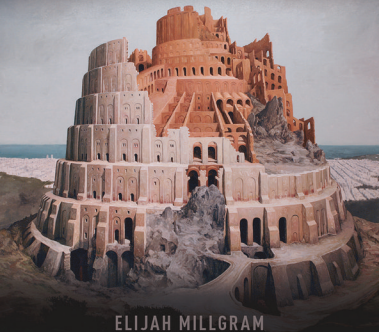


# THE GREAT ENDARKENMENT

PHILOSOPHY FOR AN AGE OF  
HYPERSPECIALIZATION



ELIJAH MILLGRAM

# **The Great Endarkenment**



# The Great Endarkenment

PHILOSOPHY FOR AN AGE  
OF HYPERSPECIALIZATION

Elijah Millgram

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

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

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

© Oxford University Press 2015

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

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

Library of Congress Cataloging-in-Publication Data  
Millgram, Elijah.

The great endarkenment : philosophy for an age of hyperspecialization / Elijah Millgram.  
p. cm.

Includes bibliographical references.

ISBN 978-0-19-932602-0 (cloth : alk. paper) 1. Philosophy, Modern—21st  
century—Methodology. 2. Specialism (Philosophy) 3. Analysis (Philosophy) I. Title.  
B805.M55 2015  
191—dc23

2014032204

9 8 7 6 5 4 3 2 1

Printed in the United States of America on acid-free paper

For my parents



# CONTENTS

Acknowledgments ix

<b>1. Introductory Remarks on the Tower of Babel</b>	<b>1</b>
<b>2. The Great Endarkenment</b>	<b>21</b>
Appendix A	44
Appendix B	49
<b>3. Practical Reasoning for Serial Hyperspecializers</b>	<b>54</b>
<b>4. D'où venons-nous . . . Que sommes nous . . . Où allons-nous?</b>	<b>77</b>
<b>5. Millian Metaethics</b>	<b>101</b>
<b>6. Why Do We Think There Are Things We Ought to Do?</b>	<b>126</b>
<b>7. Lewis's Epicycles, Possible Worlds, and the Mysteries of Modality</b>	<b>155</b>
<b>8. Progressive Necessity</b>	<b>188</b>
<b>9. Applied Ethics, Moral Skepticism, and Reasons with Expiration Dates</b>	<b>221</b>
<b>10. Segmented Agency</b>	<b>234</b>
Postscript	263
<b>11. Afterword: A Call to Arms</b>	<b>269</b>
 Bibliography	 283
Index	297





## ACKNOWLEDGMENTS

A number of the chapters in this volume have made previous appearances:

- “The Great Endarkenment” includes material from a review of Steven Hales, editor, *A Companion to Relativism*, which appeared in *Notre Dame Philosophical Reviews*, April 2012.
- “Practical Reasoning for Serial Hyperspecializers” appeared in *Philosophical Explorations* 12(3): 261–278. September 2009. Reprinted by permission of the publisher (Taylor & Francis Ltd, <http://www.tandf.co.uk/journals>).
- “D’où venons-nous . . . Que sommes nous . . . Où allons-nous?” originally appeared in D. Callcut, *Reading Bernard Williams* (London: Routledge, 2009): 141–165.
- “Millian Metaethics” is a revised and expanded version of “Mètaètica milliana,” *Quaderni di Scienza Politica* XVII, 2 (2010): 245–265.
- “Applied Ethics, Moral Skepticism, and Reasons with Expiration Dates” appeared in S. Black and E. Tiffany, *Reasons to be Moral Revisited* (Calgary: University of Calgary Press, 2009): 263–280. Reprinted by permission of the publisher (<http://www.tandfonline.com>).
- “Segmented Agency” appeared in M. Vargas and G. Yaffe, *Rational and Social Agency: The Philosophy of Michael Bratman* (Oxford: Oxford University Press, 2014): 152–189.

I am very grateful for permission to reprint. Acknowledgments are to be found chapter by chapter, but it is especially appropriate to thank the John Simon Guggenheim Memorial Foundation here, for providing fellowship support during the time the book as a whole was being assembled. Thanks also to Aliya Khan for proof checking.



## Introductory Remarks on the Tower of Babel

And the whole earth was of one language, and of one speech. And it came to pass, as they journeyed from the east, that they found a plain in the land of Shinar; and they dwelt there. And they said one to another, Go to, let us make brick, and burn them thoroughly. And they had brick for stone, and slime had they for mortar. And they said, Go to, let us build us a city and a tower, whose top may reach unto heaven; and let us make us a name, lest we be scattered abroad upon the face of the whole earth. And the Lord came down to see the city and the tower, which the children of men builded. And the Lord said, Behold, the people is one, and they have all one language; and this they begin to do; and now nothing will be restrained from them, which they have imagined to do. Go to, let us go down, and there confound their language, that they may not understand one another's speech. So the Lord scattered them abroad from thence upon the face of all the earth: and they left off to build the city. Therefore is the name of it called Babel; because the Lord did there confound the language of all the earth: and from thence did the Lord scatter them abroad upon the face of all the earth.

—*GENESIS 11:1–9*

The Kipling-like just-so story, evidently meant to explain something that has to be a central part of any acceptable philosophical anthropology, makes a miracle out of an inevitability, and let's take a moment to consider why. Putting up a skyscraper requires a great deal in the way of technical specialization; you need the architects and the code consultants, the mechanical engineers and the plumbing engineers, the civil engineers and the landscape architects, the waterproofing consultants and the hardware consultants, the lighting designers and the daylighting consultant, the interior designers and the acoustical consultants; and we haven't so much as gotten to the specialty consultants. On the client side, there have to be project managers, owner's

representatives, the financing team, and so on. Once the project goes into construction, the general contractor will engage steel fabricators, roofers, electricians, plumbers, framers, drywallers, painters . . . and we have barely gotten started on the full list.<sup>1</sup>

Now, in the course of acquiring their professional specializations, the specialists will have learned a technical vocabulary, and often enough, not just a vocabulary, but specialized notation and techniques for manipulating it, as when the architects learn to read blueprints or the engineering students take calculus. They learn discipline-specific standards, guidelines, and priorities, along with the hard-to-formulate sensibilities that allow them to work with the approximations, idealizations, and other avowedly not-quite-true assertions which make up such a large part of the information in any field. And when they have come up to speed, the normal upshot is—and this much should be familiar from almost anyone’s experience nowadays—that only a similarly trained member of that specialization can properly understand them.

The obstacles to mutual comprehension go deeper than the fact that outsiders cannot read specialist professional literature. Because professional standards end up being framed in a proprietary vocabulary, one that can take anywhere from five years to a decade to learn, outsiders cannot so much as understand those standards. Because specialists internalize those standards, outsiders cannot understand the specialists’ concerns and preferences (in part because they cannot comprehend what the preferences are for, in part because outsiders cannot see *why* anyone would have *those* preferences). And because they cannot share the technical sensibilities of the insiders, even when people *seem* to understand, say, pronouncements on factual matters in someone else’s area of expertise, they are not normally competent to *use* the information they superficially understand; the information that specialists work with is accompanied by special handling instructions. What we saw quaintly called confounding of language—the inability of one person to understand another—is entailed by the ability of a society to build a decently tall tower.

We are, and are becoming ever more, a society of specialists. As recently as the early nineteenth century, it was possible for a polymath—such as, famously at the time, William Whewell—to master all of the science of his day.<sup>2</sup> It is not nearly possible any longer; specialization is far more highly articulated than at any time in previous human history, and because this difference in degree has come to amount to a difference in kind, I’ll mark the newly extreme form of division of labor with the label *hyperspecialization*. Consequently, communication across the barriers between professions and disciplines is our own very

---

<sup>1</sup> I’m grateful to Michelle Hill for background.

<sup>2</sup> For a popular account, see Snyder (2011); Fisch (1991) is a recommended treatment of Whewell’s philosophy.

pressing problem, and it threatens not just our more ambitious enterprises but the successful management of our day-to-day lives.

The builders of the Tower of Babel are represented as committing the sin of hubris (though that is a Greek rather than a Hebrew way of putting it); so the lesson of the tale might seem to be that we should give up on high-rise construction and, in general, on overly ambitious technical projects. But we cannot walk back our commitment to hyperspecialization; whatever the drawbacks of hubris, there are at this point simply too many people on our planet to do without this sort of division of labor.<sup>3</sup> Hyperspecialization is necessary if we are to keep anything on the order of our current population alive, which is to say that the resolution of the biblical tale—geographical dispersion into distinct societies composed of people who presumably *do* speak the same language—is no longer an option for us. When the languages in question are those of communities of experts, that would be to segregate the specialists from one another, and specialists can only get the job done when they work together. We have no choice but to live together and cooperate with other individuals whom we cannot possibly understand, in executing the large-scale and very demanding projects on which our survival now depends.

## 1.1

This book is devoted to the specifically philosophical problems that we face because we live in our own version of the Tower of Babel. Now, the pro forma disclosure: I write as an analytic philosopher. Analytic philosophers, especially, are not used to the idea that division of labor poses problems, much less urgent ones, for them. And so, as I introduce this collection of essays, I will need to explain why these very basic facts about contemporary society should concern *philosophers*.

But now, if my primary audience does consist of philosophers, that has consequences for the way a volume like this has to be organized. The philosophers themselves, these days, are not only specialized, but *subspecialized*; for instance, it is now routine for a philosopher not merely to think of himself as primarily, say, an ethicist or an historian, but to confine himself to an area of moral philosophy such as virtue ethics or metaethics, or, within history, an area such as ancient (i.e., Greek and Roman) philosophy. I will take up the challenges this poses to my project in the Afterword; right now, notice that such subspecialization had better figure as a constraint on my presentation. The professional incentives that come with subspecialization make it nearly impossible to convince many analytic philosophers to read anything that is clearly outside their subspecialization; a monograph that cuts across a

---

<sup>3</sup>This is not a new observation; see, for instance, Hayek (1994, p. 110).

number of these subfields, even if it does engage one's own subspecialization, is likely never to be picked up. So to make the ideas available to subspecialized philosophers, the chapters in this book, with the exception of this introductory overview and the Afterword (Chapter 11), can be read as freestanding essays. This strategy entails a certain amount of repetition, and I apologize for it in advance: the organizing ideas of the volume will be reintroduced wherever they are needed. I do hope that once a reader has engaged the problems I am posing within the confines of his own subspecialization, he will be persuaded that only by taking a broader view of the issues can even those problems be addressed—and so will decide to read other chapters, and even the book as a whole.

The remainder of this Introduction will be given over to four tasks. I will provide a roadmap of the book, one which allows professional readers with specialized interests to locate the chapters most directly pertinent to them. I will also preview a number of themes that reappear from chapter to chapter. I will consider how much progress it is feasible to make at the present stage of the discussion, and so what both the ambitions and limits of the argument of the book must be. Last but not least—I take this very seriously indeed—when philosophical problems are important, they are not problems *only* for philosophers. Thus I would like this book to be read by a broader audience than just those PhD-bearing professionals. And so I am going to explain how it is that the specifically philosophical problems to which I want to draw our attention are also of broad, immediate and very practical interest.

Accordingly, although some of the discussion may seem exotic, I have attempted to write throughout so that a reader without an advanced degree in philosophy can follow it; this has meant explaining from scratch concepts and moves with which readers who are in-field will be familiar, and I'm going to ask them for patience while I keep other readers on board. Let me flag one issue that the alert reader will already be wondering about: I have just claimed that professional outsiders cannot properly understand the insiders, and I have also just promised accessibility to the nonphilosophers. In the Afterword to the book (Chapter 11), I will return to the question of how I can expect the nonspecialists to track the arguments I will present here.

## 1.2

Turning now to one of those threads that is going to keep on reappearing throughout the upcoming chapters: as I worked on specialization and the problems it poses, I discovered that in order to make headway I would have to push back against apriorism in philosophy, pretty much across the board. An informal introduction to the relevant bit of philosophers' terminology: what you can know a priori is what you can know up front, without looking; conversely, if you have to look and see to find something out, then your knowledge

is a posteriori.<sup>4</sup> With rare exceptions, philosophy has conducted itself as an apriorist enterprise, meaning that it is assumed that you can know most everything about the topics taken up by philosophers without going outdoors to look at anything at all.<sup>5</sup> If the reader is philosophically trained, the attempt to up-end this part of the discipline's self-conception may strike him as an agenda best left for its own occasion. So as I continue the Introduction, I'll try to show how these agendas belong together: how hyperspecialization turns out to be a very good reason to reconsider philosophical apriorism, and, conversely, that letting go of our disciplinary bias toward the a priori is needed to make sense of one after another aspect of our lives as hyperspecializers.

An invitation to turn our backs on apriorism is not the only attempt to revise current philosophical practice that the reader will encounter here. As an analytic philosopher, I grew up trained to approach problems as exercises in semantic or conceptual analysis and to understand metaphysics as having for its object descriptive theory with a distinctive look and feel. (I will describe it briefly below, and the conventionally trained philosopher learns to recognize it immediately.) I found that along with the pushback against apriorism, I needed to adopt a different way of thinking about metaphysics, in which it amounts to design analysis of intellectual or cognitive devices, and to treat metaphysics—construed this way, as intellectual ergonomics—as an alternative to semantic analysis. I understand very well that one doesn't want to be subjected to too many revolutions at once, but once again I will try to convey here how this shift in approach is necessary if we are to make sense of hyperspecialization.

### 1.3

What sort of problem does hyperspecialization put on our plate? Our reappropriation of the story of the Tower of Babel suggests that some of the most pressing issues will have to do with social coordination: that is, they will be

---

<sup>4</sup>Take the distinction with more than one grain of salt: mathematics is supposed to provide the paradigm of a priori knowledge, but Polya (1990), a thoughtful discussion of mathematical investigation, suggests that not even the paradigm stays on its side of the boundary. (See, for instance, the emphasis on observation and induction in his Chapters 1–4.)

<sup>5</sup>A recent movement, so-called experimental philosophy, presents itself as a turn away from apriorism, but let me suggest that you take that with a grain of salt, too. While work done under that heading is somewhat of a hodgepodge, the core of the movement consists in the social-science investigation of what people's "intuitions" are. Sometimes this sort of investigation is presented as debunking philosophers' claims to know things without looking, but not nearly as often as you might think. Sometimes it is presented simply as a way of determining just what the intuitions are, so that they can be used as a basis for further philosophical theorizing; see Weinberg (2013, pp. 93ff) for a description of the approach. Calling *that* a turn away from the a priori is a little like a creation science institute announcing that its research is now empirically informed, because instead of just telling you what the Bible says, from memory, they look up chapter and verse.



about getting along and managing cooperative enterprises with people whom we do not understand. If we think of the subject matter of ethics as centrally concerned with how human beings are to get along with one another, and if hyperspecialization requires us to rethink even the very basic social traffic rules, then we can expect it to set a new agenda for moral philosophers—as I argue, in Section 3.6 and in Chapter 9, that it does. If you unpack that very quick train of thought, you will see that it turns on a consideration that it's time to add explicitly into the mix.

Specialization is not in itself a distinctively human capacity, and since we were discussing the building of a tower, recall that other animals do on occasion execute relatively complicated cooperative construction projects: bees, for example, or ants.<sup>6</sup> However, the form that specialization takes in human beings is, as far as we know, unique. For a given species of ant or bee, the types of specialized task involved in setting up a nest or hive are fixed, and so where human beings sometimes decide that their next building will include features seen on no previous skyscraper (for instance, the sort of mass damper installed in Boston's John Hancock tower, and then in New York's former Citibank building), and that they will have to hire new kinds of specialists, bees do not. Previously unheard-of specializations cannot belong to a fixed repertoire; they cannot be, as we used to say, instincts. Thus people learn their specializations, and not always the same ones. They are able to work their way into any of an apparently open-ended list of disciplines or professions, and what is more, they are able to switch from one to another—although because the costs of doing so are high, these migrations do not happen very many times over the course of a life.<sup>7</sup> Thus someone might (real example) start out as a dancer, then work for an architecture firm specializing in theater design, then for an IT department, and subsequently become a vice-president of a bank. To have a way of marking this, I will say not just that human beings hyperspecialize, but that they are *serial* hyperspecializers.

