly counterparts follow from the theorems of Gödel and Cohen.

60. Again, if the if-thenist’s goal were to define ‘truth’ as ‘that which turns up in the relevant thens’, this would be a problem, but we’ve seen that our If-thenist has no interest in that project (see footnote 55).

mathematical decisions as arbitrary conventions. In short, it avoids both robust accounts of ‘mathematical truth’ and the threat of ‘anything goes’. For all this, I recommend it to your consideration.^{61,62}

61. [[Since this essay was completed, Geoffrey Hellman has published a reply (Hellman [2020]). His paper poses a range of challenges, but most of them spring from his Robust Realism (see [202?a]). He understands me to reject that position ‘as a necessary price worth paying for advantages in avoiding metaphysical and epistemological concerns that beset other philosophical approaches to mathematics’ (p. 251; see also p. 247). It is true that I allude to concerns like these in the introductory paragraphs of the present essay as motivation for exploring if-thenism as an alternative, but in [2011b], where Robust Realism and its fellows are laid out in detail, I emphasize that my second-philosophical objection is prior to, say, Benacerraf’s challenge. My concern is that Robust Realism sets an extra condition on mathematical justification—that a claim be shown to comport with its abstract metaphysics—and that this condition can outweigh considerations of mathematical advantage: ‘Sure, such-and-such an assumption would further our pursuit of deep mathematical goals, but, sadly, it happens to be false on my mathematical metaphysics’. For the Thin Realist and the Arealist (who most closely resembles the Enhanced If-thenist), mathematical considerations by themselves are justification enough, so the objection isn’t that a viable epistemology is hard to find—it’s that it’s inappropriate to impose this epistemological requirement in the first place (see [2011b], pp. 57–59). Robust Realism isn’t surrendered reluctantly under pressure of objections but regarded as a non-starter because it deviates from sound mathematical method.]]

62. My thanks to Bob Geroch, whose persistent questions about the significance of Gödel’s theorems from the point of view of a quite advanced culture of extra-terrestrials inspired the musings that led to this essay, and to Geoffrey Hellman, Juliette Kennedy, Charles Leitz, David Malament, Guillaume Massas, Christopher Mitsch, Stella Moon, Tony Queck, Michael Resnik, and Jeffrey Schatz for comments on earlier drafts. Special thanks to Massas and Moon for extremely helpful discussion of §4 and related issues.

REFERENCES

Adams, Zed

[2016] *On the Genealogy of Color* (New York: Routledge).

Anderson, Barton

[2015] 'Can computational goals inform theories of vision?', *Topics in Cognitive Science* 7, pp. 274–286.

Anderson, Lanier

[2005] 'Neo-Kantianism and the roots of anti-psychologism', *British Journal for the History of Philosophy* 13, pp. 287–323.

Anstey, Peter

[2011] *Locke and Natural Philosophy* (Oxford: Oxford University Press).

Arbeiter, Sophia, and Juliette Kennedy, eds,

[202?] *The Philosophy of Penelope Maddy* (Cham, Switzerland: Springer).

Atherton, Margaret

[1991] 'Corpuscles, mechanism, and essentialism in Berkeley and Locke', *Journal of the History of Philosophy* 29, pp. 47–67.

Austin, J. L.

[1940] 'The meaning of a word', in his [1979], pp. 55–75.

[1946] 'Other minds', reprinted in his [1979], pp. 76–116.

[1950] 'Truth', reprinted in his [1979], pp. 117–133.

[1956a] 'A plea for excuses', reprinted in his [1979], pp. 175–204.

[1956b] 'Ifs and cans', reprinted in his [1979], pp. 205–232.

[1962] *Sense and Sensibilia* (Oxford: Oxford University Press).

[1979] *Philosophical Papers*, J. O. Urmson and G. J. Warnock, eds., third edition (Oxford: Oxford University Press).

Avigad, Jeremy

[2006] 'Methodology and metaphysics in the development of Dedekind's theory of ideals', in J. Ferreirós and J. Gray, eds., *The Architecture of Modern Mathematics* (Oxford: Oxford University Press), pp. 159–286.

Ayer, A. J.

[1936] *Language, Truth and Logic*, second edition (New York: Dover, 1952).

[1946] 'Introduction to the second edition of Ayer [1936]', in his [1936], pp. 5–26.

- Baldwin, Thomas
 [1990] *G. E. Moore* (London: Routledge).
- Barrow-Green, June, and Reinhard Siegmund-Schultze
 [2016] ‘“The first man on the street”—tracing a famous Hilbert quote (1990) back to Gergonne (1825)’ *Historica Mathematica* 43, pp. 415–426.
- Benacerraf, Paul
 [1973] ‘Mathematical truth’, reprinted in Benacerraf and Putnam [1983], pp. 403–420.
 Benacerraf, Paul, and Putnam, Hilary, eds.
 [1983] *Philosophy of Mathematics*, second edition (Cambridge: Cambridge University Press).
- Berkeley, George
 [1709] *New Theory of Vision*, in his *Works on Vision*, C. Turbayne, ed. (Indianapolis, IN: Bobbs-Merrill, 1963), pp. 7–102.
 [1721] *De Motu*, in his [1992], pp. 43–107.
 [1734] *The Analyst*, in his [1992], pp. 157–221.
 [1992] *De Motu* and *The Analyst*, D. Jesseph, trans. and ed. (Dordrecht: Kluwer).
- Biener, Zvi, and Eric Schliesser, eds.
 [2014] *Newton and Empiricism* (New York: Oxford University Press).
- Bloom, Paul
 [2000] *How Children Learn the Meanings of Words* (Cambridge, MA: MIT Press).
- Bolzano, Bernard
 [1837] *Theory of Science*, R. George, ed. and trans. (Berkeley: University of California Press, 1972).
- Boole, George
 [1854] *The Laws of Thought* (New York: Dover, 1958).
- Brookes, Derek
 [1997] ‘Introduction’, in Reid [1764], pp. xii–xxv.
- Broughton, Janet
 [1983] ‘Hume’s skepticism about causal inferences’, *Pacific Philosophical Quarterly* 64, pp. 3–18.
 [1992] ‘What does the scientist of man observe?’, *Hume Studies* 18, pp. 155–168.
 [2002] *Descartes’s Method of Doubt* (Princeton, NJ: Princeton University Press).
 [2003] ‘Hume’s naturalism about cognitive norms’, *Philosophical Topics* 31, pp. 1–19.
 [2004] ‘The inquiry in Hume’s *Treatise*’, *Philosophical Review* 113, pp. 537–556.
 [2005] ‘Hume’s voyage’, in J. Jenkins et al., eds., *Persons and Passions* (Notre Dame, IN: University of Notre Dame Press), pp. 174–191.
- Burge, Tyler
 [1992] ‘Frege on knowing the third realm’, reprinted in his [2005], pp. 299–316.
 [1998] ‘Frege on knowing the foundation’, reprinted in his [2005], pp. 317–355.
 [2000] ‘Frege on a priority’, reprinted in his [2005], pp. 356–387.
 [2005] *Truth, Thought and Reason* (Oxford: Oxford University Press).
 [2010] *Origins of Objectivity* (Oxford: Oxford University Press).
- Burgess, J. A.
 [1990] ‘The sorites paradox and higher-order vagueness’, *Synthese* 85, pp. 417–474.
- Carey, Susan
 [2009] *The Origin of Concepts* (New York: Oxford University Press).
- Carnap, Rudolf
 [1934] *Logical Syntax of Language*, A. Smeaton, trans. (London: Routledge & Kegan Paul, 1937).
 [1950] ‘Empiricism, semantics and ontology’, reprinted in Benacerraf and Putnam [1983], pp. 241–257.

- [1963] 'Intellectual autobiography', in P. A. Schilpp, ed., *The Philosophy of Rudolf Carnap* (La Salle, IL: Open Court, 1963), pp. 1–84.
- Carroll, Lewis
- [1895] 'What the tortoise said to Achilles', *Mind* 4, pp. 278–280.
- Cassam, Quassim
- [2009] 'Can the concept of knowledge be analyzed?', in Greenough and Pritchard [2009], pp. 12–30.
- Chesney, Dana, and Harry Haladjian
- [2011] 'Evidence for a shared mechanism used in multiple-object tracking and subitizing', *Attention, Perception, and Psychophysics* 73, pp. 2457–2480.
- Chomsky, Noam
- [1957] *Syntactic Structures* (The Hague: Mouton).
- Cohen, Jonathan
- [2009] *The Red and the Real* (Oxford: Oxford University Press).
- Cohen, Sheldon
- [1982] 'St. Thomas Aquinas on the immaterial reception of sensible forms', *Philosophical Review* 91, pp. 193–209.
- Coliva, Annalisa
- [2010] *Moore and Wittgenstein* (New York: Palgrave Macmillan).
- [2018] 'What do philosophers do? Maddy, Moore and Wittgenstein', *International Journal for the Study of Skepticism* 7, pp. 217–234.
- Crone, Robert
- [1992] 'The history of stereoscopy', *Documenta Ophthalmologica* 81, pp. 1–16.
- Curley, Edward
- [1972] 'Locke, Boyle, and the distinction between primary and secondary qualities', *Philosophical Review* 81, pp. 438–464.
- Dear, Peter
- [2009] *Revolutionizing the Sciences*, second edition (Princeton, NJ: Princeton University Press).
- De Bary, Philip
- [2002] *Thomas Reid and Scepticism* (London: Routledge).
- Dehaene, Stanislas
- [2011] *The Number Sense: How the Mind Creates Mathematics*, revised and expanded edition (New York: Oxford University Press) (First edition, 1997).
- Dehaene, Stanislas, and Elizabeth Brannon, eds.
- [2011] *Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought* (Amsterdam: Elsevier).
- Descartes, René
- [1632] *Optics*, in his *Discourse on Method, Optics, Geometry, and Meteorology*, revised edition, P. Olscamp, trans. and ed. (Indianapolis, IN: Hackett, 2001), pp. 65–173.
- [1633] *Treatise of Man*, T. S. Hall, trans. and ed. (Amherst, NY: Prometheus Books, 2003).
- [1641a] *Meditations on First Philosophy*, J. Cottingham, trans., in J. Cottingham et al., eds., *The Philosophical Writings of Descartes*, volume II (Cambridge: Cambridge University Press, 1984), pp. 1–62.
- [1641b] 'Letter to Mersenne', A. Kenny and J. Cottingham, trans., in J. Cottingham et al., eds., *The Philosophical Writings of Descartes*, volume III (Cambridge: Cambridge University Press, 1991), pp. 171–173.

- [1644] *Principles of Philosophy*, J. Cottingham, trans; in J. Cottingham et al., eds., *The Philosophical Writings of Descartes*, volume I (Cambridge: Cambridge University Press, 1985), pp. 179–291.
- Diamond, Cora
- [1986] ‘Realism and the Realistic Spirit’, reprinted in her *The Realistic Spirit: Wittgenstein, Philosophy, and the Mind* (Cambridge, MA: MIT Press, 1991), pp. 39–72.
- Dirac, P. A. M.
- [1963] ‘The evolution of the physicist’s picture of nature’, reprinted in M. Kline, ed., *Mathematics in the Modern World* (W. H. Freeman and Co., 1968), pp. 239–247.
- Downing, Lisa
- [1998] ‘The status of mechanism in Locke’s *Essay*’, *Philosophical Review* 107, pp. 381–414.
- [2002] ‘Robert Boyle’, in Nadler [2002], pp. 338–353.
- [2005] ‘Berkeley’s natural philosophy and philosophy of science’, in K. Winkler, ed., *Cambridge Companion to Berkeley* (Cambridge: Cambridge University Press), pp. 230–265.
- [2009] ‘Locke: the primary and secondary distinction’, in R. Le Poidevin et al., eds., *Routledge Companion to Metaphysics* (New York: Routledge), pp. 98–108.
- [2011] ‘Sensible qualities and material bodies in Descartes and Boyle’, in Nolan [2011a], pp. 109–135.
- [2014] ‘Locke’s metaphysics and Newtonian metaphysics’, in Biener and Schliesser [2014], pp. 97–118.
- Eddington, A. S.
- [1928] *The Nature of the Physical World* (New York: Macmillan).
- Einstein, Albert
- [1949] ‘Autobiographical notes’, in P. A. Schilpp, ed., *Albert Einstein: Philosopher-Scientist* (La Salle, IL: Open Court), pp. 1–94.
- Ernst, Michael
- [2015] ‘The prospects of unlimited category theory’, *Review of Symbolic Logic* 8, pp. 306–327.
- Feferman, Solomon
- [1984] ‘Toward useful type-free theories I’, *Journal of Symbolic Logic* 49, pp. 75–111.
- Feigenelson, Lisa
- [2011] ‘Objects, sets, and ensembles’, in Dehaene and Brannon [2011], pp. 13–22.
- Feynman, Richard
- [1985] *QED: the Strange Theory of Light and Matter* (Princeton, NJ: Princeton University Press).
- Field, Hartry
- [1973] ‘Theory change and indeterminacy of reference’, reprinted with a new postscript in his [2001], pp. 177–198.
- [1994] ‘Deflationist use of meaning and content’, reprinted in his [2001], pp. 104–156.
- [2001] *Truth and the Absence of Fact* (Oxford: Oxford University Press).
- Fogelin, Robert
- [1976] *Wittgenstein* (London: Routledge).
- [1987] *Wittgenstein*, second edition (London: Routledge).
- [1993] ‘Hume’s skepticism’, in D. Norton, ed., *Cambridge Companion to Hume* (Cambridge: Cambridge University Press), pp. 90–116.
- [2009] *Hume’s Skeptical Crisis* (Oxford: Oxford University Press).
- Frege, Gottlob
- [1884] *Foundations of Arithmetic*, second revised edition, J. L. Austin, trans. (Oxford: Basil Blackwell, 1980).
- [1891] ‘Function and concept’, reprinted in his [1997], pp. 130–148.

- [1892] 'On Sinn and Bedeutung', reprinted in his [1997], pp. 151–171.
- [1893] *Basic Laws of Arithmetic*, vol. 1, excerpted in his [1997], pp. 194–223.
- [1897] 'Logic', reprinted in his [1997], pp. 227–250.
- [1918] 'Thought', reprinted in his [1997], pp. 325–345.
- [1997] *The Frege Reader*, M. Beaney, ed. (Oxford: Blackwell).
- Galileo Galilei
- [1623] 'The Assayer', in *The Controversy on the Comets of 1618*, S. Drake and C. O'Malley, trans. (Philadelphia: University of Pennsylvania Press, 1960), pp. 151–336.
- Gaukroger, Stephen
- [2002] 'Francis Bacon', in Nadler [2002], pp. 298–307.
- [2014] 'Empiricism as a development of experimental natural philosophy', in Biener and Schliesser [2014], pp. 15–38.
- Gibbins, Peter
- [1987] *Particles and Paradoxes* (Cambridge: Cambridge University Press).
- Goddard, Peter, ed.
- [1998] *Paul Dirac: the Man and His Work* (Cambridge: Cambridge University Press).
- Godfrey-Smith, Peter
- [2008] 'Recurrent transient underdetermination and the glass half full', *Philosophical Studies* 137, pp. 141–148.
- Goldman, Alvin
- [2015] 'Review of H. Kornblith, *A Naturalistic Epistemology* (Oxford: Oxford University Press, 2014), *Notre Dame Philosophical Reviews*.
- Gonzalez, Jessica
- [n.d.] 'Detecting negations', unpublished manuscript.
- Greenough, Patrick, and Duncan Pritchard, eds.
- [2009] *Williamson on Knowledge* (Oxford: Oxford University Press).
- Gregory, Frederick
- [1977] *Scientific Materialism in Nineteenth Century Germany* (Dordrecht: Reidel).
- Hahn, Hans
- [1933] 'Logic, mathematics and knowledge of nature', reprinted in A. J. Ayer, ed., *Logical Positivism* (New York: Free Press, 1959), pp. 147–161.
- Harnett, Patrice, and Rochel Gelman
- [1998] 'Early understandings of numbers: paths or barriers to the construction of new understandings?', *Learning and Instruction* 8, pp. 341–374.
- Hatfield, Gary
- [1988] 'Representation and content in some (actual) theories of perception', reprinted in his [2009a], pp. 50–87.
- [1991] 'Representation in perception and cognition: task analysis, psychological functions, and rule instantiation', reprinted with additions in his [2009a], pp. 88–123.
- [1992] 'Color perception and neural encoding: does metameric matching entail a loss of information?', reprinted in his [2009a], pp. 266–280.
- [1995] 'Remaking the science of mind: psychology as a natural science', in C. Fox, R. Porter, and R. Wokler, eds., *Inventing Human Science: Eighteenth Century Domains* (Berkeley: University of California Press), pp. 184–231.
- [2003a] 'Representation and constraints: the inverse problem and the structure of visual space', reprinted in his [2009a], pp. 153–177.
- [2003b] 'Objectivity and subjectivity revisited: color as a psychobiological property', reprinted in his [2009a], pp. 281–296.