Thus the social traffic rules cannot be fixed ahead of time as they are for the social insects. But then, if the configuration of differently specialized human beings changes regularly, if there are no rules for coordinating activities that are workable no matter what the roles are, and if it follows that the moral rules have to be rewritten equally regularly, then the distinctive form that

---

<sup>6</sup>For an overview, see Hölldobler and Wilson (2009).

<sup>7</sup>Perhaps other animals can switch from role to role also—see Gordon (1989) and Gordon (1986) on how colonies of social insects shift workers from one task to another—but, very importantly, not *this* way.

The observation that we can learn our way around more than one specialization does complicate my claim that we today make up a society of specialists who cannot understand one another. Some specialists *can*, on occasion, fully understand specialists of other types, when they themselves have mastered the other specialization. But, again, it is only possible for an individual to traverse at most a small handful of specializations. I will consider what to make of this complication in Chapters 2 and 11.

specialization takes in contemporary human society indeed poses problems for moral philosophers: most obviously, what the rules should be from time to time; almost as obviously, how we should go about adjusting them.<sup>8</sup>

I am going to be trying to convince you that the problems we face as serial hyperspecializers infiltrate the topics of traditional philosophical discussion at a great many points, and as I do so, to make the case that we have been doing the philosophy wrong: the chapters to come will be a problem-by-problem inventory of wrong turns that have come of overlooking specialization. Here is another of those problems, still closely related to issues in moral philosophy. Because we invent new specializations and discover new subject matters, we cannot already know what is important and what our priorities should be in those novel domains. Consequently, for serial hyperspecializers, practical reasoning (that is, thinking about what to do) involves learning what matters from experience—a claim I advance in Chapters 3 and 10. Thus, if I am right, serial hyperspecialization requires us to rethink our theory of rationality.

The conventional wisdom is that *learning* what to want and what to care about makes no sense at all. Consider an argument, which I think is often somewhere in the background but not generally spelled out, for moral theory or ethics constituting a stable subject matter. Only what's empirically observable is a posteriori. *What to do* and *what to care about* aren't anything that you can observe. Ethics and morality are about what to do, and maybe about what to care about, so they must be a priori. But an a priori subject matter can't change with the times; so ethics and moral philosophy are about what anyone could have understood at any time, even thousands of years ago. It would follow that nothing new in the way of specialization could make a difference to ethical theory.

Since I've just said that I think we learn what matters from experience, you can expect me to contest at least one of the premises of that argument.<sup>9</sup> In any case, not just our moral philosophy but our theory of practical rationality has been apriorist. The default view on this topic is still instrumentalism: the notion that figuring out what to do is means-end reasoning, and that it bottoms out in your desires. Again, to call something a priori just means that it's something you can know about up front, without looking, and put that way, it's obvious that instrumentalism is a practical version of apriorism. You just *have* these desires, from which you derive practical conclusions: because the

---

<sup>8</sup>For recent discussion of the idea that we may need to treat our moral guidelines as revisable, see Millgram (2009d), and Richardson (2013).

<sup>9</sup>For instance, Millgram (2005a, Chapter 1) describes one way we observe what is and isn't a good idea. Millgram (1997) gives an argument that we have to learn what matters from experience; the present discussion is one way of framing that argument, by providing a deeper explanation for some of its premises. I'll give a philosophers' example below, in Section 10.7, especially footnote 47, with follow-up in Section 10.9.

desires tell you what your objectives are, you know what your goals are *without looking*. You don't have to keep an eye out to see if it's time to change your desires; while they might *happen* to change, it's never required by what you see. Maybe you're born with them, or maybe not; maybe they spring to mind when you do look, but the world doesn't get to correct your desires or to tell you to desire things you don't happen to want. (Chapter 4 lays out the very sophisticated version of this view developed by Bernard Williams and tries to figure out just how it is motivated.) However, in the cognitive-function way of looking at things, desires—or whatever their replacements turn out to be—have to make design sense: they have to be configured as part of a larger design directed toward the environment of the creature in which they are components. In Chapters 3 and 4, I will argue that serial hyperspecializers could not possibly get by using desires as philosophers have characterized them. Therefore, the drivers of practical reasoning cannot be—and, more decisively, had *better* not be—desires.

These specialization-driven reconceptions of the intellectual devices on which we rely tend to spill over from problem space to the next. Another adjacent philosophical subspecialty, the theory of agency, is more or less the theory of how we do what we do. A different way of thinking about what to do, one on which you have to find out what matters and what is important as you proceed, ends up entailing a different manner of proceeding when you do it. And so agency is going to have to be rethought along with practical rationality—or so I argue in Chapter 10. Attempts to understand agency have recently been focused on characterizing the differences between—to use a representative example—wanting something and *really* wanting it. If they are worth hanging onto, those higher-octane “really wantings” must have a role within the lives of serial hyperspecializers, and one question I will broach is just what that role might be.

So here is a first round of advice on traversing the book: If you are primarily interested in practical reasoning and moral theory, start with Chapter 2, which does a second pass over what I take to be the central problems produced by specialization, at a somewhat higher level of resolution. Continue with Chapters 3 and 4, which try to show that our current understanding of practical rationality is a bad fit for the sort of creatures we are; Chapter 10 does the same for contemporary theory of agency. If your primary interest is in moral theory, Chapter 9 explains one way in which we will need to reconsider our ethics. If you have an interest in metaethics, Chapters 5 and 6 take a shot at both dislodging the standard frame in which work in this area gets done, and showing how to go about doing metaethics for serial hyperspecializers. Finally, if your primary focus *is* one specialized field within moral philosophy, Chapter 11 considers how that came about, and whether it is a philosophically defensible way of demarcating the scope of one's philosophical interests.

## 1.4

So far, perhaps so plausible, but I am also going to argue that serial hyperspecialization requires us to rethink our metaphysics. That is likely to raise some eyebrows, but before explaining why, and why it's alright to let the eyebrows come back down, let's detour to take up a preliminary question. I mentioned a few steps back that analytic philosophers don't concern themselves with the topics I propose to treat. Philosophers have been aware of the division of labor at least since Plato and have certainly noticed some of the problems it raises.<sup>10</sup> But I am going to be taking up a number of issues that have not come in for previous discussion. Why haven't they? If the problems of specialization management that I am going to engage are new to the field, that requires explanation; surely specialization is too obvious not to notice. We do already have a fallback explanation: philosophy's longstanding apriorist bias makes it hard for philosophers to take notice of social change. But even if that explanation is correct as far as it goes, we should want to deepen it: what accounts for the apriorism in the first place?

I'm going to be arguing that philosophers have in fact encountered a number of problems arising out of specialization, but have almost always failed

---

<sup>10</sup>For instance, in the *Ion*, Plato's Socrates argues that literary criticism confined to a particular figure can't be an area of specialization. A representative argument has it that the person best placed to assess discussions of a particular subject matter is an expert on that subject matter; that the subject matters discussed by different authors will overlap; that consequently anyone who is best-placed to assess one author's work will be best-placed to assess many other authors' work as well. Since Ion's profession, that of rhapsode, was built around the assumption that you can be an expert in just *one* literary figure (Homer, in his case), it follows that the profession of rhapsode was a misguided institution—and if Plato was right, today's departments of languages and literatures, along with much contemporary history of philosophy, are likewise misguided (Plato, 1997, 531a–532b). This is Plato giving us reading instructions to which, remarkably, Plato scholars seem oblivious: he's telling us that there's no such thing as an expert on Plato; that if you're reading Plato on ethics, the relevant background to bring to it is ethics, not history of ancient philosophy; that if you're reading Plato on metaphysics, the relevant background to bring to it is metaphysics, and not history of ancient philosophy. For our own purposes, the problem Plato is highlighting is that of distinguishing genuine areas of possible expertise from pseudospecializations. See LaBarge (1997) for a recent overview of Plato's take on related questions.

Again for instance, Nietzsche (1988, vol. iv, pp. 178f) complains about 'inverse cripples' (in Walter Kaufmann's translation); the concern is, in part, that a specialist's hypertrophied personality exhibits spiritual deformities.

One more example: Adam Smith assumes that the upshot of division of labor is "the man whose whole life is spent in performing a few simple operations," which Smith expects will undermine the intelligence, character, and civic role of "the great body of the people" (1981, pp. 781f, 788). If I am right, Smith has gotten the problem almost exactly backwards—perhaps because he was writing so early on in the process of specialization. Our hyperspecialized experts build up ever more demanding skill sets, which they are frequently required to exercise with ever more intelligence and ever more creativity.

This is a good place to register that not all division of labor involves the sort of specialization of interest to us here; the fragmentation of activity characteristic of old-time assembly lines, or assignment of generic tasks by gender, can be put to one side.

to recognize them as having to do with division of labor.<sup>11</sup> Instead, they have imagined themselves to be theorizing about, to put it a bit baldly, exotic invisible objects. One of my objectives, over the coming chapters, will be to make the case that work in metaphysics has on more than one occasion involved this sort of misunderstanding. If analytic philosophers haven't seen the problems posed by hyperspecialization, or seen the problems as their own problems, one very important reason is that they have taken them as an occasion for metaphysics—while misinterpreting what the enterprise of metaphysics is about and what it can reasonably be expected to accomplish.

Metaphysics as it's practiced today by analytic philosophers—*old-school metaphysics*, just to have a short label for it—has three features that concern us. First, it is a theoretical enterprise, that is, one that attempts to settle questions of fact. For instance, metaphysical theories purport to tell you what the world is really, at bottom, composed of; or what sorts of things there are in general; or how the things that aren't the basic ones depend on or are derivative from the things that are.<sup>12</sup> Or—and I'll discuss this example at length in Chapters 7 and 8—they purport to provide a theory of modality, that is, of what we're talking about when we say that something *might* have happened, or that it *would* have happened if something else had, or that something *has* to happen. By way of conveying to first-timers what comes under the heading of those exotic invisible objects I mentioned a moment back, one such explanation of modality, very respectable within the field, but which properly strained the credulity of outsiders, was given in terms of configurations of parallel universes, which were claimed to exist alongside our own. Finally for now, there's a branch of metaphysics that is cross-classified as a part of moral philosophy: metaethics is the metaphysics of ethics or morality. Over the past century, metaethicists devoted themselves to the analysis of a remarkable force, "normativity," which I will discuss in Chapter 5. So, and finally for now, metaphysical theories purported to explain the very special oomph in virtue of which you *ought* to do one thing or another.

Second, within analytic philosophy, the default method of solving a problem has been to figure out what some class of expressions means and, when the topic is metaphysical, to then read one's metaphysical theory off of one's semantic theory. To arrive at that theory of meaning, they typically ask themselves what we would say about one or another imagined circumstance. There's a secondary method, called "reflective equilibrium," which I won't emphasize just now, but which amounts to adjusting whatever prior opinions one has about a topic in the direction of greater coherence. Streamlining the

---

<sup>11</sup>"Almost always": among the exceptions, Putnam (1975b) investigates the adaptation of reference, meaning, and natural-kind concepts to what at the time he called the linguistic division of labor; however, we will not be taking up these topics in philosophy of language here.

<sup>12</sup>Schaffer (2009) can serve as a representative pronouncement; he asks metaphysicians to spend less time on what exists (thereby acknowledging that that's most of what they've been doing recently) and more on what depends on what else.

discussion by putting reflective equilibrium to one side, and in a somewhat dated way of saying it, philosophy is treated, and not just by the metaphysicians, as conceptual analysis. Since it's not only concepts that get analyzed, I'll say that in the analytic tradition problems are addressed in the first place by *semantic analysis*.<sup>13</sup>

Third, and this follows from the features we've just enumerated, old-school metaphysics has been apriorist. When metaphysics is driven by semantic theory, we are being recommended one or another account of the way things are as *implicit in the way we speak*. When opinions ("intuitions") are being adjusted into a reflective equilibrium, the ensuing account is being recommended as *implicit in the way we are accustomed to think*. Because the theory is arrived at on the basis of how we *already* speak, and what we *already* think, it is arrived at without looking; that is, it is treated as *a priori*.<sup>14</sup>

---

<sup>13</sup>For a critical description of reflective equilibrium, see Millgram (2005a, pp. 8–10). My discussion here will be streamlined in part because the remarks in the penultimate paragraph of the section will stick to reflective equilibrium, *mutatis mutandis*; in part because the pros and cons of reflective equilibrium deserve their own discussion, which I hope to give them elsewhere; in part because it does seem to me that within old-school metaphysics, the lion's share of the work has in fact been allocated to semantic analysis.

Having introduced that description, let me qualify it: not all old-school metaphysics can be plausibly construed either as semantic analysis or as an attempt to arrive at a reflective equilibrium. To show why, I'll use an example introduced in the previous note: Schaffer's pronouncement that metaphysics should be not be about what exists, but what grounds what, and what items are fundamental.

You might think that if "grounding" is to be the focus of the enterprise, we should be getting an explanation of what it is, but Schaffer says that he won't tell you: "Grounding should . . . be taken as *primitive* . . . [It] is an unanalyzable but needed notion— . . . the *primitive structuring conception of metaphysics*." Again, "it is clear what we mean . . . the notion of grounding may be *unfamiliar* to some metaphysicians raised only on Quine and Carnap. The best advice I can give is *work with the notion*, and see if you then come to grasp it" (Schaffer, 2009, pp. 364, 376).

Now, I'm told that Ernst Gombrich used to play a game with his classes that went like this. He would introduce two classifiers, *ping* and *pong*, but without telling the group anything about them at all; then he would have them vote on whether one item or another was ping . . . or pong. (He would start off with mice and elephants, but eventually work his way up to Rembrandts.) In these rounds of voting, it was possible to elicit a remarkable degree of convergence as to what was ping and what was pong. (For a short description, see Gombrich 1984, pp. 370f.) Gombrich, oddly enough, seemed to think that there were *correct* answers; but since nothing at all had been said about what the terms meant when they were introduced, what was being exhibited was a phenomenon of potential interest to cognitive scientists or social psychologists, not the meaning of the predicates.

To invite metaphysicians to start using the primitive notion of grounding, when you haven't done anything to introduce it, is just playing another round of "ping and pong". What Schaffer is doing is not all *that* uncommon in the practice of old-school metaphysics, and whatever you think of it, it's not semantic analysis.

Nonetheless, I think my characterization of the drivers of contemporary analytic metaphysics is fair in the main.

<sup>14</sup>But here is a further qualification: not quite all current metaphysics is apriorist. For instance, Ladyman and Ross (2007) exemplify a minority approach; they dismiss metaphysics that bottoms out in "intuitions," linguistic or otherwise, but share the view that metaphysics is in the business of producing a true and very general theory about the objective world. Since science produces objective truths (physics, the general ones, and the special sciences, the local ones), it follows, they think, that scientific results are to replace the intuitions. (In this case, the proposed unifying background theory lends itself to a thumbnail caricature: "Can't you see? It's all *patterns* . . . patterns all the way down!") I'll return briefly to this variation on the metaphysics-as-theory approach in Chapter 11, note 13, below.