- [2009a] *Perception and Cognition* (Oxford: Oxford University Press).
- [2009b] 'Getting objects for free (or not): the philosophy and psychology of object perception', in his [2009a], pp. 212–255.
- Hatfield, Gary, and Sarah Allred, eds.
[2012] *Visual Experience: Sensation, Cognition, and Constancy* (Oxford: Oxford University Press).
- Hauser, Marc, Noam Chomsky, and Tecumseh Fitch
[2002] 'The faculty of language: what is it, who has it, and how did it evolve?', *Science* 298, pp. 1569–1579.
- Heck, Richard
[2000] 'Cardinality, counting, and equinumerosity', reprinted in his *Frege's Theorem* (Oxford: Oxford University Press, 2011), pp. 156–179.
- Hellman, Geoffrey
[1989] *Mathematics without Numbers* (Oxford: Oxford University Press).
[2020] *Mathematics and Its Logics* (Cambridge: Cambridge University Press).
- Henry, John
[2008] *The Scientific Revolution and the Origins of Modern Science*, third edition (New York: Palgrave Macmillan).
- Hesiod
[70 bc?] *Works and Days*, in his *Theogony and Works and Days*, M. L. West, trans. (Oxford: Oxford University Press, 1988), pp. 35–62.
- Hollingworth, Andrew, and Ian Rassmussen
[2010] 'Binding objects to locations: the relationship between object files and visual working memory', *Journal of Experimental Psychology: Human Perception and Performance* 36, pp. 543–564.
- Hume, David
[1739] *A Treatise of Human Nature*, D. F. and M. J. Norton, eds. (Oxford: Oxford University Press, 2000).
[1748] *An Enquiry concerning Human Understanding*, T. Beauchamp, ed. (Oxford: Oxford University Press, 1999).
- Hurlburt, Russell, and Eric Schwitzgebel
[2007] *Describing Inner Experience* (Cambridge, MA: MIT Press).
- Ichikawa, Jonathan, and Matthias Steup
[2018] 'The analysis of knowledge', *Stanford Encyclopedia of Philosophy* (Summer 2018 edition), E. Zalta, ed., <https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/>
- Jackson, Reginald
[1941] *An Examination of the Deductive Logic of John Stuart Mill* (Oxford: Oxford University Press).
- Jacovides, Michael
[2007] 'Locke's distinctions between primary and secondary qualities', in L. Newman, ed., *Cambridge Companion to Locke's Essay concerning Human Understanding* (Cambridge: Cambridge University Press), pp. 101–129.
[2017] *Locke's Image of the World* (Oxford: Oxford University Press).
- Kahneman, Daniel, Anne Treisman, and Brian Gibbs
[1992] 'The reviewing of object files: object-specific integration of information', *Cognitive Psychology* 24, pp. 175–219.

- Kant, Immanuel
 [1781–1787] *Critique of Pure Reason*, P. Guyer and A. Wood, trans. and eds. (Cambridge: Cambridge University Press, 1997).
 [1783] *Prolegomena to Any Future Metaphysics*, G. Hatfield, trans. and ed., revised edition (Cambridge: Cambridge University Press, 2004).
- Kennedy, Juliette, and Jouko Väänänen
 [2015] ‘Aesthetics and the dream of objectivity: notes from set theory’, *Inquiry* 58, pp. 83–98.
- Kitcher, Philip
 [1978] ‘Theories, theorists and theoretical change’, *Philosophical Review* 87, pp. 519–547.
 [2002] ‘On the explanatory role of correspondence truth’, *Philosophy and Phenomenological Research* 64, pp. 346–364.
- Kitcher, Philip, and Kyle Stanford
 [2000] ‘Refining the causal theory of reference for natural kind terms’, *Philosophical Studies* 97, pp. 99–129.
- Klement, Kevin
 [2010] ‘The functions of Russell’s no class theory’, *Review of Symbolic Logic* 4, pp. 633–664.
- Kline, Morris
 [1972] *Mathematical Thought from Ancient to Modern Times* (New York: Oxford University Press).
- Kornblith, Hilary
 [2002] *Knowledge and Its Place in Nature* (Oxford: Oxford University Press).
- Kripke, Saul
 [1982] *Wittgenstein on Rules and Private Language* (Cambridge, MA: Harvard University Press).
- Ladyman, James, and Don Ross (with David Spurrett and John Collier)
 [2007] *Every Thing Must Go* (Oxford: Oxford University Press).
- Lehrer, Keith
 [1976] ‘Reid’s influence on contemporary American and British philosophy’, in S. Barker and T. Beauchamp, eds., *Thomas Reid: Critical Interpretations* (Philadelphia, PA: Philosophical Monographs), pp. 1–7.
 [1989] *Thomas Reid* (London: Routledge).
- Leng, Mary
 [2010] *Mathematics and Reality* (Oxford: Oxford University Press).
- Linnebo, Øystein
 [2017] *Philosophy of Mathematics* (Princeton, NJ: Princeton University Press).
- Locke, John
 [1689] *An Essay concerning Human Understanding*, P. Nidditch, ed. (Oxford: Oxford University Press, 1975).
- Loeb, Louis
 [2007] ‘The naturalisms of Hume and Reid’, *Proceedings and Addresses of the American Philosophical Association* 81, pp. 65–92.
- Lüthy, Christoph
 [2000] ‘What to do with seventeenth-century natural philosophy? A taxonomic problem’, *Perspectives on Science* 8, pp. 164–195.
- Lycan, William
 [2007] ‘Moore’s anti-skeptical strategies’, in S. Nuccetelli and G. Seay, eds., *Themes from G. E. Moore* (Oxford: Oxford University Press), pp. 84–99.
- Maddy, Penelope
 [1988] ‘Believing the axioms’, *Journal of Symbolic Logic* 53, pp. 481–511, 736–764.