If that's right, we shouldn't be surprised that analytic metaphysics is treated as having as stable a subject matter as moral philosophy. That old-school metaphysics is oblivious to recent changes in human social life follows from its a priori methodology. If you permit yourself the apriorism, you won't be attending to social change: like other forms of change, you only find out about it by observing it. In addition, because the focus of metaphysics was taken to be the most basic things that compose the very bottom layer of the world, and because philosophers don't expect those most basic things in the world to change, it has been taken for granted that nothing new in the social order (such as novel forms of specialization) could make a difference to metaphysics. All in all, analytic metaphysicians have adopted the attitude of Adam Sedgwick: "Were nature changeable, there could be no philosophy."<sup>15</sup>

Perhaps the way we are accustomed to speaking and thinking is a repository of folk wisdom. But because semantic theories at best elicit what we collectively *already* mean (usually, since languages change slowly, what we have meant for a long time), the most we will uncover through semantic analysis are aspects—those having to do with choices of what to mean by what—of solutions to problems that were old enough to have been seamlessly woven into the language.<sup>16</sup> Likewise, reflective equilibrium systematizes what we collectively think *already*; it largely draws on opinions that are old enough to have taken on that patina of respectability registered by phrases like "common sense." But suppose that I am right in thinking that specialization has very recently and very suddenly become much more extreme and has turned into a different sort of beast than it used to be. If the problems posed by serial hyperspecialization are new, then semantic analysis is at best part of a ground-clearing operation: a survey of older solutions to past problems, which one might reasonably conduct as preparation for moving on. And systematizations of older common-sense opinions have that very same status as well.

Because resisting apriorism is a theme of the chapters to come, I am going to owe an account of the a priori, and now is a good occasion to preview it. If semantic analysis amounts to reconstructing aspects of old solutions to old problems, solutions so old that we have forgotten the reasons our ancestors adopted them, we should expect aprioricity to turn out to be a sort of institutionalized amnesia. In Chapter 8 (and secondarily in Chapter 6), I am going to make a case, albeit a preliminary case, that that is what it is.

---

<sup>15</sup>Quoted in Hilton (2006, p. 447).

<sup>16</sup>Only aspects, because when you are solving a real-world problem, assigning meanings to symbols is generally just one component of a more complex arrangement. For instance, you don't understand how the traffic is being managed if you confine yourself solely to figuring out what the inscriptions on the signage—"STOP," "YIELD," "ONE WAY"—*mean*.

## 1.5

There is a much better way to understand the enterprise of metaphysics: as *intellectual ergonomics*. Throughout its history, or so it seems to me, metaphysics has been the answer to the question: how do we have to understand the world for reasoning about it to be possible?<sup>17</sup> There is a *practical* spin to put on the question: how can we make reasoning and inference tractable and effective? One approach to the question so understood will focus on the design—and on the redesign—of intellectual devices that make it feasible to think about the world through which we must navigate ourselves. If that is what we are interested in, when we are doing metaphysics, then we should be attending not in the first place to what expressions *mean*, but to what the devices *do*. I aim to demo what this looks like over the course of several of the chapters in this volume.<sup>18</sup>

When I say that metaphysics is intellectual ergonomics, I mean that it is in large part an *applied* science, and important precisely for that reason.<sup>19</sup> However, I'm not suggesting that all of the ergonomics is directed toward problems

---

<sup>17</sup>Metaphysics, construed as a theoretical enterprise, has the look and feel of a very peculiar would-be science, one that seems to share a subject matter with physics. That such an academic enterprise could so much as exist should strike you offhand as a scandal. Who could take seriously such a parallel physics—a science that gives answers to some of the very same questions to which physics addresses itself, but *different ones*? (And worse, without any of the empirical predictive power that we have come to expect from real science.) Construing the field as attempting to answer the question I have just flagged explains why, despite its ups and downs, metaphysics hasn't merely been incompetently executed physics.

<sup>18</sup>I originally became convinced that metaphysics had to be understood as an enterprise with this sort of practical point while trying to find a philosophical home for the great many decisions on our part as to which partial truths—what idealizations, approximations, and so on—will best facilitate thinking our way through the problems we face. (Partial truth is also not the conventional wisdom, but I argue for it in Millgram 2009a.) Although this is not the topic of the present book, it now seems to me that serial hyperspecialization is the correct frame for those earlier arguments. Because our specializations come and go, so do the systems of representation we use to navigate them; even collectively, we don't have the time it takes to get our vocabularies neatly tailored to their subject matter. And so we have to think using descriptions that often don't quite fit what we are trying to describe. One consequence is that we need to broaden our understanding of reasoning about matters of fact to accommodate steps in arguments that we know not to be true, but which nonetheless are true *enough*.

Thus one branch of metaphysics has to do with choosing those misrepresentations—when we are self-aware about what we are doing, those *intentional* misrepresentations—of our surroundings that make it possible to reason effectively about them. However, as will be clear from the discussion to come, it is not my view that all of metaphysics has to do with designing and implementing approximation techniques.

If all that is correct, this is a further way in which serial hyperspecialization requires us to rethink our understanding of rationality—in this case, theoretical rather than practical rationality.

<sup>19</sup>It's worth mentioning in passing a further explanation for apriorism in metaphysics, one that John Dewey diagnosed as a vestige of ancient Greek class consciousness (Dewey, 2008a, ch. 1): the insistence that metaphysics isn't at all practical. Not just in ancient Greece but today, to engage in an activity that is thoroughly useless counts as a form of conspicuous consumption. And so we see philosophers—who in fact have to work for a living as salaried schoolteachers—presenting their



arising out of specialization. Some such devices are best understood as handling other issues entirely, and some of these improvised expedients resemble Swiss army knives in being compromises between many simultaneously imposed ergonomic desiderata and implementation constraints.<sup>20</sup> The present choice of theme is the way that division of labor makes sense of our intellectual toolkit, but there are other parts of the toolkit that no doubt have to be understood by attending to very different features of human life.

As I promote the strategy of replacing semantic treatments with cognitive function analyses, I will use two of the traditional problems I have mentioned as my testbed: again, what to make of the force of an “ought,” and what to make of modal thought. I also earlier remarked that analytic metaphysicians have often been in the business of providing theories of invisible objects, and here is a plausible hypothesis as to how that might have come to pass. Treatments both of modality and of that very special force of an “ought” are good examples. Something that *has* to happen (that’s *necessary*) looks precisely the same as something that merely *does* happen; again, you can see what *happens*, but you can’t ever see the *ought*. If someone starts looking for what the “has to” or “should” represents—for a semantic value that is analogous to the semantic values of the other parts of the sentences in which they figure—we shouldn’t be all that surprised when it turns out that they are construed as picking out very peculiar things that are *there*, but which you can’t *see*. Thus, if I am right, the problem arises out of taking understanding necessities and oughts to be a matter of saying what such terms mean.<sup>21</sup>

In Chapters 6 and 8 I will develop my alternative: both “oughts” and necessities can be seen as intellectual devices with a cognitive function, and they turn out to be much less mysterious than the invisible objects posited by the semantic analysis approach. But I don’t know that it’s reasonable simply to demand of philosophers who have been trained on one method suddenly

---

discipline as being of solely ornamental value. I’ve registered that the deliverances of old-school metaphysics are generally out of step with the sensibilities of people who live in the real world—in an example I adduced a few paragraphs back, and which I’ll discuss in Chapter 7, one very famous theory of what it is for something to be *possible* amounted to saying that it is taking place or is located in a parallel universe—and that too strikes me as being of a piece with this sort of class positioning. To display that you are exempt from having to be level-headed in your opinions has the same sort of social payoffs as showing that you are exempt from making contributions of practical value. I suggested a moment ago that resolving problems raised by the distinctively human form of division of labor was of intense practical interest. Philosophers who act as though working on problems of practical interest counts as a demotion will be disinclined to see such problems as their own.

<sup>20</sup>For some discussion of personhood as an ergonomic expedient, but one that does not focus on specialization, see Millgram (2014b); I point toward the complications introduced by hyperspecialization for our older concept of the person in Millgram (2011b).

<sup>21</sup>However, I don’t mean to suggest that there were not other motivations as well; one of these, which will appear around the edges of our treatment of “ought,” is the sense that one can only be properly confident about the force of a requirement or demand if it is understood as a *thing*: roughly, the mistake made by proponents of the gold standard when they insist that unless it represents a special sort of object that is intrinsically valuable, the money is just *paper*.

to abandon it. Thus Chapters 5 and 7 are meant to show how and why the semantic analysis approach to these metaphysical problems is in any case unsuccessful. I will try to strike a balance between taking account of the ins and outs of particular old-school views and framing the treatment in a way that generalizes—so that what is seen to be unworkable is not merely one or another metaphysical theory, but rather the old-school way of solving philosophical problems.

As I proceed, I'll be attempting to exhibit our "oughts" and "musts" as cognitive expedients improvised in part to handle older and less extreme forms of specialization than we face today.<sup>22</sup> Normally, when we reconstruct an intellectual device in our repertoire that addresses specialization in one way or another, we are examining a solution to a predecessor of the problem as we currently have it, although sometimes one in the process of being exapted to our more demanding social and technical environment. We want to understand how our older intellectual tools addressed the predecessors of our current problems so that we can invent—this time around, in full awareness of what we are doing—better successors for more demanding circumstances. Metaphysics, understood as intellectual ergonomics, is first of all a *design science* (like architecture, or computer science, or mechanical engineering).

To be sure, we are not interested solely in successors to the ergonomic equipment we now have; not all of the problems that we owe to our Tower of Babel have precedents with even improvised solutions. For example, because we are now in the business of deploying cooperatively assembled arguments, where no one is competent to assess the argument as a whole, older ideals of autonomy, intellectual and otherwise, no longer have much to do with today's realities—or so I will argue in Chapter 2. We will need to devise new methods of managing the interfaces between disciplinary specializations; and this is, from the vantage point of philosophy, an entirely new task.

Returning to a roadmap for the reader: If your primary focus is metaphysics and epistemology, you can traverse the volume as follows. First read Chapter 2, which provides an overview of the central problems. Then proceed to Chapter 4, which talks through the intellectual trajectory of Bernard Williams's work, in the service of showing that a number of assumptions used in picking out the subject matter of old-school philosophy of language and epistemology—I mean, the notion that the concepts we want to analyze first and foremost are truth, knowledge, and so on—make sense only in the prespecialized world. If you feel at home in analytic metaphysics, Chapter 7 will try

---

<sup>22</sup>We should not assume that the older solutions to those older problems are prehistorical: that Strawsonian "descriptive metaphysics" (Strawson, 1971) is *caveman* metaphysics. Lee (2005) argues that what now seem to us to be fundamental distinctions in the metaphysics of mind were innovations dating to roughly Aristotle's generation, and Williams (2002, Chapter 7) makes the case that our present folk metaphysics of time emerged only slightly earlier; Mann (2000) holds that the basic intellectual equipment of individuals and kinds which we still deploy is due to Aristotle himself.

to convince you that you should be having second thoughts, by showing how Lewisian theories of modality fail in their own terms. Then proceed to (or, if you're not invested in Lewisian modal realism, proceed directly to) Chapter 8, which shows how to approach the topic of modality from the standpoint of serial hyperspecialization. If you find the approach plausible and would like to see a second illustration, Chapters 5 and 6 provide one from the metaphysics of (so-called) normativity. And as before, if your primary focus *is* a specialized field in metaphysics and epistemology, Chapter 11 considers how that came about, and whether it is a philosophically defensible way of demarcating the scope of one's philosophical interests.

## 1.6

Before Aristotle started in on one topic or another, he would survey earlier work, in pretty much the mode of someone taking inventory of building materials on a construction site.<sup>23</sup> He can serve us as a role model: I've just suggested that once we recognize previous responses to predecessors of the problems we ourselves face, we will have made some headway, in that those intellectual devices are raw materials that we can adapt and incorporate into our own ergonomic designs. We should not think of them in the first place as theories; the theories of old-school metaphysics are, most of them, misunderstandings of cognitive tools, and it is those tools that I have in mind as resources, to be altered and integrated into new and more adequate tools. But there is another sort of resource to which I also want to turn our attention, which does not look so much like a device as a capacity.

A serial *hyperspecializer* is a good deal like an organism that is tightly adapted to a narrow ecological niche—a comparison I will develop further in Chapters 3 and 4. (In the human case, the ecological niches often end up looking like professions or disciplines or social roles.) However, *serial* hyperspecializers are not only adapted, but adaptable, in that they can move from one niche to another. Other niche dwellers—think of a species of ant that is specialized to the roof canopy of tropical rain forests—do not move from one niche to another (the ants do not decide that the time has come to leave the trees and settle into a prairie). Humans do, and what is more, they do not merely *find* new niches in their complex ecology: they *create* them. Moreover, although we are serial hyperspecializers, we are not merely niche dwellers; the world contains niches, but not all of the world is composed of them, and

---

<sup>23</sup>Readers familiar with Aristotle's contrast between potentiality and actuality should be aware that the comparison can bear a good deal of weight. His predecessors' views are being assessed to determine their potential for being incorporated into his own: for their potentialities, in his technical sense of the term.

sometimes, when we have to make our way in the less structured zones of our environments, specialization is not a viable strategy.

It's not well understood how we do these things, but since we *do* do them, we must have whatever it takes to perform these tasks. Chapters 3, 4, and 10 are contributions to a task analysis that has something like this form: to explore a novel environment in this way, you have to be doing such and such; we explore our environments that way; so it follows that we are able to do such and such. Those such and suches—that is, the capacities I am going to be identifying—are available when we are attempting to design new ways of handling serial hyperspecialization.

Bear in mind that one has these capacities even when one is not moving from specialization to specialization; even if you end up never specializing yourself, or never switching jobs, you are still a serial hyperspecializer, in that you have the wherewithal to do both. Moreover, even if you never switch jobs, the constellation of social roles that makes up your environment is produced by other people exercising those capacities. So making sense of these capacities is an essential step in apprehending both what sort of environment you are coping with and what aspects of yourself you will be mobilizing when you do.

The primary reason these capacities are so little understood is that our apriorist legacy consigns them to neglect. As a matter of strategy, apriorism about both practical rationality and our intellectual apparatus amounts to presupposing that the world as a whole is a very large, very stable ecological niche. Return for a moment to the instrumentalist model of practical rationality, on which your goals provide all of your reasons for action. Your goals designate what you *already*, without looking, know to pursue; they are usable guides to action only against a quite restricted range of features of your environment. (A trivial example: being a coffee snob tells you what to do if you are living somewhere coffee is available, comes in better and worse, and you are able to afford the better coffee.) Simply to take for granted that the goals you happen to have are effective guides is for practical purposes to presuppose that the enabling background is both pervasive and stable: that even if you move from one location to another within your world, those features will still be present, and that they are not going to change if you stay where you are. That is effectively to presume that they are deep structural features of your world, which is in turn to presume that your very variegated and fluid world has the structure of a niche: a sort of structure that is only found in smallish sectors of—*inside*—the world. In the initial chapters of this volume, I will be criticizing what looks very much like an old-time category mistake (the world is not a niche within it), but for now, the point is this: because we philosophers have been behaving as though it were not a mistake, we haven't so much as inventoried the abilities of which we avail ourselves when we are not inside very familiar territory. Before we go ahead and take a shot at replacing our jerry-rigged metaphysics

and our obsolete ethics, we need to have on hand a much more inclusive list of the resources (both the cognitive devices and the native abilities) that we can draw upon.

## 1.7

I've been introducing the philosophical problems involved in managing serial hyperspecialization—only introducing them, however, because a proper description will only emerge from the chapter-by-chapter discussion. And I have just previewed a survey of the resources available for solving them. So you might expect that I am about to promise a solution. This may seem like sudden swerve, but such an announcement is not in the offing, and I want to wrap up the Introduction by explaining why.

There is an open-and-shut reason for not advancing a series of proposals for managing serial hyperspecialization here, and I will discuss it in Chapter 11. To anticipate, once we have seen what these problems amount to, and how they arise, it will turn out that they must be addressed piecemeal. Because the solutions must be positioned on the interfaces between different disciplines, and because addressing different problems requires different training, the work of developing them has to be done by a great many differently trained people.

But there is also a deeper and more important reason that no one philosopher should want to move too quickly toward a set of solutions to the problems posed by serial hyperspecialization. Because adopting responses to *these* challenges amounts to changing what we are, it is a very big decision. Moreover, the adoption must be collective in order to be effective, which means that no individual can make the decision. We can pursue these points by considering further just what sort of decision we are looking at.

Immanuel Kant gave a famous list of questions to which philosophy is centrally addressed: What can I know? What ought I to do? What may I hope? And at some point he supplemented that list with a further and final question: What is Man?<sup>24</sup> There seem to me to be two related ways of answering this last question that are to the present point. One would be something on the order of a description I have presented here, of human beings as serial hyperspecializers. That is, the first sort of answer gives a characterization of what Aristotle would have called the human *ergon*: it says what people are by saying what they *do*, or *how they work*. There are many quite familiar characterizations of this form: Man is the animal that cooks.<sup>25</sup>

---

<sup>24</sup>The shorter list is in his first *Critique*, at A804f/B832f; the addition is in the *Jäsche Logic* (Kant, 1992, p. 538), that is, in Kant (1902–, vol. 9, p. 25).

<sup>25</sup>This is a much more revealing characterization than one might at first think, for reasons to be found in Sterelny (2010).

The other is less familiar and requires brief stage setting. A review of the history of philosophy makes it apparent that the problems of philosophy are perennial: Plato's problems have not been solved, and they are still our problems; our problems were, whether he realized it or not, Plato's as well. (Mostly he did realize it: Plato was one sharp cookie.) The problems of philosophy are closely connected to one another, in that solving one presupposes having a solution to others, and in that taking a stance on one has consequences for what you think about other problems that depend on it.<sup>26</sup> And some of the problems of philosophy have to do, quite directly, with what it is to be a human being trying to get around in the world. Just so you have a sense of what I mean here, Kant's first two questions—"What can I know?" and "What should I do?"—arise only for creatures that face the sorts of epistemic and practical challenges we do: ants, bees, and termites don't need answers to either of them. If I'm right, the other sort of answer we can give to Kant's fourth question is: human beings are the creatures for whom *these* (the problems we philosophers have not been able to get away from or solve for over two thousand years) are the perennial philosophical problems.

The two sorts of answers to Kant's fourth question that I've just juxtaposed are themselves connected. If I am right in thinking that metaphysics is intellectual ergonomics, and that the problems to which we need ergonomic solutions crop up as we try to get along in the way that is characteristic of us—when we do what people do, working the way that people work—then the perennial philosophical problems are problems for creatures with a particular *ergon*. Ethics clearly enough has always been specific to humanity; again, ants and bees would not be good clients for, and do not need, any moral philosophy that might suit us. Metaphysics, ethics, and even epistemology have been attempts to make sense of intellectual devices that themselves only have a point within the context of, to borrow a phrase from Ludwig Wittgenstein, a particular form of life.

The cognitive devices we deploy and that old-school metaphysics misunderstands belong to that older design characterization of humanity. But if people are—and have only recently become—serial hyperspecializers, our design characterization has changed. Our older cognitive repertoire is being exapted and expanded to meet the new problems we face; and I have been recommending innovation that will equip us with a new intellectual toolkit. The new toolkit will pose new problems, and these will be specifically *philosophical* problems—recognizably the successors to the former collection of perennial problems. For what it is worth, this does seem to promise progress in philosophy: the new problems may be more tractable than the old, but in any case they will displace the old.

---

<sup>26</sup>I talk through the pattern at greater length in Millgram (2013).

For instance, Kant's first question, "What can we know?" is tied to an epistemological success concept that is suitable for creatures whose information is stored as *truths*: knowledge, after all, consists in things you think that are true (and satisfy controversial further conditions as well). In Sections 4.3–4.6, I will suggest that serial hyperspecializers are in the business of thinking with discipline-local idealizations, approximations, and other representations that they understand perfectly well to be, strictly speaking, false—but true *enough* to work with. Accordingly, knowledge is not nearly as central a success concept for us today as it no doubt once was for our predecessors.<sup>27</sup> Instead, we need success concepts that will allow us to assess and regulate our uses of such partial truths; those concepts will pose philosophical problems of their own; we are becoming the creatures for whom developing, controlling, and understanding those success concepts is a central philosophical task.

The problems that serial hyperspecializers face are in the first place practical. Thus to choose solutions to them is to make a choice about what the *ergon* of a serial hyperspecializer is: it is to make a choice about what we do and how we work. And it is also to opt for one rather than another set of philosophical problems; it is to choose to be the creature for which *those* (rather than these) are the philosophical problems. Thus it is to change what counts as a correct answer to Kant's fourth question. That would be a very big deal, and before we start redesigning ourselves—before we even start to sketch out possible makeovers, much less settle on one of them—we had better pause and take stock. This book is meant as a start on the stock-taking.<sup>28</sup>

---

<sup>27</sup>See also Millgram (2009a, Section 6.5). Of course, even our ancestors made use of partial truths of various kinds; the point at hand has to do with how much more we have come to rely on them in the meantime. Previous philosophers approached epistemology with the impression that understanding reasoning with approximations and the like could be postponed until the pivotal concept, knowledge, had been analyzed; that posture is no longer tenable, and it is rather the analysis of knowledge that can be backburnered.

<sup>28</sup>For comments on an earlier version of this Introduction, I'm grateful to Chrisoula Andreou, Heather Douglas, Matt Haber, Aubrey Spivey, and Cynthia Stark; I was also helped by feedback from an audience at the Università degli Studi di Modena e Reggio Emilia.

## The Great Endarkenment

Pretty much all of us take the Enlightenment for granted. And when I say “all of us,” I mean not just this essay’s primary audience, analytic philosophers, but also we academics, we participants in Western market economies, we citizens of Western democratic republics. We acknowledge that there are cultures that have not assimilated this initially European event, as well as pockets of pre-Enlightenment ways of thinking here at home—New Age spiritualism, “alternative” medicine, palm readers, feng shui, astrology, and so on—but we think of them as belonging to an unassimilated periphery. Here in the center (in mainstream political debate, in the corridors of power, and in the ivory tower), we act like it’s a done deal. Just for instance, the triumph of the Enlightenment seems so secure that we no longer pursue mopping up operations: why even bother denouncing the residual superstitions I just mentioned?<sup>1</sup>

---

<sup>1</sup>Certainly there have been attacks on the Enlightenment from within the academy: for instance by Michel Foucault and Alasdair MacIntyre (e.g., Foucault, 1988, 1995; MacIntyre, 1988, 1990). But these are positioned as dissent against an overwhelmingly dominant paradigm. This isn’t the place to amass evidence of its dominance, but here are a couple of observations meant to give you a sense of what sort of data one would compile.

In what seems to me to be of a piece with our unwillingness to complain to the newspapers about their astrology columns, the old ideologues and polemicists (think François-Marie Arouet, or Julius Wellhausen, or Jeremy Bentham) strike us as quaint and undignified, in part just because they *were* polemicizing; we find it hard to imagine ourselves into a world in which you had to fight, and fight dirty, for what every reasonable person knows is plain common sense.

A further such indicator of where we have arrived is the status of naturalism, and its more specific relative, physicalism, within analytic philosophy: on the one hand, naturalism (and usually physicalism) seems to almost all card-carrying philosophy professors to be not just true, but *necessarily* true. On the other, I have yet to meet a philosopher able to give a coherent rendering of the content of either doctrine. (Stoljar 2010 is a useful overview of the apparently insuperable difficulties involved in spelling out the content of physicalism.) It might seem remarkable that the adherents of such widely shared views are unable to so much as put them into words, but the history explains how this peculiar state of affairs came about. The naturalist stance is the rejection of pre-Enlightenment superstition, and the victory has been so overwhelming that, like the early monotheists described by Yehezkiel Kaufman



But the very victory of the Enlightenment has been its undoing—and not at the cultural and political periphery, but, I am going to argue, within the academy, the corporate world, technocratic bureaucracies, and democratic institutions of governance. I am going to explain how it has happened that a side effect of the Enlightenment's success has made its ideals no longer attainable, and how, slowly but surely, what the Enlightenment took to be its most important achievements have been undone. We are now living in a transitional period, which we might as well call the Great Endarkenment: the verge of the new age of superstition that we will enter if we do not understand and come to terms with the problems to which the Enlightenment has given rise.

## 2.1

What *was* the Enlightenment? Many things, of course, but as it wrapped up, one of its most distinguished theoreticians summarized its central commitment as growing up into independence of judgment. “Enlightenment,” Kant announced, “is man's emergence from his self-incurred immaturity,” and he further explained what he meant by that: immaturity is a matter of letting other people do your thinking for you.<sup>2</sup> The Enlightenment's many rejections—of prerepublican modes of political organization, of inherited theological doctrines and ecclesiastical privilege, of philosophical traditions that seemed to be too tightly entwined with social forms of the past, and of just plain magical thinking—can be seen, in retrospect, as upshots, or even just symptoms, of that central commitment to thinking for yourself. I will shortly suggest that the commitment to thinking for yourself has turned out to be self-undermining. But first, and now I am redescribing the Enlightenment from perhaps an unusual point of view, that of a philosopher of logic, I want to consider what assumptions could reasonably motivate the demand.

Imagine what it would be like if confirmation theory, as envisioned by logical positivists such as Rudolph Carnap, had worked out: if there were a routinized method for systematically deploying evidence to arrive at conclusions that the evidence supports. (In their program, deductive logic was held up as the model for analytic treatments of inductive inference; the aspiration was to make conclusions about matters of fact mechanically checkable.) If it were possible to say, across the board, what evidence can consist in and how it can be collected (as in the logical positivists' so-called protocol sentences), and if the technique of bringing it to bear were proceduralized, and if the

---

(1953/1956), we simply no longer understand the views being rejected. Just as the mischaracterization of sophisticated pagans as worshippers of sticks and stones witnesses the ruthlessly effective suppression of the pagan tradition in early Israelite society, so our own inability to say what naturalism means is the starkest witness to how decisively we *won*.

<sup>2</sup>Kant (1991, p. 54); I have deleted his emphasis.

procedure reflected an independently specifiable criterion of correctness, then there would *just* be a right answer and a way to get it: empirical inference correctly done could always be the mere calculation of what is already, prior to the calculation, the correct result. Similarly, imagine what it would be like if a calculation—perhaps the sort of arithmetic exercise envisioned by the early utilitarians—sufficed for making social and individual choices. If the inputs to the procedure determined what choices (of acts, or of rules, or for that matter of social institutions) were correct, and if correctness of those practical conclusions, given those inputs, were independent of any deliberation or choices made by any actual human, then, like the operations of deductive logic, all practical conclusions would be mechanically checkable.<sup>3</sup> Let's call what we're imagining the *procedural utopia*.

Now, often there is a point to doing something for yourself that others could just as well do for you, in something like the way that the cake is a more special gift if you bake it yourself. We sometimes think it's good for children to learn how to do something themselves, and in something like this spirit, Kant took the only significant obstacle to thinking for yourself to be a childish timidity—that is, a moral and ethical failing—and his own prescription for training adults into the “courage” and “resolution” to think for themselves was the freedom to address the reading public, which he called “the public use of reason.” One side of his idea has become very familiar: in the so-called marketplace of ideas, the best of them will win out. The inspiring flip side has been neglected; as one's ideas are tested in the marketplace, that is, as they undergo public intellectual scrutiny, one will overcome that timidity and attain intellectual autonomy, and not just *intellectual* autonomy: freedom of expression, of a certain carefully characterized sort, will put you on the road to moral adulthood.

But, regardless of such benefits, you only have the latitude to insist on autonomy when you can endorse two complementary assumptions. The first, of course, is that you *can* think things through for yourself. The second is that you *have* to do it yourself, because we do not live in the procedural utopia. After all, if there were such a transparent test for correctness, if there were a calculation that would *just* get it right, insisting on doing it yourself would be allowable and even praiseworthy on occasion (when doing it your own way is on a par with making the gift by hand, all by yourself), but precious, self-indulgent, and irresponsible when anything important was at stake. Imagine your fully grown child complaining to you that, while she understands that you meant well, you had messed up her life irretrievably by insisting on

---

<sup>3</sup>To be sure, when I say that correctness is independent of human mental activity, I don't mean that cognition cannot be *about* human mental activity. Likewise, in those decision procedures, the inputs are in large part facts about preferences, or hedonic responses, or, more generally, facts about the psychologies of agents. For in-period worries about the program, see Neurath (1959).

doing it your way, rather than just following the instructions in the manual. If there *were* such a manual, that sort of complaint would be perfectly in order.<sup>4</sup> Kant's position was a reasonable one to adopt only because—with a handful of exceptions such as deductive logic—there simply is no way to determine what theoretical or practical conclusion to draw, other than to have someone examine the considerations in play and to figure it out.

Psychologism in philosophy of logic holds that the subject matter of logic is the operations of the mind: logic is concerned with what counts as correctly performing a particular mental activity, namely, reasoning; it is not in the first place about what abstract objects (such as propositions) stand in what entailment relations. The contrary view, almost universally endorsed today, is that the correctness of a conclusion is independent of human mental activity.<sup>5</sup> Autonomy, intellectual or otherwise, is a requirement when an investigator or deliberator who has lined up all the considerations he can, and examined whatever entailment relations he can make out, inevitably faces the moment depicted in a famous Sidney Harris cartoon: where, in the middle of a blackboard, a step has been drawn in and labeled “Then a miracle occurs.”<sup>6</sup> That is, after all of the preparation, he must *make up his mind* what he believes or is going to do, and, in almost all cases that matter, what he has managed to line up cannot itself *tell* him what that will be.<sup>7</sup> The Enlightenment makes sense only if psychologism is the right approach in philosophy of logic: if the best we can do is to assess how well the mental activity was performed, because part of the process being assessed is that miracle moment. Deference to the individual's autonomy, both practical and intellectual, is a substitute for a full

---

<sup>4</sup>That said, the second assumption isn't a component of the views of all Enlightenment figures. Jeremy Bentham, for instance, thought that there was a way to specify the correctness of a choice that was independent, in the relevant sense, of agents' deliberations. His own argumentative practice did not, however, reflect his theoretical views (Millgram, forthcoming), and I expect inevitably so.

<sup>5</sup>Antipsychologism was one of the McCarthyisms of the previous century's academia, and philosophers are brought up with the article of faith that psychologism is (or, rather, was) an unforgivable doctrine; few of them are aware of the nature of the debate from which that condemnation emerged. But see Kusch (1995), and Ringer (1990, pp. 295–298) for the social role of accusations of psychologism in, respectively, German philosophy and German intellectual culture more generally.

<sup>6</sup>Reproduced in Dennett (1991, p. 38).