- [1990] *Realism in Mathematics* (Oxford: Oxford University Press).
- [1997] *Naturalism in Mathematics* (Oxford: Oxford University Press).
- [2001] 'Naturalism: friends and foes', *Philosophical Perspectives* 15, pp. 37–67.
- [2007] *Second Philosophy* (Oxford: Oxford University Press).
- [2008] 'How applied mathematics became pure', *Review of Symbolic Logic* 1, pp. 16–41.
- [2011a] 'Naturalism, transcendentalism, and therapy', in J. Smith and P. Sullivan, eds., *Transcendental Philosophy and Naturalism* (Oxford: Oxford University Press), pp. 120–156.
- [2011b] *Defending the Axioms* (Oxford: Oxford University Press).
- [2014a] 'A second philosophy of arithmetic', *Review of Symbolic Logic* 7, pp. 222–249.
- [2014b] *The Logical Must: Wittgenstein on Logic* (Oxford: Oxford University Press).
- [2017a] *What Do Philosophers Do? Skepticism and the Practice of Philosophy* (New York: Oxford University Press).
- [2017b] 'Set-theoretic foundations', in A. Caicedo, J. Cummings, P. Koellner, and P. Larson, eds., *Foundations of Mathematics: Essays in Honor of W. Hugh Woodin's 60th Birthday* (Providence, RI: American Mathematical Society), pp. 289–322.
- [2018] 'Replies to Coliva, Leite, and Stroud', *International Journal for the Study of Skepticism* 8, pp. 231–244.
- [2019] 'What do we want a foundation to do? Comparing set-theoretic, category-theoretic, and univalent approaches', in S. Centrone, D. Kant, and D. Sarikaya, eds., *Reflections on the Foundations of Mathematics: Univalent Foundations, Set Theory, and General Thoughts* (Cham, Switzerland: Springer, 2019), chapter 13.
- [2020] 'A plea for natural philosophy', *Proceedings and Addresses of the American Philosophical Association* 94, pp. 41–67. [This presidential address was extracted from Essay #1.]
- [202?a] 'Reply to Leng', to appear in Arbeiter and Kennedy [202?].
- [202?b] 'Reply to Wilson', to appear in Arbeiter and Kennedy [202?].
- Magnus, P. D., and Craig Callender
- [2004] 'Realist ennui and the base rate fallacy', *Philosophy of Science* 71, pp. 320–338.
- Malamet, David
- [1996] 'In defense of a dogma: why there cannot be a relativistic quantum mechanics of (localizable) particles', in R. Clifton, ed., *Perspectives on Quantum Reality* (Dordrecht: Kluwer), pp. 1–10.
- Marcus, Russell
- [2014] 'How not to enhance the indispensability argument', *Philosophia Mathematica* 22, pp. 345–360.
- Marr, David
- [1982] *Vision* (San Francisco: Freeman).
- Maudlin, Tim
- [2007] *Metaphysics within Physics* (Oxford: Oxford University Press).
- Metzinger, Thomas, ed.
- [2000] *Neural Correlates of Consciousness* (Cambridge, MA: MIT Press).
- Mill, John Stuart
- [1843] *A System of Logic* (London: Longmans, 1959).
- Miller, Richard
- [1987] *Fact and Method* (Princeton, NJ: Princeton University Press).
- Montano, Ulianov
- [2014] *Explaining Beauty in Mathematics* (Cham, Switzerland: Springer).
- Moore, G. E.
- [1910–1911] 'What is philosophy?', in his *Some Main Problems of Philosophy* (London: George Allen & Unwin, 1953), pp. 1–27.

- [1925] 'Defence of common sense', reprinted in his [1959], pp. 32–59.
- [1933–1934] 'Lecture notes, 1933 and 1934', in his *Lectures on Philosophy* (London: George Allen and Unwin, 1966), pp. 153–196.
- [1939] 'Proof of an external world', reprinted in his [1959], pp. 32–59.
- [1941] 'Certainty', first published in his [1959], pp. 227–251.
- [1959] *Philosophical Papers* (London: George Allen and Unwin).
- Moyal-Sharrock, Danièle
- [2004] *Understanding Wittgenstein's On Certainty* (New York: Palgrave Macmillan).
- Nadler, Steven
- [2002] *Companion to Early Modern Philosophy* (Malden, MA: Blackwell).
- Newton, Isaac
- [1687] *Principia: Mathematical Principles of Natural Philosophy*, I. B. Cohen and A. Whitman, trans. and eds. (Berkeley: University of California Press, 1999).
- [1704] *Opticks* (New York: Dover, 1952).
- [1962] *Unpublished Scientific Papers*, A. R. and M. B. Hall, eds. (Cambridge: Cambridge University Press).
- Nichols, Ryan
- [2007] *Thomas Reid's Theory of Perception* (Oxford: Oxford University Press).
- Nieder, Andreas
- [2011] 'The neural code for number', in Dehaene and Brannon [2011], pp. 103–118.
- Nolan, Lawrence
- [2011] 'Descartes on "what we call color"', in Nolan [2011a], pp. 81–107.
- Nolan, Lawrence, ed.
- [2011a] *Primary and Secondary Qualities* (Oxford: Oxford University Press).
- Olive, David
- [1998] 'The monopole', in Goddard [1998], pp. 88–107.
- Olson, Ingrid, and Yuhong Jiang
- [2002] 'Is visual short-term memory object based? Rejection of the "strong-object" hypothesis', *Perception and Psychophysics* 64, pp. 1055–1067.
- Orzack, Steven, and Elliott Sober, eds.
- [2001] *Adaptionism and Optimality* (Cambridge: Cambridge University Press).
- Ostwald, Wilhelm
- [1904] *The Principles of Inorganic Chemistry*, A. Findley, trans. (London: Macmillan, 1914).
- Pais, Abraham
- [1998] 'Paul Dirac: aspects of his life and work', in Goddard [1998], pp. 1–45.
- Palmer, Stephen
- [1999] *Vision Science* (Cambridge, MA: MIT Press).
- Pasnau, Robert
- [2017] *After Certainty* (Oxford: Oxford University Press).
- Perrin, Jean
- [1913] *Atoms*, D. Hammick, trans. (New York: Van Nostrand, 1923).
- Piazza, Manuela
- [2011] 'Neurocognitive start-up tools for symbolic number representations', in Dehaene and Brannon [2011], pp. 267–285.
- Pinker, Steven
- [1994] *The Language Instinct* (New York: William Morrow and Company).
- Pritchard, Duncan
- [2016] *Epistemic Angst* (Princeton, NJ: Princeton University Press).

- Pryor, James
 [2000] 'The skeptic and the dogmatist', *Nous* 34, pp. 517–549.
- Putnam, Hilary
 [1967a] 'The thesis that mathematics is logic', reprinted in his [1979], pp. 12–42.
 [1967b] 'Mathematics without foundations', reprinted in his [1979], pp. 43–59.
 [1968] 'Is logic empirical?' also known as 'The logic of quantum mechanics', reprinted in his [1979], pp. 174–197.
 [1973] 'Philosophy and our mental life', reprinted in his *Mind, Language and Reality: Philosophical Papers*, volume 2 (Cambridge: Cambridge University Press, 1975), pp. 291–303.
 [1975] 'What is mathematical truth?', reprinted in his [1979], pp. 60–78.
 [1976] 'Reference and understanding', in his *Meaning and the Moral Sciences* (Boston: Routledge & Kegan Paul, 1978), pp. 97–119.
 [1979] *Mathematics, Matter and Method, Philosophical Papers*, volume I, second edition (Cambridge: Cambridge University Press).
 [1995] 'Mathematical necessity reconsidered', in P. Leonardi and M. Santambrogio, eds., *On Quine* (Cambridge: Cambridge University Press), pp. 267–282.
- Quine, W. V. O.
 [1936] 'Truth by convention', reprinted in his [1976], pp. 77–106.
 [1951] 'Two dogmas of empiricism', reprinted in his *From a Logical Point of View*, second edition (Cambridge, MA: Harvard University Press, 1980), pp. 20–46.
 [1954] 'Carnap and logical truth', reprinted in his [1976], pp. 107–132.
 [1969] 'Epistemology naturalized', in his *Ontological Relativity and Other Essays* (New York: Columbia University Press), pp. 69–90.
 [1975] 'Five milestones of empiricism', in his *Theories and Things* (Cambridge, MA: Harvard University Press, 1981), pp. 67–72.
 [1976] *The Ways of Paradox*, second edition (Cambridge, MA: Harvard University Press).
- Reichenbach, Hans
 [1944] *Philosophical Foundations of Quantum Mechanics* (Berkeley: University of California Press).
- Reid, Thomas
 [1764] *Inquiry into the Human Mind on the Principles of Common Sense*, D. Brookes, ed. (University Park: Pennsylvania State University Press, 1997).
 [1785] *Essays on the Intellectual Powers of Man*, D. Brookes, ed. (University Park: Pennsylvania State University Press, 2002).
- Resnik, Michael
 [1980] *Frege and the Philosophy of Mathematics* (Ithaca, NY: Cornell University Press).
- Rickless, Samuel
 [1997] 'Locke on primary and secondary qualities', *Pacific Philosophical Quarterly* 78, pp. 297–319.
- Rogers, Brian Bruce
 [2010] 'Wittgenstein's philosophical methods in *On Certainty*', in V. Munz, K. Puhl, and J. Wang, eds., *Proceedings of the 32nd International Wittgenstein Symposium: Language and the World* (Berlin: De Gruyter), pp. 360–362.
 [2011] *Philosophical Method in Wittgenstein's On Certainty*, PhD dissertation, UC Irvine.
- Roth, Paul
 [2003] 'Review of Kornblith [2002]', *Notre Dame Philosophical Reviews*.
- Russell, Bertrand
 [1903] *The Principles of Mathematics* (New York: W. W. Norton and Co.).

- [1912] *The Problems of Philosophy* (Oxford: Oxford University Press, 1959).
- [1919] *Introduction to Mathematical Philosophy* (New York: Simon and Schuster, 1971).
- [1935–1936] ‘The limits of empiricism’, *Proceedings of the Aristotelian Society* 36, pp. 131–150.
- Saunders, Simon, Jonathan Barrett, Adrian Kent, and David Wallace, eds.
- [2010] *Many Worlds? Everett, Quantum Theory, and Reality* (Oxford: Oxford University Press).
- Sellars, Wilfred
- [1962] ‘Philosophy and the scientific image of man’, reprinted in his *Science, Perception and Reality* (London: Routledge & Kegan Paul, 1963), pp. 1–40.
- Shapin, Steven
- [1996] *The Scientific Revolution* (Chicago: University of Chicago Press).
- Shapiro, Stewart
- [2000] *Thinking about Mathematics* (Oxford: Oxford University Press).
- Sluga, Hans
- [1980] *Gottlob Frege*, (London: Routledge).
- Smith, John-Christian
- [1990] ‘Reid and the contemporary view of consciousness’, in J.-C. Smith, ed., *Historical Foundations of Cognitive Science* (Dordrecht: Kluwer), pp. 139–159.
- Smith, Sheldon
- [2014] ‘The mystery of applied mathematics?: a case study in mathematical development involving the fractional derivative’, *Philosophia Mathematica* 22, pp. 35–69.
- Sorensen, Roy
- [2012] ‘Vagueness’, *The Stanford Encyclopedia of Philosophy* (Summer 2012 Edition), Edward N. Zalta, ed., <http://plato.stanford.edu/archives/sum2012/entries/vagueness/>.
- Spelke, Elizabeth
- [1988] ‘Where perceiving ends and thinking begins: the apprehension of objects in infancy’, in A. Yonas, ed., *Perceptual Development in Infancy* (Hillsdale, NJ: Lawrence Erlbaum), pp. 197–234.
- [2000] ‘Core knowledge’, *American Psychologist* 55, pp. 1233–1243.
- Spelke, Elizabeth, Grant Gutheil, and Gretchen Van de Walle
- [1995] ‘The development of object perception’, in S. Kosslyn and D. Osherson, eds., *Visual Cognition* (Cambridge, MA: MIT Press), pp. 297–330.
- Stanford, Kyle
- [2000] ‘Refining the causal theory of reference for natural kind terms’ (with Philip Kitcher), *Philosophical Studies* 97, pp. 99–129.
- [2006] *Exceeding Our Grasp* (New York: Oxford).
- [2010] ‘Getting real: the hypothesis of organic fossil origins’, *Modern Schoolman* 87, pp. 219–243.
- [2014] ‘Reading nature: realist, instrumentalist, and quietist interpretations of scientific theories’, in L. Sklar, ed., *Physical Theory: Method and Interpretation* (New York: Oxford University Press), pp. 94–126.
- [2015a] ‘Catastrophism, uniformitarianism, and a scientific realism debate that makes a difference’, *Philosophy of Science* 82, pp. 867–878.
- [2015b] ‘“Atoms exist” is probably true, and other facts that should not comfort scientific realists’, *Journal of Philosophy* 112, pp. 397–416.
- [2016] ‘Naturalism without scientism’, in K. J. Clark, ed., *Blackwell Companion to Naturalism* (New York: Wiley-Blackwell), pp. 91–108.
- [2018] ‘Unconceived alternatives and the strategy of historical ostension’, in J. Saatsi, ed., *Routledge Handbook of Scientific Realism* (New York: Routledge), pp. 212–224.

- [2021] 'Realism, instrumentalism, particularism: a middle path forward in the scientific realism debate', in T. Lyons and P. Vickers, eds., *Contemporary Scientific Realism: The Challenge from the History of Science* (Oxford: Oxford University Press), pp. 216–238.
- Stein, Howard
- [1989] 'Yes, but ... some skeptical remarks on realism and anti-realism', *Dialectica* 43, pp. 47–65.
- [2004] 'The enterprise of understanding and the enterprise of knowledge', *Synthese* 140, pp. 135–176.
- Stenius, Erik
- [1960] *Wittgenstein's Tractatus* (Oxford: Basil Blackwell).
- Stern, David
- [2003] 'The methods of the *Tractatus*', in P. Parrini et al., eds., *Logical Empiricism* (Pittsburgh, PA: University of Pittsburgh Press), pp. 125–156.
- [2004] *Wittgenstein's Philosophical Investigations* (Cambridge: Cambridge University Press).
- Stroud, Barry
- [1977] *Hume* (London: Routledge).
- [1984] *The Significance of Philosophical Scepticism* (Oxford: Oxford University Press).
- [1996] 'Epistemological reflection on our knowledge of the external world', reprinted in his *Understanding Human Knowledge* (Oxford: Oxford University Press), pp. 122–138.
- [2009a] 'Review of *Second Philosophy*', *Mind* 118, pp. 500–503.
- [2009b] 'The transparency of "naturalism"', reprinted in his *Philosophers Past and Present* (Oxford: Oxford University Press, 2011), pp. 240–255.
- Tapper, Alan
- [2003] 'Reid and Priestley on method and the mind', in J. Haldane and S. Read, eds., *The Philosophy of Thomas Reid* (Malden, MA: Blackwell), pp. 98–119.
- Teller, Davida
- [2003] 'Color: a vision scientist's perspective', *Behavioral and Brain Sciences* 26, pp. 48–49.
- Ullman, Shimon
- [1980] 'Against direct perception', *Behavioral and Brain Sciences* 3, pp. 373–381.
- Van Atten, Mark
- [2004] *On Brouwer* (Stamford, CT: Thomson Wadsworth).
- Van Cleve, James
- [2002] 'Thomas Reid's geometry of visibles', *Philosophical Review* 111, pp. 373–416.
- [2004] 'Reid's theory of perception', in T. Cuneo and R. Woudenberg, eds., *Cambridge Companion to Thomas Reid* (Cambridge: Cambridge University Press), pp. 101–133.
- van Fraassen, Bas
- [2002] *The Empirical Stance* (New Haven, CT: Yale University Press).
- [2009] 'The perils of Perrin, in the hands of philosophers', *Philosophical Studies* 143, pp. 5–24.
- [2015] 'Naturalism in epistemology', in R. Williams and D. Robinson, eds., *Scientism, the New Orthodoxy* (London: Bloomsbury Academic Publishers), pp. 63–95.
- Vogel, Edward, Geoffrey Woodman, and Steven Luck
- [2001] 'Storage of features, conjunction, and objects in visual working memory', *Journal of Experimental Psychology: Human Perception and Performance* 27, pp. 92–114.
- Von Neumann, John
- [1947] 'The mathematician', reprinted in J. R. Neuman, ed., *The World of Mathematics* (New York: Simon and Schuster, 1956), volume 4, pp. 2053–2063.
- Wagner, Mark
- [2006] *The Geometries of Visual Space* (New York: Taylor and Francis).