<sup>7</sup>Kant has been interpreted as laying down a procedure for determining what to do—the so-called CI-procedure. (See Millgram 2005a, pp. 90–95, 141–142, for an overview.) But, first, the alleged procedure operates on maxims that it does not itself generate—intentions that an agent has in some other way to come to see as live options—and the CI-procedure does not work well unless the range of maxims on which it operates is already constrained. Second, one of the steps of the CI-procedure is constructing a “perturbed social world,” in which one's maxim guides the actions of all agents; however, there is no algorithm for generating such a construction, which requires ingenuity and insight in arbitrarily large doses. Finally, one of the steps of the CI-procedure is having the agent consider whether he can “will” a given perturbed social world. When imperfect duties are at issue (i.e., in most cases), the agent must simply *decide* whether he can so will: whether the perturbed social world is supportive *enough* of his agential capacities. In short, the dubiously named CI-procedure involves at least three then-a-miracle-occurs moments. It is not a procedure in the technical sense, that is, an algorithm.

understanding (the sort of understanding you could recast as an algorithm) of how effectively to solve theoretical and practical problems.

There are still neo-Carnapian confirmation theorists, and there are still crude Benthamite utilitarians, but the procedural utopia has never been more than a philosophical fantasy. The Enlightenment was right: we *don't* have a mechanical procedure for figuring out what to believe and what to do; we *don't* have an independently applicable criterion that determines what the correct theoretical and practical conclusions are; we neither know how we in fact do it, nor how to do it right. Psychologism *is* the right approach in philosophy of logic. The Enlightenment's implicit view was that, if we put our minds to it and allowed ourselves such aids as a healthy dose of unfettered criticism, we wouldn't do that bad a job of drawing our conclusions.<sup>8</sup> And, as I will shortly explain, for a little while that turned out to be true enough. While it is important to acknowledge the many disappointments, not least in the political arena, nonetheless, astonishingly, individuals were able to take up the responsibilities imposed on them by Enlightenment ideology—until the Enlightenment itself changed the human species form. Let me pause to explain what I mean by that.

## 2.2

Imagine a space alien to have visited our planet, any time during the course of human history prior to the last three hundred years or so, and to have written up field notes on humanity in the logically distinctive register of the natural historian.<sup>9</sup> The alien would have noticed that human beings, like a number of other terrestrial species, manage to occupy several distinct ecological niches and that they exhibit specialized adaptations to these niches. Some humans, he would have observed, feed themselves by tending animals (in something like the way that some ants tend aphids, he might note); others harvest plants; still others prey on fish; others confine themselves to parenting and alloparenting; yet others do none of these things, but occupy a handful of exotic, hard-for-an-alien-to-describe ecological roles in which they sometimes prey on, sometimes exchange products or services with, the food-producing humans. (We would call these specialized adaptations blacksmiths, priests, warriors, and the like.) Although caste memberships of these kinds might initially have seemed hereditary to the alien, at some point he would have been in a position to record that, perhaps modulo sexual specialization, just about any human, if placed from birth in the correct social location, is able to specialize into any one of

---

<sup>8</sup> An example of this cast of mind can be found in Mill's *On Liberty*, in the seventh paragraph of its second chapter (Mill, 1967–1989, XVIII:231f).

<sup>9</sup> For an elegant characterization of that register, see Thompson (2008, Part I); for a short overview, see (Millgram, 2009e).

the limited number of roles available. However, these creatures would have seemed to the alien to exhibit something very much like imprinting; exposed to that initial social environment, and having become a farmer, or a fisherman, or a wet nurse, they do not subsequently respecialize, and come later to inhabit a different ecological niche.

Now imagine that the alien has returned for a follow-up study, just about now. He would note that the human beings' adaptive strategy has become startlingly different. Because from a naturalist's point of view, a species is to be characterized in terms of its implementation of an ecological strategy, similarities in appearance notwithstanding, they might as well be a different species.<sup>10</sup>

The new species-wide strategy still involves specialization to ecological niches, but where before these niches were fairly simply defined, and quite limited in number, now they are much more highly articulated, and there are many more—apparently indefinitely many more—of them: some of the names the alien might list, taken from the creatures' own vocabulary, would perhaps be: website designer, pastry cook, tactile interpreter, network administrator, registered lobbyist, roastmaster, sommelier, oncologist, radiologist, philosopher of biology . . . And whereas, earlier on, the phenomenon of a human being's moving from one specialized role to another was rare enough to be neglected in a fully adequate natural history, in the meantime it has become common enough to count as a trait without which the form of life cannot be successfully described. People don't do it all that often (and sometimes get through life without doing it at all), but they retrain, and occasionally switch careers as many as five times.

I will return in a moment to the characterization of the new ecological strategy and, in particular, to aspects of it that the alien, standing outside of it, is unlikely to get right. In the meantime, let me give a name to the later strategy. The human beings that the alien observes on his return visit today are *serial hyperspecializers*.<sup>11</sup>

---

<sup>10</sup>So why might the alien naturalist identify them as the same species after all? Presumably, looks and lineage. The former isn't a principled ground, and the latter is only temporarily so. Species concepts form a Ravetzian lattice (Ravetz, 1979, pp. 193ff); different versions of the concept will be appropriately used by different specialists for different purposes. But let's take Mayr's version of the species concept: a species is a reproductively closed population. That's a good concept to use when you're dealing with sexually reproducing organisms, because the gene frequencies in one generation allow you to explain and predict gene frequencies in subsequent generations, and because genotypes explain phenotypes. We have already started to see technological interventions that rupture those links: as these accumulate, in human beings, gene frequencies in one generation will no longer explain gene frequencies in later generations, and genotypes will very shortly no longer explain phenotypes. The Mayr species concept will soon cease to be a useful intellectual tool for thinking about human beings. And as we can expect this to happen to one species concept after another, we need to contemplate the possibility that human beings will, in short order, no longer by any reasonable lights make up a biological species.

<sup>11</sup>For further pieces of the story, see Chapters 3, 4, and 9.

We who know the history of this transformation from the inside should occasionally remind ourselves what it looks like from the alien's perspective. And we should also remind ourselves how it happened. It was the exercise of public reason, so important to Kant, which produced the bodies of knowledge and technique built into these hyperarticulated specializations and which promises to produce indefinitely many more of them: indeed, by the mid-nineteenth century, William Whewell needed to invent a word for the new classes of specialists who were most responsible for this proliferation; he called them "scientists."<sup>12</sup> The phenomenon of people thinking for themselves has dramatically changed what, from the point of view of a natural historian, would in an Aristotelian vocabulary be called the human species form. If you like, there has been an unnoticed extinction event: humanity has vanished, we are their replacement, and that drastic ecological reconfiguration was the work of the Enlightenment.

## 2.3

As it turns out, when you grow up and start thinking for yourself, you don't merely leave older baggage behind; you start finding things out; that's especially true when you think out loud, addressing the public as, in Kant's phrasing, "a man of learning." And as it further turns out, when it's not just you, but many people finding things out, the way human beings generally live in their social and natural environment changes. Back before the Enlightenment kicked off, a reasonably intelligent, reasonably diligent person could know his way around pretty much everything there was to know, enough to form thoughtful judgments on the basis of that knowledge, and be sufficiently at home in it to make his own contributions, more or less across the board. (We call some of those people "Renaissance men.") But as the pool of information grew—and, more importantly, as the repertoire of skills needed to access and deploy different parts of that pool grew as well—the shared body of knowledge and technique was inevitably divided up into much narrower domains, consigned to the expertise of specialists. Specialists in such a society do not merely memorize information that other specialists or generalists do not. First, they have to develop and master proprietary systems of representation, in order to control their part of the pool of information; the upshot is that, increasingly typically, no one who is not the same sort of specialist can so much as *understand* the information in that part of the pool. Second, specialists internalize standards and guidelines that govern both their thinking about matters of fact and their choices of what to do; these standards and guidelines are also unintelligible to nonspecialists and, consequently, different

---

<sup>12</sup>Whewell (1847, vol. ii, p. 560).

sorts of specialist have startlingly different priorities and concerns.<sup>13</sup> A slightly exotic example that will turn out to matter presently: specialists of different sorts are adept at—and have internalized standards for—very different forms of argumentation; if you are not the relevant sort of specialist, you cannot tell whether a given argument of the kind they trade in is any good.

Consequently, autonomy is rarely a live option for serial hyperspecializers. To explain how and why, I will step through a series of progressively deeper obstacles to thinking for yourself.

First and most obviously, a serial hyperspecializer is in the business of relying on what he is told by serial hyperspecializers in other specializations; he relies on them both when he must take action outside of his own area of specialization, and when his own specialization invokes outside expertise. Part of thinking for yourself is rethinking the assumptions you are relying on; you are not equipped to rethink your assumptions when they are the deliverances of expertise not your own; thus, much of the time, serial hyperspecializers are not equipped to think for themselves.

Perhaps just a little less obviously, the problem is compounded by a common technique for passing information over disciplinary barriers, namely, dumbing it down. Normally, within one of these highly specialized fields, there is a great deal of nuance worked into both representations of fact and action-guiding assessments; accordingly, extensive specialized training on the part of users of those representations and assessments is almost always presupposed. For instance, within a field, a model may be understood to give usable results only sometimes, results which must be intelligently interpreted. Since outsiders cannot make head or tail of the hedges and qualifications, and cannot be expected to supply the intelligent interpretation on their own, when information and assessments are exported from a specialized discipline, they have the nuance, hedging, and qualifications stripped out. This problem turns up one step earlier than the inability of a would-be autonomous agent to rethink assumptions; quite frequently, for the very reasons he is unable to, those assumptions are already low-quality surrogates for the premises that would be needed as the basis for responsible and intelligent deliberation.<sup>14</sup>

---

<sup>13</sup>For a version of the first two of these points, see Fisch and Benbaji (2011, p. 7).

<sup>14</sup>For instance, financial institutions use VaR (value-at-risk) limits to manage near-term exposure to risk. (If the two-week 1% VaR is, say, one million dollars, that means that the chance of losing more than that in the designated time frame is less than one percent.) VaRs compress a great deal of modeling into a single figure; the model is built around exogenous variables that affect the institution's economic value and therefore might generate losses, and there are in addition various methods of computing VaRs: on the basis, *inter alia*, of historical data or via Monte Carlo simulations. These models are constructed by professionals with a strong mathematics background, but the VaR measurements serve as decision making tools for business managers—thus, they are used by people with insufficient mathematical training to understand the intricacies of the model.

Preemptively extending the illustration to cover points I'm going to make in Section 2.5, the financial institution's internal client is unlikely to know what scenarios have been taken into account,

Thus an agent aspiring to think for himself normally starts off not with what he needs to know, but with simplified—often ruthlessly simplified—substitutes. Even if he somehow looked up what he needs to know, he would be unable to understand it: a specialized system of representation can only be mastered through lengthy and intensive training. Even if he in some superficial sense understood it, he would not know what to do with it: the ability to competently deploy information generated by and for specialists depends on acquiring the relevant standards and guidelines, and that too normally presupposes lengthy and extensive training. So such an agent is not in a position to rethink his own views responsibly, which is to say that he is not in a position to think for himself.

## 2.4

That first lap of argument was straightforward, but it's not too early to get a couple of objections out of the way. First of all, a champion of the idea that we are pretty much alright as we are might allow that perhaps we cannot rethink the pronouncements of other specialists for ourselves, but reply that most of real life is lived within a shared public space, that that public space is not the domain of experts, and that autonomy matters most when it has to do with decisions precisely in the public space.

I don't myself know that activity within the shared public space is more important than any other part of life. It no doubt gets more media play, but that is because the shared space is a lowest common denominator of our cultural market; Hollywood makes romantic comedies rather than movies about civil engineering, because everyone is expected to experience romance, but very few people are engineers. However, supporting the industrial production of a class of cultural artifacts in this manner doesn't make an activity more important.

In any case, that objection misunderstands what serial hyperspecialization has inevitably made of the public arena in which we interact with our fellow

---

nor understand the type of calculation used. As the VaR result is a single number, as each financial institution has a large number of exposures, and as many exposures can generate losses under differing scenarios, the business manager will in many cases be unaware of the variables that influenced the result. Moreover, the process of constructing and validating such a model is time-consuming and detail-oriented; normally, even the people who produced the model will not remember many of the assumptions incorporated in it, short of redoing their work, which means that the client cannot simply ask them what went into it. Finally for now, many financial institutions utilize software from external IT providers, which further distances the end-user decision maker from the technical staff that created the model. To anticipate, this sort of ignorance can matter a good deal; for example, if the historical data used as input in the model differ systematically from present data, someone using an historical VaR might significantly under- or overestimate his actual risk, leading to incorrect business decisions. (I'm grateful to Jeremiah Millgram for talking me through the example.)



citizens. Because the machinery that supports even the most generic interactions is now managed by teams of specialists, serial hyperspecializers find themselves confronted, not with a public arena whose workings they understand and can reason intelligently about, but with a user interface.<sup>15</sup> When what serial hyperspecializers count as the generic environment they share has come to take its structure from the many specializations that impinge on it, its inhabitants will not understand how that public space works—even when they are well trained in its user interface. An example: everyone knows how to pay for goods and services; very few people understand what sort of information traces are left by their bank card, what uses are made of them, and how a pattern of purchases can subtly affect the shape of one's life. Another example: helping those less fortunate than yourself is, in our current user interface, typically a matter of making a donation to a charitable organization; as it turns out, you have to be intimately familiar with the workings of NGOs to have any idea whether your contribution is likely to make things better or worse—and sometimes, a further layer of anthropological expertise is needed to figure those workings and their upshots out.<sup>16</sup> A third example: everyone knows how to vote, and the issues and allegiances on which they express their opinions are vigorously debated within the shared public space. But only specialists understand how voting works in our political system; only specialists understand how that public debate is managed. And of course no matter how vivid the voters' opinions are, generally their correctness turns on specialized knowledge that almost no voters will possess or understand.<sup>17</sup>

Next, it is likely to be objected that I am overstating the requirements on autonomy, intellectual and otherwise. After all, one must in any case trust the testimony and practical assessments of others.<sup>18</sup> Thinking for yourself, the objection continues, requires that you can *rely* on your premises, not that you can rethink them. And surely the point of having experts is that we can rely on them.

I am about to turn to the mistake embedded in the objection: that if you can believe what an expert tells you, that should be enough. However, in the service of showing how pervasive the difficulty I am pressing is, the first appendix to this chapter will explain why identifying experts on whom you can rely is

---

<sup>15</sup>Our public environment recapitulates part of the recent history of computing. Not that long ago, operating systems presented users with interfaces that put on display what was taking place computationally. The Lisp Machine, a workstation developed at MIT during the 1980s, was an extreme instance: its built-in editor provided a keystroke command that allowed any user to see the definition of any function in his environment (Stallman et al., 1984). Today, users interact with computers by pointing at and clicking on icons—pictures—that are systematically misleading about the underlying workings of the environment and the programs running in it.

<sup>16</sup>Mosse (2005).

<sup>17</sup>For an introduction to issues pertaining to voter knowledgeability and voters' openness to argument and other forms of persuasion, see Converse (1964), and Zaller (1992).

<sup>18</sup>As I have argued myself; see Millgram (1997, Chapter 7).

harder than it looks. Surely, thinking for yourself involves at least being able to take responsibility for the choice of experts on whom you rely. But to ascertain who the right experts are, you have to be that sort of expert yourself.<sup>19</sup>

## 2.5

The point of helping yourself to the pronouncements of specialists is to deploy them in your own arguments; anyway, that is the point if we are trying to live up to the Enlightenment ideal of thinking for yourself. Moreover, it is normal to have to integrate the deliverances of different categories of authoritative specialists into a single train of thought, in order to support or dispose of some conclusion which you are considering. To see why this sort of deployment and integration is a nontrivial challenge, let's take as the central case the one in which the specialists' deliverances are supported by arguments.