- Warren, Richard
 [2003] 'Confusion of sensations and their physical correlates', *Behavioral and Brain Sciences* 26, p. 51.
- Warren, William
 [2012] 'Does this computational theory solve the right problem? Marr, Gibson, and the goal of vision', *Perception* 41, pp. 1053–1060.
- Weatherall, James Owen
 [2017] 'On G. E. Moore's "Proof of an external world"', *Pacific Philosophical Quarterly* 98, pp. 219–250.
- Weinberg, Steven
 [1987] *Elementary Particles and the Laws of Physics* (Cambridge: Cambridge University Press).
- Weiskrantz, Lawrence
 [1997] *Consciousness Lost and Found* (Oxford: Oxford University Press).
- Wells, H. G.
 [1908] *First and Last Things* (Teddington, UK: Echo, 2011).
- Whittle, Paul
 [2003] 'Why are we still asking this question?', in R. Mausfeld and D. Heyer, eds., *Colour Perception* (Oxford: Oxford University Press), pp. 203–204.
- Williamson, Timothy
 [1994] *Vagueness* (London: Routledge).
 [2000] *Knowledge and Its Limits* (Oxford: Oxford University Press).
 [2007] *The Philosophy of Philosophy* (Malden, MA: Blackwell).
 [2009] 'Replies to critics', in Greenough and Pritchard [2009], pp. 279–384.
 [2017] 'Review of *What Do Philosophers Do?* *Skepticism and the Practice of Philosophy*', *Journal of Philosophy* 114, 492–502.
- Wilson, Margaret
 [1979] 'Superadded properties: the limits of mechanism in Locke', reprinted in her [1999], pp. 196–208.
 [1992] 'The history of philosophy today; and the case of sensible qualities', reprinted in her [1999], pp. 455–494.
 [1999] *Ideas and Mechanism* (Princeton, NJ: Princeton University Press).
- Wilson, Mark
 [1982] 'Predicate meets property', *Philosophical Review* 91, pp. 549–589.
 [2006] *Wandering Significance: An Essay on Conceptual Behavior* (Oxford: Oxford University Press).
 [202?] 'The flight from application', to appear in Arbeiter and Kennedy [202?].
- Winkler, Kenneth
 [1989] *Berkeley* (Oxford: Oxford University Press).
- Wittgenstein, Ludwig
 [1921] *Tractatus Logico-Philosophicus*, C. K. Ogden, trans. (London: Routledge & Kegan Paul, 1922), D. Pears and B. McGuinness, trans. (London: Routledge & Kegan Paul, 1961). [Quotations in the text are from the Pears and McGuinness translation.]
 [1930] *Philosophical Remarks*, R. Rhees, ed., R. Hargreaves and R. White, trans. (New York: Barnes and Noble, 1975).
 [1937–1944] *Remarks on the Foundations of Mathematics*, revised edition, G. Anscombe, trans, G. von Wright, R. Rhees, G. Anscombe, eds. (Cambridge, MA: MIT Press, 1978).
 [1949] *Philosophical Investigations*, fourth edition, G. E. M. Anscombe, P. M. S. Hacker, and J. Shulte, trans. (Oxford: Blackwell, 2009) (First published in 1953).

REFERENCES

- [1949–1951] *On Certainty*, G. E. M. Anscombe and G. H. von Wright, trans. and eds. (Cambridge, MA: MIT Press, 1969).
- Wolterstorff, Nicholas
- [1987] ‘Hume and Reid’, *Monist* 70, pp. 398–417.
- Wright, Crispin
- [2002] ‘(Anti-)sceptics simple and subtle: G. E. Moore and John McDowell’, *Philosophy and Phenomenological Research* 65, pp. 330–348.
- Zagzebski, Linda
- [1994] ‘The inescapability of Gettier problems’, *Philosophical Quarterly* 44, pp. 65–67.

INDEX

For the benefit of digital users, indexed terms that span two pages (e.g., 52–53) may, on occasion, appear on only one of those pages.

- abductive inference, 50, 57–58, 61, 62–90
Achinstein, Peter, 60–61
Adams, Zed, 37
Allred, Sarah, 45n.92
Anderson, Barton, 241n.21
Anderson, Lanier, 197n.20
Anscombe, Elizabeth, 148–49, 155, 157–58
Anstey, Peter, 17n.20
applied mathematics. *See* mathematics in application
Arealism. *See* set theory
Argument from Illusion, 27, 52n.10, 102–3, 104–5, 115–16, 122–26, 131–32, 133
Aristotle, 14–15, 17n.18, 21, 185, 208
arithmetic, 8–9, 49, 61–62, 93–94, 151, 189, 206–7, 216–17, 237, 244–52, 253–61, 267–68, 271–72, 285, 286–87, 288–93
Atherton, Margaret, 19n.31, 20n.35
atoms, 5, 46n.98, 50, 50n.5, 54–55, 57, 58, 59–60, 62–90, 218–19, 222–23n.9, 283n.45
Austin, J. L., 1n.3, 6–7, 13–14, 24–25, 29, 41, 42–44, 45–47, 111–14, 118, 123n.54, 157–58, 173n.2, 174n.4, 180n.13, 181–87
Ayer, A. J., 198–203, 205, 206, 215

Bacon, Francis, 15–16, 95, 100–1
Baldwin, Thomas, 135n.4
Bangu, Sorin, 187n.29, 252n.52
Benacerraf, Paul, 20n.36
Berkeley, George, 4, 5–6, 18, 19n.31, 19–21, 22nn.42–43, 32n.66, 43–44, 45–46, 97, 103n.18, 106–7, 115, 122n.52, 122n.53, 128

Bloom, Paul, 248–49
body, 17, 29–30, 32–34, 35–36, 41. *See also* object concept
Bolzano, Bernard, 208
Boole, George, 197n.21
Boyle, Robert, 4, 16–17, 19n.31, 19–20, 22nn.42–43, 30–31, 33–34, 37–38, 45–46
Broughton, Janet, 97–99, 100, 105n.20, 107n.23, 107–8n.24
Burge, Tyler, 191n.6, 192–93n.10, 193n.11, 239nn.10–12, 240–42, 245n.36, 245–46
Burgess, J. A., 231

Callender, Craig, 67n.35
Cantor, Georg, 190–91, 250–51n.51
Carey, Susan, 238nn.7–9, 244nn.30–31, 245–46nn.38–40
Carnap, Rudolf, 48n.103, 59, 169n.32, 203–6, 210n.29, 215
Carroll, Lewis, 205–6, 210n.29, 280n.41
Cassam, Quassim, 101, 46nn.96–98,
Cavell, Stanley, 152
Chakravartty, Anjan, 90n.69
Chesney, Dana, 245n.37
Chin, Adam, 90n.69
Chomsky, Noam, 25, 249–50
Cohen, Jonathan, 36, 37n.76
Cohen, Sheldon, 15n.8
Coliva, Annalisa, 135n.4, 148n.1, 158n.14, 159n.19, 162n.23, 163n.26, 169n.32, 170n.34
color, 23, 29–41, 43–45, 47–48, 150–51, 155