Now, any argument that is not deductive is *defeasible*; that is, even if the argument is reasonable on the face of it, and would defaultly go through, it can be aborted by any of indefinitely many defeating conditions. Here is a sample entailment which you might use to build nondeductive arguments: because academics compete to place their work in highly ranked journals, the more selective the journal in which it is placed, the better the work. Although this generalization can be used to construct arguments for assessing the quality of academic output, those arguments will not necessarily go through—for instance, if the supply of journal reviewers has been exhausted and isn't being replenished, or if the reviewers in a small subspecialty are able to recognize each other and play tit-for-tat strategies in an iterated prisoners' dilemma. (In fact, these conditions obtain most of the time: an argument that goes through, *ceteris paribus*, is not to be confused with an argument whose conclusion is *probably* true.) As is standardly the case when an argument is defeasible, there are—obviously—indefinitely many other circumstances in which arguments constructed using this entailment would fail.<sup>20</sup>

Doing a decent job of thinking for yourself requires sensitivity to the defeating conditions of the arguments supporting your views, and if you are not an expert in the subject matter of those arguments, you do not generally control those defeating conditions. Continuing the example, an academic outsider, who is assured that so-and-so publishes in highly ranked journals in his field, who draws the conclusion that so-and-so's work must be of high quality, and

---

<sup>19</sup>A variation on this problem: when you have to make a choice as to which of two different leads to pursue, where these draw on different areas of expertise, your investigation confronts a resource allocation issue. How do you tell which one merits an investment of scarce resources? Only an expert, someone who is up to speed in both, is equipped to make the judgment call.

<sup>20</sup>For a more leisurely introduction to defeasibility, see Section 6.2; for tit-for-tat strategies, see Axelrod (1984).

who then relies on that work himself just doesn't know enough to have the right sort of second thoughts—although every now and again an Alan Sokol will remind people that those second thoughts ought to be pretty pressing. In cases where considerations drawn from different areas of expertise are being integrated into a single train of thought, thinking for yourself requires a still more demanding version of that sensitivity, to whether the reasons supporting one consideration are compatible with the reasons supporting another consideration. There is no reason to expect a serial hyperspecializer to be able to exhibit this sort of sensitivity to the mutual incoherence of his premises; they are, after all, adopted from diverse specialists whose trains of thought he cannot follow.

So the difficulty faced by the Enlightenment ideal goes deeper than the fact that serial hyperspecializers are not normally able to assess the quality of the expert descriptions, assessments, and, for that matter, direct instructions which they have no alternative but to consume, and it goes deeper than the fact that they cannot rethink them on their own. When you are assembling a defeasible argument, you should be confident in your conclusions only to the extent that you control the argument's defeasibility conditions; in a less fancy way of saying it, your confidence should not outrun your ability to catch problems and bugs that crop up in your argument. On the one hand, you do not control the defeasibility conditions of arguments that are conducted in other areas of expertise and that supply you with premises for arguments in your own area of expertise. On the other, you cannot expect experts in other areas to be sensitive to defeaters to their arguments that crop up when they are applied in your own area of expertise. Thus the level of confidence that serial hyperspecializers should have in the conclusions of even moderately demanding arguments is low. When you try thinking for yourself, and you have to deploy the conclusions of specialists in other disciplines, you can't trust what you come up with.

The problems of cross-disciplinary defeasibility management are amplified and deepened by a further obstacle. As I have already observed, induction into one of the many specializations characteristically involves adopting its proprietary modes of argumentation. Now, there is a dramatic philosophical term of art for someone who has a different logic than you do: he is a *logical alien*. The notion is usually introduced on the way to an argument that there could be no such thing: logical aliens, the received wisdom has it, are inconceivable, and the very notion is incoherent.<sup>21</sup> But once we have serial hyperspecializers clearly in view, because specialists' standards for correct argumentation are not shared, we should be prepared to find that we live in a society of people

---

<sup>21</sup>For example, Quine (1966, p. 102), or Davidson (1984, p. 137). Lear (1990, pp. 190ff) is one of the more interesting rebuttals. Wittgenstein (1983, I-143, 148–152), at any rate on its standard reading, is also meant as such an argument.

who are logical aliens with respect to one another. And that is what we *do* find.<sup>22</sup> When philosophers work at imagining logical aliens, they tend to tell stories about visits from outer space, or faraway primitive tribesmen. But if you are an academic employed by a university, and you want to meet a logical alien, you don't need to walk any further than the other end of the hall—or at most, to an adjacent building on your very own campus.

The psychologism implicit in the Enlightenment's demand that you think for yourself has as an upshot that you can judge whether an argument is any good only when you can, as we now say, wrap your mind around it. Perhaps you don't need to come up with that judgment for yourself. But when you rely on someone, in a manner that we understand to preserve your autonomy, you are delegating the investigation of your question to him, and you accept his conclusion because the way he went on to conduct the investigation was pretty much the way you would have done it—say, if you had had the time, and were where you needed to be to take a look. That is, you accept the results because the investigation is conducted correctly by your own lights.<sup>23</sup>

However, when you delegate part of your deliberation to a logical alien—put less dramatically, to a specialist whose idea of what a good argument is differs from yours—what he comes up with will not normally conform to standards you accept. When an outsider is aware of another discipline's internal standards, he may well and is even likely to think they are wrong-headed. (Not sure what I mean? Try explaining a transcendental argument to

---

<sup>22</sup>This isn't the place for a catalog, but let me indicate the sort of thing I have in mind: Statistical argument has a home in social science, and the recent attempt, in so-called experimental philosophy, to appropriate social science methodology within philosophy is a close-to-home reminder that what philosophers have regarded as properly an argument isn't what social scientists are trained to construct.

Again, but still staying close to home: The calculus of probabilities is central to various sciences; Bayesian epistemologists take probabilistic inference as their research topic. I have never seen those philosophers' own inferences, in their own publications on their chosen research topic, take a Bayesian form.

A further sampling: Millgram (2005a, Chapter 8) contrasts the demands placed on argument by philosophers, historians and politicians; Millgram (2013, p. 394 and n. 11) contrasts the demands placed on argument by philosophers and practitioners of literary studies; Auerbach (2003, pp. 547f), in a meditative aside toward the end of what is one of the acknowledged classics of twentieth-century literary criticism, suggests that the argumentative technique appropriate to his task mirrors that of the modern realistic novel. Rabin (1976, 1980) famously proposed using proofs that establish only that their conclusions are probable, and thus made vivid a clash between the standards for argumentation internalized by mathematicians, and those emerging in computer science during the late twentieth-century. For tensions between the standards of mathematicians and those of theoretical physicists, see Jaffe and Quinn (1993, especially at p. 5). Oreskes and Conway (2010, pp. 269ff) describe the way the different professional standards of scientists and journalists complicate the transfer of information from producers to clients.

<sup>23</sup>Thus, Raz (1990, p. 129) remarks: "There is likely to be ready agreement that experts of all varieties are to give advice based on the very same reasons which should sway ordinary people who wish to form their minds independently. The experts' advantage is in their easy access to the evidence and in their better ability to grasp its significance."

a developmental psychologist or a chemist.) Because you're not delegating to someone who thinks as you do, we no longer have an explanation for how delegation of this sort preserves your autonomy—or, perhaps more carefully, we haven't yet elaborated a notion of autonomy that makes room for such an explanation.<sup>24</sup>

Our third version of the threat to autonomy, then, is that as far as any one sort of serial hyperspecializer is concerned, many other serial hyperspecializers are logical aliens. Any serial hyperspecializer is, willy-nilly, the epistemic and practical client of serial hyperspecializers of other types. As we now understand thinking for yourself, when you accept a premise from someone else, you allow it to be reasonably and intelligently arrived at only if you would endorse the reasoning that produced it—that is, if that is the way you would have thought through the question yourself. But because the arguments deployed by experts in other specializations don't conform to standards of argumentation that you have internalized and that make sense to you, you are not in a position to make your willingness to adopt them contingent on this sort of endorsement.

## 2.6

In my exposition to this point, I've been adopting the perspective of an individual who is forced to rely on others. It's now time to take a step back and remind ourselves that the members of a society of serial hyperspecializers are symmetrically dependent on one another.

The Enlightenment insistence on thinking for yourself was tied up with the concern that those who did not would have their thinking done for them by other people who *were* thinking for themselves; Jeremy Bentham, typically, thought that you would be taken advantage of by “sinister interests.” Back in the day, someone else could have, in principle, done just as good a job of figuring things out as you would have yourself, and the job you would have done might well have been perfectly adequate. However, the problems we have been identifying are problems for *anybody*. So that, nowadays, when you let someone else do your thinking for you, while there is still coordination of interests to worry about, the main worry ought to be that, no matter how sincerely the person you are delegating your thinking to has your interests at heart, he is no better equipped than you are to perform the task that you are unable to

---

<sup>24</sup> As both Hilary Putnam and Samuel Fleischacker have reminded me, Kant follows up the demand that one think for oneself with two others: that when one thinks, one put oneself in the position of just anybody, and that one think in accord with oneself. (See Fleischacker 2013, p. 22, for references.) The presence of and the need to rely on logical aliens makes it as hard to see how these latter requirements can be met as it does the first of them.

adequately execute yourself.<sup>25</sup> It is not just that, after a couple of centuries of *Aufklärung*, thinking for yourself, when it comes to just about anything that really matters, is no longer an option: that no matter how resolute or courageous you are, you have to let others do most of your thinking for you. Rather, the most pressing problem is that, when you do, it's the blind leading the blind.

Let's stick with the view from above, and continue to represent the very general pattern of dependencies in terms of arguments. In a society of serial hyperspecializers, an individual starts out most moderately serious trains of thought by accepting premises from specialists in other disciplines; he conducts the train of thought, typically, within his own area of expertise; while he may be the end user, ordinarily his conclusion will be conveyed to a client who is himself a specialist in yet a further discipline. The boundaries between one specialist's stretch of argument and another's are merely artefacts of the intellectual division of labor: in the view from above, all these stretches of argument make up *one* (likely very long) argument. So the question we are really considering is how it is possible to monitor the quality of lengthy and complex arguments that are, so to speak, draped over indefinitely many areas of expertise.

My own very blunt assessment is that while various improvised expedients are locally available, we don't have any general and well-understood way of managing such arguments. However, I'm aware that this take on where we are is likely to appear too hasty to the adherents of one important intellectual tradition, and in the second appendix to this chapter I will explain why I think otherwise; this will serve to make a formal point I would like to have in view.

## 2.7

I began by suggesting that the Enlightenment is being undone, and now that we have a clearer rendering of the dynamics of the process, we had better ask ourselves just how urgent a concern it really is. Maybe we should just leave behind the moralistic demand that each of us grow up and think for himself. After all, it's my experience, and probably yours also, that when you ask philosophers to give you an *argument* for being autonomous, you walk away empty-handed—even when it's Kantian moral philosophers, the ones you'd think would have those arguments if anyone did. Maybe there are actually no good reasons to regret the passing of the Enlightenment, in which case the tone of world-historical drama with which I began this chapter was no more

---

<sup>25</sup>Those conflict of interest problems, which I'm putting to one side, are however themselves infected by specialization: as when malpractice insurance gives the physician incentives to overtreat you; pharma junkets give him an incentive to overprescribe; financial interests tied to his practice give him an incentive to overdiagnose.

than theatrics. With this challenge in mind, I want to take a little time out to make the costs of the Great Endarkenment clear, and to briefly gesture at symptoms that we have started paying those costs.

Let's first recap the problem as it has emerged up to this point. There is the initial difficulty with living up to the demands of the Enlightenment: that you can't tell for yourself whether your views about what the world is like or your decisions about what to do make sense, because, first, you can't rethink the assumptions of your trains of thought for yourself; the merits of your position are opaque to you—but, it might seem, perhaps not to an expert on whom you are relying.

Second, once you accept that you have to rely on authorities, you face the problem of successfully integrating premises supplied by expert authorities into your own arguments. That requires sensitivity to potential defeaters that, lacking the relevant expertise, you cannot reasonably be expected to have.<sup>26</sup> And third, the assertions and advice that you appropriate from other specialists to serve as premises for your own arguments are supported by trains of thought that by your own lights are not properly put together.

When you trust the social structures of the Endarkenment to produce socially reasonable outcomes, and to enable intelligent decision making on the part of individuals, there is no theory or model that shows why it should work (and when it won't).<sup>27</sup> On the contrary, we have a pretty good *prima facie* argument that it would be a miraculous accident if it *did* work. When we appreciate the symmetries in the ways that members of a society of serial hyperspecializers are positioned with respect to one another, the overall difficulty is not just that the merits of any position we adopt are opaque to us: they will be equally opaque to *anyone*. We are, all of us, playing a variant of the children's game of telephone, in which, at each stage, a player receives a handful of premises that he is not equipped to properly understand, reassess, or reason with; his move is to deploy those premises to produce a conclusion in his own area of expertise; he then dumbs it down, stripping out the hedges, qualifications, and so on needed to make intelligent use of it, and makes it available to a different player (one who is not equipped to properly understand, reassess, and reason with it), for that next player's turn. The outcome is a series of commitments (both decisions and views as to how the facts stand) whose merits no one is in a position to assess—where that is true recursively: each decision, whose merits no one can assess, is made on the basis of further decisions, whose merits no one can assess either. Trusting in the outputs of *this*

---

<sup>26</sup>When you have to rely on different kinds of authority at once, you face the problem of integrating inputs from those very different experts: for instance, of determining whether their inputs can be deployed jointly, and, when they conflict, which is of higher quality.

<sup>27</sup>By way of a contrast, we have a clear explanation of why markets are supposed to coordinate economic activities successfully—anyway, when they aren't being tasked with the problems of serial hyperspecialization. (I discuss that latter case in Appendix B.)

process is on a par with settling what you are going to do by reading entrails or casting hexagrams. It is blind, terrified superstition.

In the course of getting through the day, or the year, we have to make decisions; those decisions, ranging from life-and-death choices on down, have consequences for the way lives go; so how our lives go, individually and collectively, depends on the quality of those decisions. Here as elsewhere, quality must be constantly wrested from the jaws of entropy, and we should expect that unless we are continually assessing the merits of our decisions, they will be almost all no good. But we have gradually put ourselves in a position where no one is equipped to assess the merits of *most* decisions. So we have put ourselves in the position of making decisions, most of which we can expect to be no good. Leaving to one side the irony that all that hyperspecialization was, in retrospect, an inevitable side effect of the Enlightenment itself, we're in a pretty bad fix. The Great Endarkenment is the onset of a period during which individual lives will go badly, and the societies they constitute will suffer through yet another of those dreary eras of misery that seem to fill the history books.

You might be wondering whether I am overstating the costs. How could our circumstances be as bad as all that? Life goes on. Our cars are not *always* falling apart on the road; airplanes do not *usually* fall out of the sky; skyscrapers *mostly* stay up; all of these artifacts require the successful cooperation of highly specialized experts of many different kinds.<sup>28</sup> So successful cooperation must after all be possible.

Life does indeed go on—except, as the credits soundtrack for a popular videogame reminds us, for the ones who are dead.<sup>29</sup> The record shows that there can be far too many of the latter without anyone so much as noticing, and let me remind you of a recently documented case. In the world of pre-modern medicine, life went on, just as it does now. In retrospect, patients were being tortured and slaughtered, in the ghastliest ways imaginable, for no good reason at all. But no one could tell, because they weren't equipped to assess the theories, inferential practice, and effectiveness of the procedures performed by members of a specialized professional guild.<sup>30</sup> So patients continued to turn to physicians for expert medical assistance, even though, until about the turn of the twentieth century, by far the best medical choice anyone could make was to have nothing to do with doctors. If we find ourselves once again in the

---

<sup>28</sup>For an attempt to marshal statistics showing that life doesn't just go on, but has been getting better and better, see Lomborg (2001, *passim*, with recaps at pp. 328, 348). The volume as a whole itself constitutes a very instructive illustration of the point of the present essay. The author adduces evidence from a wide range of sources in order to show that environmentalist initiatives are almost always misconceived. Since environmentalism covers a very wide range of issues, the evidence inevitably draws on many different types of expertise. The upshot is that the book's reader will almost always be incompetent to assess both the quality of evidence itself and the use that is made of it in the argument in which it is brought to bear.

<sup>29</sup>Coulton (2011), which concludes the first version of *Portal*.

<sup>30</sup>See Wootton (2007) for a respectable overview.



predicament of being unable to manage quality control for theory and practice in a great many disciplines, I would not put a great deal of stock in the impression that life goes on nonetheless—and, as I have been arguing, that is, to an ever greater extent, our predicament.

If we do not see it that way, that is no doubt for two reasons. First, we are dealing with an *emerging* problem. The track record of the Enlightenment has been a pattern of successes—skyscrapers that stay up, ever-better mortality statistics, and so on—that has only recently started to be subverted by hyper-specialization. But if my argument is on target, the pattern will more and more become one of failure; we should not be lulled into complacency by the legacy of better times. And second, we exhibit bias toward the visible.<sup>31</sup> I have been arguing that we're in trouble because there are problems that are in principle hard to notice. Because we consequently don't, for the most part, notice them, we can't count them; thus, we are not in a position to measure the severity of our situation.

Nevertheless, every now and again, one or another of these problems does come to light and is brought to the attention of the general public. Because illustrations can go some way toward showing how pressing the failures we are considering can be, I will remind you of a couple of recent and widely publicized instances.

Internal combustion engines are a remarkably effective method of powering vehicles and, more generally, burning fossil fuels is a way to meet the energy needs of an industrial society. Some of the especially dramatic defeaters of the arguments behind those conclusions are only properly understood by climate scientists.<sup>32</sup> It's notable that one of the central obstacles to formulating and implementing a policy response to the problem of anthropogenic climate change is that outsiders are unable to understand and assess the science for themselves.<sup>33</sup>

There has been a secondary obstacle to producing a coherent climate policy that is less discussed, but just as important. A minority cluster of positions on this topic amount to agreeing that human activity is raising the average temperature of the planet, but demurring from the view that this is an urgent matter; the background assumption at work in such a posture is that, and I'll put this in an appropriately down-home register, *I* know what's important for

---

<sup>31</sup>For a classic statement of the notion, see Bastiat (1995).

<sup>32</sup>Clark et al. (2013); as it happens, it is possible to package a simplified explanation of the basic mechanism that drives global warming (Ranney, 2013), but if I am right, even as much as that is no more than a fortunate accident. (Lomborg 2001, Chapter 24, attempts to provide defeaters for the arguments supporting the dramatic defeaters.)

<sup>33</sup>Oreskes and Conway (2010, Chapter 6) emphasize the role of political motivation; however, it is clear on their own account that the strategies adopted for delaying a policy response to climate change depend on the inability of the general public to discriminate specialized expertise from a thin simulacrum thereof, which in turn is only an important factor because both the public and policy makers must rely on what experts tell them. (See also Chapter 4, footnote 32, below.)

*me*. But that assumption, we can already observe, is a mistaken holdover from a less demanding era; in the world of hyperspecialization, we have to live with not just intellectual division of labor, but *evaluative* division of labor. We have to outsource the production of not just the information but assessments to experts, and we—the philosophers and those down-home folks—are unaware that we have to do it this way, and in any case don't know how to manage outsourced assessment.<sup>34</sup>

And our second instance: the so-called Great Recession was in large part due to the failure of financial institutions to price collateralized debt obligations (CDOs) correctly. It had been shown that in principle these financial instruments could be fairly priced, and it was concluded that sophisticated institutional players would do so successfully. The arguments for that conclusion are defeated—it was belatedly realized—by considerations from computational complexity theory: correctly pricing CDOs is computationally intractable.<sup>35</sup> Who would have thought, ahead of time, that financial professionals and economists needed the input of specialists in that particular subdiscipline of computer science?

Once again, we only know about the bugs that we've caught, and the argument tells us that we can't count the ones we've missed.<sup>36</sup> I don't have anything like a full account of our present ways of coping with serial hyperspecialization. But I can point to one mechanism that I am guessing plays a substantial role in keeping the wheels on those cars. I'll quickly describe it, adapting an amusing discussion by Bruno Latour. Then I'll give a real-world example; lastly, I'll say why, if this is how we do it, we are indeed living out the undoing of the Enlightenment.

Latour reminds us of a story on which previous generations were brought up, about how it is that science can put men on the moon.<sup>37</sup> During the Dark Ages, physicians, natural philosophers, and so on believed the pronouncements of dead authorities over the evidence of their own eyes; science, the story continues, is now able to put men on the moon because we pay attention to the evidence of our eyes and disallow arguments from authority. Latour argues that that is just a bedtime story. What makes science *work*, he tries to show, *is* arguments from authority. Scientific progress requires results to be stable (enough); the relevant unwritten rule within the science

---

<sup>34</sup>I'll take up the philosophers' version of the error, sometimes called "internalism," in the second appendix to this chapter, in Chapter 3 and in Section 6.7.

<sup>35</sup>Arora et al. (2011).

<sup>36</sup>And notice that in both of these illustrations, neither of the defeating conditions that proved to matter lay within specializations that were on the paths of the defeated arguments. The arguments for using internal combustion engines drew on a great deal of specialized knowledge and skill, but not on climate science; the arguments for slicing and dicing mortgages into CDOs didn't bring computer scientists to the table.

<sup>37</sup>Latour (1987).

and engineering community is roughly that, in order to challenge someone's result, you have to build a bigger laboratory than he had; so results stay authoritative long enough, on average, for the enterprise to function.

This much in the story seems right: when clients in one hyperspecialized discipline take over a result, device, or technique from another, even when they are not equipped to properly understand it or deploy it in inference, sometimes they will find that they can get away with using it: they may not know why, but it doesn't get them into trouble, and so it becomes a staple of the client field. Stability matters because we for the most part don't know *why* some results or techniques work in our field and others don't. It was once explained to me why brake assemblies for trucks are not manufactured in China—a fact that amounts to a puzzle, because it would after all be much cheaper. The firms that assemble the vehicles themselves don't, as it turns out, understand brake assemblies, and so they don't know why any particular supplier produces reliable brakes. They do know that when a truck doesn't stop, it's a bad thing; so they're not going to try to save a few dollars by switching suppliers. The brake manufacturers in turn don't understand the components *they* use, or what makes them reliable; once again, they won't switch suppliers . . . and this stickiness in the supply chain stretches all the way down to a particular gasket factory. A gasket is just a piece of rubber, and you would think it could be made anywhere at all; nonetheless, one gasket manufacturer is especially reliable; their clients don't understand why, but they're not going to try to save a few cents by switching suppliers.<sup>38</sup>

In the Dark Ages, the peasants did as their parents had done; on those rare occasions when they walked to the next village over, they made sure not to stray off that path through the woods; who knew what would happen if you did? That's *us*, when we assemble trucks and when we do a great deal else. Like medieval peasants, we find rituals that we can repeat, and stick with them in something like the way that people walk so as not to step on cracks in the sidewalk. So far, we have somehow gotten by; but to the extent that we are doing so on brute trial-and-error and conservatism, it is hard to believe that we would not be much, much more successful if we actually understood why some of our cross-disciplinary borrowings worked and others did not.

Say that freedom is, *inter alia*, the ability to do what you decide, and say that liberty is centrally a matter of protecting freedom from governments and other authorities. Surely freedom is valuable primarily because the ability to do as you decide is valuable (although I want to acknowledge the many secondary ways in which it also matters, such as avoiding the humiliation of having obnoxious bosses tell you what to do). But the ability to do as you decide to

---

<sup>38</sup>I'm grateful to Maneesh Modi for the background. Notice the role of modularization in this quality control strategy: if the truck doesn't stop, you can be pretty sure it was the brakes, and so on down the component tree.

is valuable in the first place because you can think for yourself, and that when you do, having enough confidence in your own conclusions to act on them is not simply unreasonable. If it *were* unreasonable, your best option would be to hope for a Jeeves to whom to entrust the management of a life as lived by Bertram Wooster.<sup>39</sup>

The Great Endarkenment thus directly undercuts our commitment to freedom and to liberty. And it undercuts those commitments indirectly as well. Eternal vigilance, we are told, is the price of liberty; vigilance involves thinking things through for yourself, and if you can't, the question of what your liberty is *worth* is likely to be moot. It is small consolation that those to whom we lose our liberty are bound to be about as incompetent as we are ourselves. Keeping our freedom, keeping our liberty, and keeping the both of them worth hanging onto mean figuring out what it could be to take realistic and meaningful responsibility for the conclusions of our theoretical and practical arguments—for our assessments, opinions, and choices—within a society of serial hyperspecializers. That is to say, it means arresting and reversing the Great Endarkenment.

## 2.8

What would that *take*? Many efforts on many different fronts will be necessary, and we had better be thinking them out from scratch; it's clear from the get-go that older ways of addressing the challenges posed by older forms of specialization—for instance, education in the classics—are not going fix the problems we have surveyed.<sup>40</sup> We do need to inventory and analyze the devices that we are currently using to cope, however clumsily, with hyperspecialization as we now have it; but we should not imagine that such a survey can be more than preliminary ground clearing.<sup>41</sup> An intelligent response to a challenge of this order of magnitude begins with understanding and successfully characterizing it. Only once we understand the problem will we be able to devise and verify solutions to it.

---

<sup>39</sup> As, for instance, in Wodehouse (1984); for a darker depiction of misguided autonomy, see Laxness (1997).

<sup>40</sup> A classically oriented education was not—as it is likely to appear to most people nowadays—merely silly. As MacIntyre (2009, p. 142) observes, “the small-scale . . . physical size of [Greco-Roman] culture, its relatively self-contained character, made it a suitable object of study. It was far enough away to be viewed dispassionately, near enough to provide a model for social behavior.” Still, if an economist is having a hard time understanding why work in computational complexity renders some of his own results unusable, ensuring that both the computer scientist and the economist share a dead culture as a common reference point will not repair *that* cognitive failure.

<sup>41</sup> For treatments of a handful of devices used in cross-disciplinary interface management, see Section 4.7 and Chapters 5, 6, 7, and 8.

Recall the slant that I put on the Enlightenment early on in this chapter, on which the difficulties that are well along to undoing its achievements are problems in philosophy of logic. Logic, conceived in the psychologistic manner demanded by the Enlightenment itself, is the theory of reasoning done right (where I want to allow that this sort of theory may amount to a system of directives); philosophy of logic, then, will be the discussion and theorizing meant to eventuate in and frame such a theory. (For instance, the dispute at which we gestured earlier, as to whether logic should be construed psychologically or antipsychologically, is squarely within the ambit of philosophy of logic.) Recall that the first of the problems on our list comes under the heading of cross-disciplinary interface management: when an argument traverses multiple specializations, how can information and guidance be moved over those disciplinary barriers? The second was the question of how to manage the defeasibility conditions of an argument, when a stretch of the argument and the intellectual control of its defeaters belong to different areas of expertise. Here we were considering especially the problem of how the different specialists already mobilized when articulating such an argument can cross-check its various stretches for defeaters; but there is also the broader problem, of how to bring to bear necessary expertise in areas that did not lie on the pathway of the argument.

The third problem has to do with our need to accept the expert pronouncements of specialists that are arrived at on the basis of procedures and in conformity with standards that we either do not understand, or that conflict with the standards and procedures we ourselves endorse. Facing up to it will require a change of approach in philosophy of logic, over and above the turn back to psychologism.

The variations in sensibility across the natural sciences—especially and in particular with respect to what can count as a good argument—are both commonplace and awkward enough to have given rise to a genre of joke.<sup>42</sup> (“How do a mathematician, a physicist and an engineer”—and here’s one typical continuation—“prove that all odd numbers are prime?”) Disciplinary specialization looks like relativism come home; but what we need to understand is how, if the way physicists solve problems looks to an engineer like a *joke*, and vice versa, the both of them are willing to use each other’s solutions as premises in their own arguments. That willingness is of course modulated; sometimes a specialist (or a layperson) will correctly accept expert pronouncements produced to standards that he does not endorse, and sometimes not; what makes the difference, and why? Philosophers of logic have almost always been concerned to delineate what reasoning done correctly is—that is, the *one* thing that counts as reasoning done correctly—and have not been figuring

---

<sup>42</sup>For a less light-hearted expression of such a clash of sensibilities, see Krantz (1989).

out how to make use of the results of reasoning that does not live up to the standards of argumentation that we accept ourselves.

If the problems at the bottom of the Great Endarkenment are problems in the philosophy of logic, then they are, first of all, problems for philosophers. Here I'm going to convey my own view on this point tersely, and through a series of bald assertions.

All cultures have wisdom literatures (or oral traditions that occupy that role); these often take up questions that we regard as typically philosophical: for instance, how a good life is to be lived. The ancient Hebrews left us *Job* and *Ecclesiastes*; we have produced, less impressively, the books lining the shelves of the self-help section of your local bookstore. Philosophy began when one individual, whom we still remember, made a nuisance of himself by asking those with opinions about such matters what their argument was. Thus argumentation has been central to philosophy from its very beginnings, and it does seem to me that there is an Aristotelian way of saying just how it figures into philosophical views and positions, namely, as its material cause: just as clay is the matter out of which ceramics are made, so argument is the material out of which philosophy is crafted. (I should say precisely here that because I have not yet argued for that claim, it remains only potentially philosophy—the form or design of a combination of philosophical ideas.) Indeed, if you look at any philosophical tradition, you will find that it has a view of what it is to conduct argument correctly. Typically, that view is distinctive, and a great deal of the shape of the tradition is to be explained by looking to its view of argumentation. Thus I am often tempted to say that the remainder of philosophy stands to logic, or to philosophy of logic, as applied ethics traditionally conceived stands to straight moral theory; all the rest of philosophy is applied philosophy of logic.<sup>43</sup>

Because of the role of argument within the craft of philosophy, philosophers are almost always explicitly trained in their methods of argumentation to an extent that is at least out of the ordinary in other fields. Moreover, the theory of rationality comes up again and again as a topic of explicit discussion in all walks of philosophy. This is a heritage we can bring to bear on the management of methods of argumentation that vary from specialized discipline to specialized discipline, but which nonetheless must be integrated with one another.

---

<sup>43</sup>If philosophy does not always seem this way to its historians and practitioners, there are various reasons. In our own doggedly unselfaware and antipsychologistic tradition, we have almost forgotten how to argue about how to reason. And because philosophers trained in the mode of argumentation endorsed by their own tradition generally know only it, they are often unprepared to recognize argumentation conducted in a different mode, by a different tradition, as argument at all. (Just for instance, one reason that analytic philosophers are by and large inept readers of pragmatist texts is that they simply do not recognize the pragmatists' arguments as *arguments*.)

Philosophy of logic is first philosophy. The answers we need to stave off the Great Endarkenment have to do with how serial hyperspecializers can conduct reasoning that traverses multiple areas of specialization; the discipline we must turn to for those answers is philosophy of logic. Philosophy of logic is, and as I have suggested moments ago, not merely as a verbal or classificatory matter, philosophy. And so reversing the Great Endarkenment is the job, in the first place, of philosophers.

And in the end, if the challenges of the Great Endarkenment are met, and the series of problems I have listed are solved, it will be by philosophers, for a further reason. Philosophers have to solve these problems, because anyone who does solve them will, whether or not he has been trained by the academic discipline that now bears the name, *thereby* count as a philosopher.