- common sense, 2–3n.6, 13, 18, 54–55, 93, 108–34,
137, 141–42, 143, 146, 170n.33, 218–19,
237–38, 243, 265n.5, 283–84
- Conant, James, 148–49
- conceptual analysis, 21n.38, 22–23, 28–29, 30,
31, 32, 35–36, 41, 46–48, 68–69, 77, 116–
17n.43, 136, 137–38, 142, 143, 182, 194,
204, 220, 239, 247
- constructivism, 7, 190–91, 194–95, 196–97,
197n.18, 206, 210, 215, 232–33, 285–86
- corpuscular mechanism. *See* mechanism
- correspondence. *See* truth and reference
- Crone, Robert, 122n.52
- Curley, Edwin, 31
- Dalton, John, 67, 73, 78–79
- Darwin, Charles, 27, 67, 209
- Dear, Peter, 14–15nn.6–7, 15n.10
- De Bary, Philip, 123n.54, 128–29n.60
- Dehaene, Stanislas, 245–46nn.38–40,
246nn.41–42, 251
- demarcation criterion, 1n.2, 3n.7, 95–96
- Descartes, René, 4, 5–6, 16–17, 20–21, 22nn.42–
43, 27, 30, 31, 32, 32n.66, 37–38, 43–44,
45–46, 53, 97, 97n.9, 109n.26, 115, 126–34,
140–41, 145–46, 214n.38, 262
- developmental psychology. *See* psychology
- deviant logic. *See* logic
- Diamond, Cora, 148–49, 153–54
- dinosaurs, 63, 70, 80–81, 84, 143, 146, 147,
170n.33
- Dirac, Paul, 275
- disquotation. *See* truth and reference
- Downing, Lisa, 17, 18n.28, 20n.35, 30nn.61–62,
30, 31, 31n.64, 32n.65, 32n.66, 33–34,
35n.72, 37, 48n.104
- dream argument, 52n.10, 128–29, 132–33, 135,
140–41, 158–59, 162n.23, 166, 167
- Dumas, Jean-Baptiste, 55n.18
- Eddington, A. S., 7–8, 219–20, 222–23n.9,
237–38n.4, 238n.8
- Einstein, Albert, 46n.98, 59, 64–65, 67, 68–69, 70,
80–81, 89, 209
- eliminative induction. *See* abductive inference
- empiricism, 2, 88–89, 101n.15, 122n.52, 122n.53,
206–7, 210
- constructive empiricism, 5, 50, 51–62,
65, 89n.66
- epistemology (disciplinary), 20–23, 25, 28–29,
46–47, 182, 228, 243
- epistemology naturalized, 36n.73, 101n.16,
224n.12
- hinge epistemology, 6, 148–70
- Ernst, Michael, 17–18n.26
- Euler, Leonhard, 19–20, 262, 266–67
- Evil Demon. *See* dream argument
- Feigenson, Lisa, 244n.30
- Feynman, Richard, 42n.83
- Field, Hartry, 6n.11, 174–75, 176n.8, 178–79n.11
- Fine, Arthur, 58
- Fogelin, Robert, 106n.22, 107–8n.24, 152
- Formalism, 7, 190–91, 200, 203, 206, 215, 264
- foundations of the sciences, 4, 17n.18, 42–46, 54,
189, 190–91, 193, 194–95, 266–67, 272,–
73nn.23–25, 274, 276–77, 279–80
- in Hume, 18n.30, 99, 107, 127, 128–29
- in Reid, 128–29n.60, 132n.62
- in Wittgenstein, 202
- Frege, Gottlob, 7–8, 190–93, 197, 198, 203,
206, 207, 212–13, 215, 217, 247n.43, 263,
283n.46
- Galebach, Ethan, 252n.52
- Galileo Galilei, 13, 16–17, 19–20, 22nn.42–43,
29, 31, 32
- Gaukroger, 16n.12, 20n.35
- Gelman, Rochel, 247–48, 249
- geometry, 110, 121–22, 123, 149, 189, 194, 200–1,
210, 216–17, 262, 267, 277–78
- Geroch, Robert, 293n.62
- Gibbins, Peter, 210n.32, 211
- Gibson, James J., 240
- Glymour, Clark, 60–61
- Gödel's incompleteness theorems, 288–93
- Godfrey-Smith, Peter, 67n.35
- Goldman, Alvin, 46–47n.99
- Greenberg, Sean, 134n.66
- Gregory, Frederick, 197n.19
- Grice, Paul, 163n.26
- group theory, 39, 264–65, 272–73n.25, 273, 276,
276n.32, 279–80
- Gubelmann, Reto, 252n.52
- Hahn, Hans, 198
- Harnett, Patrice. *See* Rochel Gelman
- Hatfield, Gary, 19–20n.34, 44n.89, 44–45nn.91–
92, 110–11n.29, 226n.15, 239nn.10–12,,
240–43, 245n.36, 252n.52
- Hauser, Marc, 249–50
- Heck, Richard, 247, 250–51n.51
- Heis, Jeremy, 192–93n.10
- Hellman, Geoffrey, 265n.6, 293nn.61–62
- Helmholtz, Hermann von, 240
- Henry, John, 14n.6

- Hilbert, David, 8–9, 41, 263, 264, 289
 hinge epistemology. *See* epistemology
 Hollingworth, Andrew, and Ian Rasmussen, 245n.35
 Hume, David, 5–6, 18n.30, 22n.42, 25–26, 27, 28–29, 48
 Hurlebert, Russell, 120n.48
 Husserl, Edmund, 197n.18
 Huygens, Christiaan, 17
 Hylton, Peter, 208n.28
- Ichikawa, Jonathan, 22n.41
 idealism, 7–8, 28–29, 135, 136n.9, 159n.20, 194–200, 202–3, 206, 212
 indispensability arguments, 23–27, 49–50, 89, 264
 inference to the best explanation. *See* abductive inference
 introspection, 2, 15n.8, 19–20, 27, 77–121, 116–17n.43, 127, 128–29, 130, 131n.61, 131–32, 133
 intuition, 213–14, 237–38n.4, 265nn.5–6, 267, 271–72
 Cartesian, 129
 as heuristic, 191n.6, 250–51, 258, 259–61, 288–89
 Kantian, 194–96, 200–1
 rational, 239n.12
 in Wittgenstein, 152
 Intuitionism. *See* constructivism
- Jackson, Reginald, 208n.27
 Jacovides, Michael, 25, 17nn.19–20, 20n.35, 31n.64, 37, 43n.86
- Kahnemann, Daniel, 245n.32
 Kant, Immanuel, 4, 5n.9, 7–8, 25–26, 28–29, 41, 46–47, 48, 55–56, 59, 60–61, 107–8n.24, 115n.41, 136–38, 141, 149, 194–97, 200–1, 202–3, 206, 212–14, 215, 217n.4, 283n.46
 Kassam, Alysha, 90n.69
 Kennedy, Juliette, 275, 293n.62
 Kitcher, Philip, 175–79, 180–81, 183–84
 Klement, Kevin, 208n.28
 Kline, Morris, 266–67, 276
 Koellner, Peter, 45n.94
 Kornblith, Hilary, 46–47nn.98–99
- Lacroix, Sylvestre, 266–67
 Ladyman, James, 7–8, 222–23n.9, 229n.19, 237–38n.4
 large cardinals, 27, 276–77, 285
 Lavoisier, Antoine, 68–69, 73, 175–76
 Leeds, Stephen, 6n.11, 57–58, 60–61
- Lehrer, Keith, 112
 Leibniz, Gottfried, 262, 266–67
 Leite, Adam, 170n.34
 Leitz, Charles, 90n.69, 187n.29, 293n.62
 Leng, Mary, 271n.21
 Linnebo, Øystein, 263n.3, 276n.32
 Locke, John, 4, 5–6, 17, 19n.31, 19–21, 22nn.42–43, 28, 30, 31–36, 37–38, 41–42, 45–46, 97, 115, 128
 Loeb, Louis, 128–29n.60
 logic
 deviant, 215, 229n.18, 231–33, 284
 rudimentary, 8, 215, 217–19, 221–25, 228, 232n.25, 232–33, 237–38, 239–40, 242–44, 284–87
 Logicism, 190–91, 264
 Lüthy, Christoph, 16n.13, 19n.31
 Lycan, William, 146n.18
- Magnus, P. D., 67n.35
 Malament, David, 79n.52, 90n.69, 187n.29, 215n.39, 293n.62
 Malcolm, Norman, 155–56, 157–58, 163n.25, 168n.30
 Marcus, Russell, 268n.15
 Marcus, Ruth, 215n.39
 Marino, Patricia, 215n.39, 233n.31
 Marr, David, 44–45n.91
 Massas, Guillaume, 293n.62
 mathematics in application, 2, 262, 266–72, 285, 287, 289–90
 Maudlin, Timothy, 43n.85
 Maxwell, James Clerk, 68–69, 73, 79
 mechanism
 Corpuscular Mechanism, 15–17, 19–20, 29–36, 41–42
 19th century mechanism, 79
 Metzinger, Thomas, 45n.93
 Mill, John Stuart, 206–9, 210, 211–12, 215
 Mitsch, Chris, 90n.69, 187n.29, 293n.62
 model/modelling, 7–8, 39–292, 59n.23, 65n.31, 183n.22, 233n.27, 265n.7, 268n.13
 Molland, George, 19n.31
 Montano, Uliyanov, 275
 Moon, Stella, 90n.69, 293n.62
 Moore, G. E., 6, 13n.2, 28–29, 76n.47, 112, 135–47, 155, 157–58, 160, 162–64, 165, 170
 Moyal–Sharrock, Danièle, 154–55, 162n.24
- Newton, Isaac, 4, 15–18, 19–20, 22n.43, 24, 25, 27, 30, 33n.67, 40, 40n.80, 42n.82, 43–44, 45–46, 71–72, 73, 95, 95n.7, 97n.9, 98n.13, 100–1, 109, 209, 262, 266–67, 269–70

INDEX

- Nichols, Ryan, 110–11n.29, 119n.47, 123n.54
 Nolan, Lawrence, 37n.77
- object concept, 7–8, 26, 107–8n.24, 219–29, 237–43. *See also* body
- Olive, David, 275n.28
 Olson, Ingrid, 243n.25
 Ordinary Language Philosophy, 13–14, 45–47, 111–14, 118, 157–58, 181–87
 Orzack, Steven, 45n.95
 Ostwald, Wilhelm, 59, 64, 65, 72, 82
- Pais, Abraham, 275n.28
 Palmer, Stephen, 44nn.88–90,
 Pasnau, Robert, 15nn.8–9, 21–23, 23n.45, 182–83
 Pears, David, 148–49
 perceptual relativity. *See* Argument from Illusion
 Perrin*, 65–90
 Perrin, Jean, 5, 46n.98, 50, 57, 59, 60–61, 63–65, 66–67, 72, 80–81, 86, 89
 Piaget, Jean, 224–25, 238n.7, 239n.10, 246n.41
 Piazza, Manuela, 246n.42
 Pinker, Steven, 258n.6
 Plain Inquirer, 1n.3, 19n.32, 96n.8, 141n.14, 145n.17, 216n.1
 Plato, 21, 263
 Platonism, 7, 190–93, 198, 203, 206–7, 208, 211–12, 213, 214, 215, 221, 263
 Poincaré, Henri, 59
 Popper, Karl, 71
 Prichard, Duncan, 148n.1, 162n.22, 170n.34
 Priestley, Joseph, 77–78, 110–11n.29, 175–249, 183n.24
 primary/secondary, 17, 29–45, 103n.18
 priority of sense, 149, 150, 154, 168–69, 170
 Pryor, James, 135nn.1–5,
 Psillos, Stathis, 68–35
 psychology, 8, 18, 19–21, 24, 25, 28, 36n.73, 44–45, 46n.97, 47–48, 98–99, 110, 110–11n.29, 113n.38, 136–37, 155, 168–69, 182n.19, 195, 197, 199–200, 205n.25, 213, 224, 237–52, 255–58, 274, 277, 289–90
 Psychologism, 198, 199–200, 206, 213, 215
 transcendental psychology, 28, 197, 202–3
 Putnam, Hilary, 34–35n.70, 173n.1, 210–12, 215, 263n.3, 264, 265nn.5–7, 267–69, 270–71, 271n.21, 271n.20, 276n.32
- quantum mechanics, 25, 42, 50, 58, 71–72, 79, 82, 88, 209, 210–12, 213–14, 215, 221–23, 229, 232–33, 237–38, 243–44, 270–71, 285
 Queck, Toni, 293n.62
- Quine, W. V. O., 6n.11, 22n.42, 36n.73, 46n.97, 49–50, 53, 58n.22, 89, 101n.16, 204n.24, 205–6, 209–10, 211–12, 215, 224n.12, 268, 274n.27, 277–78, 280–82
- Reck, Erich, 215n.39
 reference. *See* truth and reference
 Reichenbach, Hans, 210
 Reid, Thomas, 4, 5–6, 13n.2, 18, 19–21, 22nn.42–43, 23n.44, 25–27, 28–29, 38–40, 43–44, 45–46, 47, 48, 93, 104–5, 107–8n.24, 108–34, 137
 representative theory of perception, 104, 105–7, 117–18, 133
 Resnik, Michael, 263n.3, 266, 273–74, 282–83, 285, 288–89, 290, 293n.62
 Rickless, Samuel, 31
 Robust Realism. *See* set theory
 Rogers, Brian, 6, 154–58, 161, 162–63, 163n.25, 168n.30
 Rohloff, Waldemar, 215n.39
 Rosen, Gideon, 8–9
 Ross, Don, 7–8, 222–23n.9, 229n.19, 237–38n.4
 Roth, Paul, 46–47n.99
 rudimentary logic. *See* logic
 Rush, Penelope, 7–8, 233n.31
 Russell, Bertrand, 24, 25, 28–29, 41, 42–44, 45–46, 142n.15, 145, 148–49, 203, 208, 211–12, 248n.47, 264, 267–68
- Salmon, Wesley, 60–61
 Saunders, Simon, 43n.84
 Schatz, Jeffrey, 48n.104, 90n.69, 187n.29, 293n.62
 Scholasticism, 14–16, 17, 31
 Schönbaumsfeld, Gina, 170n.34
 Schrödinger, Erwin, 275n.28
 Schwitzgebel, Eric, 120n.48
 Searle, John, 163n.26
 Sellars, Wilfred, 7–8, 220–21n.7, 222–23n.9, 237–38n.4
 set theory, 8n.12, 9, 49, 54, 190–91, 237n.2, 250–51n.51, 264–65, 271–72, 273, 274, 276–77, 277n.35, 279–80
 Arealism, 259–61, 274–93
 Thin Realism, 54n.16, 293n.61
 Shapin, Steven, 14–15nn.6–7, 15nn.10–11, 16n.12
 Shapiro, Stewart, 263n.3, 263n.4, 264
 Sluga, Hans, 197n.19
 Smith, John-Christian, 111n.30
 Smith, Sheldon, 276n.29
 Sober, Elliott, 45n.95
 Sommers, Evan, 90n.69, 187n.29
 Sorensen, Roy, 231n.23

INDEX

- Spelke, Elizabeth, 226, 227, 239n.11
 Stafford, Will, 292n.58
 Stanford, Kyle, 5, 50, 57, 62–90, 174–79, 180–81, 215n.39
 Stein, Howard, 39–40, 42n.82, 44n.87, 48, 76–78, 80
 Stenius, Eric, 148–49, 203n.23
 Stern, David, 148n.4, 151n.6, 152, 200n.22
 Steup, Matthias, 22n.41
 Stroud, Barry, 3–4, 5n.9, 18n.30, 26n.52, 52n.10, 93n.3, 96n.8, 97, 97n.11, 102n.17, 125–27, 129, 135, 167n.29
 Tapper, Alan, 110–11n.29
 Teller, Davida, 47–48
 theory of ideas, 27, 43–44, 96–98, 102–3, 104–8, 111n.30, 115–27, 128–29, 131–32, 133
 therapeutic philosophy, 13, 93, 148–70, 251–52
 Thin Realism. *See* set theory
 transcendentalism, 5, 7–8, 28–29, 55–56, 57, 59–62, 88–89, 93, 194–97, 202–3, 206, 212, 213
 Ullman, Shimon, 240n.19
 Väänänen, Jouko, 30
 Van Atten, Mark, 197n.18
 Van Cleve, James, 122n.52, 123–24n.55
 Van Fraassen, Bas, 5, 48n.103, 50, 51–62, 65n.31, 65, 66–67, 69, 72, 77–78, 88–89
 vision, 4, 18, 21–22n.40, 23, 40, 42n.82, 43–48, 54, 58, 103, 105n.21, 110, 119n.46, 120n.48, 123–24, 125–26, 132n.62, 147, 150, 162n.23, 167, 182n.19, 224–25, 228n.16, 237, 240n.18, 241–43, 244n.30, 245, 246, 254
 Vogel, Edward, 243n.25
 Von Neumann, John, 274–75
 Von Wright, 155
 Wagner, Mark, 44n.89
 Warren, Richard, 48n.102
 Warren, William, 241n.21
 Weatherall, James Owen, 134n.66, 162n.23
 Weiskrantz, Lawrence, 121n.50
 Wells, H. G., 231n.23
 Wheatstone, Charles, 122n.52
 Whittle, Paul, 48n.102
 Williamson, Timothy, 21n.38, 22n.41, 46nn.96–98, 231n.24, 284–85n.48
 Wilson, Margaret, 26, 17n.21, 18–19, 37, 42n.82
 Wilson, Mark, 77, 80, 82, 83n.58, 173n.2, 174–75, 177–78, 180, 181–83, 184n.25
 Winkler, Kenneth, 19n.31, 20n.35
 Wittgenstein, Ludwig, 6, 13, 135, 148–70, 180n.13, 200–4, 205, 206–7, 244n.29, 250–51, 283n.43
 Wolterstorff, Nicholas, 128–29n.60
 Woodin, Hugh, 45n.94
 Wright, Crispin, 6n.11, 135n.1
 Zagzebski, Linda, 22n.41
 Zermelo, Ernst, 272