## 2.9 Appendix A

If you are to be thinking for yourself, then if you have to rely on authorities, you must be able to manage some form of quality control; in the first place, you have to identify and vet the authorities on whom you rely—to have an intelligent view as to whether they are the *right* authorities.<sup>44</sup> But if you are not yourself a \_\_\_\_\_, how do you know which \_\_\_\_\_ to believe, and how far?

It's hard to get a sense of how deep the problem goes without considering an example one understands well; because such examples span different areas of expertise, and the problem I hope to exhibit has to do with one's inability to acquire many different types of expertise, accessible examples are hard to come by. As a substitute for this most direct sort of illustration, consider a variant of the problem faced by academic administrators. Academia is familiar to the primary audience of this book, and it is in the relevant respects a microcosm of society at large; its administrative structure explicitly reflects the way in which the once shared body of knowledge and technique has been subdivided into narrow areas of expertise. Showing how deans, university vice-presidents, and so on attempt to cope with the challenges to which this sort of division of labor gives rise will provide a clearer sense of what the difficulty is, and why it is so hard to address.

Serial hyperspecializers depend on the intellectual products and practical instructions generated by other highly specialized experts. But autonomous deliberation requires assessing the quality of the expertise on which the deliberator depends. In an academic setting, one (slightly ivory-tower) form the

---

<sup>44</sup> A representative expression of this view, in fact produced in the course of a discussion of the Enlightenment (Fleischacker, 2013, p. 21): "I can't (responsibly) *delegate* responsibility until I first *take* responsibility for the decisions by which I figure out what to delegate and to whom. I can't *decide* to trust authorities, as opposed to trusting them blindly, until I first figure out what, for me, will count as a good reason to trust someone."

problem takes is an academic administrator's need to assess the performance of departments within a university. For instance, a dean needs to allocate limited resources among the different departments, and if he is going to do this intelligently, he needs to decide whether a department is doing better or worse than a competing department; perhaps he needs to decide whether the department is functioning well, and can be left to govern itself, or whether an intervention on his part is necessary. The form the problem of assessment takes, in the first instance, is that the administrator is not himself a poet, or a physicist, or a philosopher, and he cannot tell whether what his faculty are doing is any good—or, if it is any good, just how good it is. Or rather, it is that, even if he was trained as one or another of these, he is unable to assess work in the other disciplines in which he was not trained. Each discipline has its own generally quite nuanced modes of assessing the activities of its members, and in order to apply the standards of \_\_ ologists, you have to be a \_\_ ologist yourself.

Traditionally, university administrators turned to assessments on the part of experts both within and without the university. But doing so just reproduced the evaluative task: Which experts do you rely on? If you are not sure how good the Department of \_\_ ology is—after all, that's what you're trying to assess—where do you get off trusting its professional judgments? Moreover, appealing to the very experts one is trying to assess obviously creates a conflict of interest. However, when a dean turns to external referees, he finds that these referees write reports that only convey real information to readers already trained in the standards of the discipline: you have to be a physicist yourself to understand what the external report on a faculty's tenure case is actually telling you; you have to be a philosopher yourself to understand what the outside letters for a potential hire are telling you . . . and no administrator could be a physicist *and* a philosopher *and* all the rest. And the dean still does not know *which* external experts to turn to; his primary source of expert advice on this matter is the very department whose judgment he does not know he can trust.<sup>45</sup> Skepticism is live and well in the offices of your upper administration.

Members of an academic specialty cannot be expected to assess work in fields not their own, and so they cannot usefully compare their own work to work in other fields. But academic administrators need a way of comparing the performance of academic units with different areas of expertise. In the past, desperate administrators substituted rankings across institutions of similar academic units for the direct comparisons of different types of academic

---

<sup>45</sup>In addition, there is the further concern that, in a small enough field, the external assessment can be easily captured by insiders: specialists outside that dean's university are likely to be beholden to their colleagues within the university in a way that they are not to administrators such as himself. But I am putting obstacles that depend on participants' bad faith to one side for the present.



unit within their own institution—the comparison they really need. (That is, what they wanted to know was whether the budget should go to the English Department or the French Department; but what they ended up looking at was the standing of their English Department in a nationwide comparison to other English Departments.) And they turned for those rankings to third parties: government agencies, self-appointed members of various disciplines, sometimes even popular magazines (think *US News and World Report*). But doing so is no better than an exercise in self-deception unless there is reason to think the person or agency generating the rankings has solved the problem one was unable to solve oneself; and just as a dean cannot be a physicist and a philosopher and a PhD in all those other fields too, neither can a journalist or civil servant. To exhibit how the problem plays out, I'm going to help myself to a new-fangled version of the desperate move; at my home institution, the administration has purchased a subscription to a commercially developed software product meant to assess the research dimension of the performance of academic units within a university.

As it turns out, the company marketing this product has exactly the same problem that a dean has. Each discipline has a different way of measuring performance, and to understand it, you have to have taken out a decade or so of your life as an apprentice, in order to internalize the practices and standards of the discipline. No programmer could have done that, and so the software package ends up looking, to anyone within any particular academic discipline, wildly off-base.<sup>46</sup> That is at any rate true of its current incarnation, and while the firm has been responsive to suggestions and pushback, the software doesn't come anywhere near to producing an assessment of faculty productivity that reflects the internal disciplinary standards of those faculty. It is both remarkable and informative that the failure of its benchmarking options to track disciplinary standards came to the managers of the firm (and I can vouch for this personally) as a *surprise*.

It is easy to leap to the conclusion that we are seeing a problem with one particular startup: perhaps they don't know their way around the business they are in; perhaps they are insufficiently motivated to get it right; in any case, another firm would do the job better and incorporate domain experts from the get-go. And in that case, there would be no philosophical lessons in the

---

<sup>46</sup>Since my primary audience is philosophers, I'll use the philosophy-oriented aspect of the illustration: as of the time of writing, the software tallies up the number of journal articles and books published by members of a department in a given year, but it does not weight articles for the ranking or selectivity of the journal (very important, in philosophy); it does not weight book publications by the ranking of the press (somewhat important, in philosophy); it does not tally up chapters published in edited volumes at all (in philosophy, a very important publication venue); and I could go on in this vein. (Due in part to pressure from yours truly, these are now ongoing projects.)

However, don't assume that the problem is that they merely happened to overlook the philosophers, who are not after all the center of the universe. One more very terse example: because law reviews are student-edited rather than peer-reviewed, the vendor initially failed to harvest law review publications for its database, even though the product was intended for and marketed to law schools.

offing. But I expect that the difficulties this firm has encountered are not just an accident. Notice that the software developer trying to hire experts has the same problem your university's vice president for research has: he cannot be all of the specialists himself, and so he cannot tell which experts to rely on for assessments of quality; he cannot interpret the assessments himself; he has no magic solution to a problem that is starting to look intrinsically hard.

And notice further that even if the developer had attempted to build in domain expertise from the outset, it would not have provided what its clients in a university administration really need. When a dean requests a departmental self-assessment, he is in the first place trying to figure out how a department is doing compared to other departments covering different fields. (Again, your dean is never faced with the choice of giving the budget to *your* English department, or a *different* English department elsewhere; he needs to decide whether to give the money to your English department or your chemistry department.) Tailoring the performance indicators for each specialization to the internal standards of the respective disciplines would simply make such comparisons harder. In any case, the developer cannot expect to hear about the need to incorporate domain expertise from his own clients: these are administrators, and preponderantly staff in offices of institutional research at various universities. The latter in particular have scarcely any contact with their faculty, and are only very rarely academics themselves. Thus they do not themselves have a sense of those discipline-internal standards; moreover, getting their own job done requires somehow bypassing all that variation in those standards. Accordingly, what looks at a cynical first glance like incompetence, or perhaps choosing to market a defective product, is best construed as the intelligent decision not to try to do a job that no one understands how to do, and to focus instead on a task whose parameters are clear.

Again, you might be thinking: how does an administrator come to sign the purchase order for such a product? But if the administrator could discern the failings of the product himself, he wouldn't need to purchase it in the first place. The problem he is trying to solve is that of assessing the performance of specialists, and the reason it is a problem is that no one who is not that type of specialist is able to do such assessment himself. The most straightforward way to determine that the product is not doing what it purports is to notice that it is not generating the correct assessments; but to do that, one would have to be able to assess the performance of academic units in different disciplines oneself. When people are faced with problems they don't know how to solve, they do things that don't address their problems; however, in such cases, their responses are an indicator of the difficulty of those problems and not of anyone's personal failings.<sup>47</sup>

---

<sup>47</sup>Thus I am complicating the diagnosis of Ginsberg (2011), who observes that within American universities, the ranks and the power of academic administrators have grown dramatically in recent decades and that administrators tend to make decisions not on the basis of the professors' priorities,

Attempts to solve the problem of assessing the work of experts, when one is not oneself that kind of expert, almost inevitably reproduce the problem rather than solve it. We saw that when you hire a consultant to tell you whether the expert is any good, you get the new and equally intractable problem of assessing the consultant; the consultant gets the so-far-unsolved problem of assessing expertise across different fields; and while the problem gets moved around, it does not get solved. In fact, if we stop and think about it, this is not by any means the first time the bump in the carpet has gotten moved. University administrators are not themselves directly consuming the products of the expertise they are trying to assess; the dean who wants to know whether his chemistry department is doing high-quality research is not about to use the chemistry faculty's publications in his own research. Rather, the university has gatekeeping as one of its central institutional functions; it exists in part to certify to further inexpert consumers of the expertise—such as students who propose to embark on apprenticeships that will make chemists of them, or employers who will hire those chemistry majors upon graduation—that the relevant information and skills are of high quality. But this means that the problem which the academic administrators are today outsourcing to a private contractor is precisely the problem their own institution existed to solve for others in the first place—the problem that, long ago, was outsourced to the universities. The cautionary tale reminds us of something we should already know: that when a problem is not well understood, merely designating an institution as responsible for solving it is not enough to get it solved.

---

but their own—which he puts down to a mix of conflicts of interest, comic incompetence, and sheer corruption. While I am not dismissing the principal-agent problem here, underlying it is a deeper issue: even with the best will and all the good faith in the world, administrators are not in a position to *understand* the academics' priorities and concerns; they are specialists themselves, and they act on the basis of the niche-specific system of standards, guidelines, and priorities that are constitutive of their own professional specialization.

That said, we should not underestimate the very real threat to academia that the current movement toward data-driven benchmarking presents. The uses to which these tools are almost inevitably put amount to *neo-Taylorism*. You will recall that Frederick Winslow Taylor, the father of the most memorable management fad of the turn of the twentieth century, enthusiastically promoted time-motion studies as a way of imposing redesigned and more efficient workflows on employees. Now, the case studies that he announced as such dramatic successes were uniformly manual and menial labor: for instance, hauling coal onto railroad cars, or bricklaying. The common denominator of these tasks was being so very simple that the question of whether the time-motion expert understood the process he was trying to optimize, or what it was supposed to accomplish, clearly did not arise (Taylor, 1997).

However, in the world of serial hyperspecialization, the updated versions of time-motion studies are used to redesign a process that the managers do not understand and whose outputs they don't know how to assess. Today's equivalents of time-motion studies are themselves performed using the sort of tools we have just been discussing; these are likewise designed and implemented by specialists who do not understand the processes the tools are supposed to measure, and who are likewise themselves unable to assess the outputs those tools are supposed to assess. It is hard to believe that delicate craft practices and skill sets can long survive this sort of management.

## 2.10 Appendix B

The most important reason for markets, according to Friedrich Hayek and his followers, is that the information necessary to make decisions about resource allocation is not all available to any single decision maker. Markets successfully exploit dispersed information; they amount to a distributed information processing and collective decision-making technique that produces highly coordinated activities across a society. While the results of these activities are not necessarily optimal, they are observed to work out reasonably well, and in any case are better than the alternative, which in this tradition is taken to be central planning.<sup>48</sup>

All of that is true enough, but while the famous invisible hand of the market can be expected, in many circumstances, to produce socially satisfactory outcomes, in those circumstances we have a very clear explanation of why those satisfactory outcomes are in the offing, and of why the prices that individuals see enable them to make economically prudent choices. The reality check here is that our models of markets exhibit the limitations of market mechanisms: so that not only do we understand why markets work when they do, we can anticipate the conditions under which they will fail. Thus, when you trust in the market to produce—often enough, nearly enough—products that satisfy consumer demand, that is more than unwarranted optimism.

However, those models do not explain how the problems we have found serial hyperspecializers to face are solved by markets. On the contrary, our models suggest—and inspection shows—that the apparently intractable difficulties that we presented in the vocabulary of argumentation reappear when we consider the sorts of transactions in which serial hyperspecializers will have to engage, now clothed in the alternative idiom of buying and selling. Markets work well when buyers are in a position to assess the products they acquire from sellers—that is, when a purchaser knows what he wants, and can tell whether he got it. In an economy of serial hyperspecializers, that is not ordinarily the case. The relevant standards of assessment are all-too-often not available to the purchaser; the purchaser is unable to identify, or so much as describe, (the analog of) defeating conditions for his purchase. And frequently enough, a given purchaser is not the final point of consumption; in the world of serial hyperspecializers, a purchase is typically intended for incorporation in a product to be sold to a differently specialized client, who will use it in ways that no actor in his supply chain properly understands.<sup>49</sup>

---

<sup>48</sup>See, for example, Hayek (1948), or Kirzner (1963, Chapter 3).

<sup>49</sup>When we examine work in the Austrian tradition, we find that the problems we are considering have not come in for careful attention. Confining ourselves to the works I mentioned in the previous footnote, Hayek (1948, p. 97) holds that “in actual life . . . our inadequate knowledge of the available commodities or services is made up for by our experience with the persons or firms supplying them . . . competition is in a large measure competition for reputation or good will”. This insightfully describes

When we look around our economy, we repeatedly encounter disastrously suboptimal solutions to economic problems.<sup>50</sup> Because the problems arise precisely because it's hard for any one person to notice them, illustrations are hard to present; nonetheless, here is one. For almost all of my adult life, I have watched computer scientists attempt to explain to outsiders why they should not be purchasing this or that product. There is no point in discussing particular instances, precisely because it would be as hard to explain the assessments to most readers as it was for the experts to explain them to the outsiders in the first place. However, it's important to register that the misguided choices were not merely occasional; on the contrary, they shaped almost all of the market in the personal and business computing segments. (Minicomputers

---

one solution to an earlier relative of the problem we face, but it assumes, first, that your supplier is better positioned than you are to assure the quality of the item he is supplying; thus it overlooks the symmetry considerations highlighted in Section 2.6: he is as dependent as you are on his own specialist suppliers, and often no one will be in a position to certify his product for its ultimate use. It further assumes that clients can reliably assess their specialized suppliers, as opposed to assessing the products or services directly. What empirical evidence we have suggests otherwise: patients, for instance, turn out to evaluate their health care providers on the basis of the upholstery in the waiting area and amount of eye contact they receive from their physician—neither of which are correlated with clinical outcomes.

Or again, Kirzner (1963, p. 34) casually remarks that “for an individual, the economic problem consists in ensuring that the resources at his disposal be utilized in the most effective manner possible—from the point of view of the goals which he has set up”; this is to take for granted an approach to practical rationality that has little to do with the problems that face serial hyperspecializers. (For additional explanation and argument, see Chapters 3 and 4.) Sticking with the medical domain, it is *typically* the case that patients neither understand the options with which they are faced, nor have set up goals that will support intelligent decision making with respect to those options.

The most direct way of making the point that consumers of specialized products are unable to tell whether they're getting what they want is unavailable: I would have to first explain the options a medical consumer does not understand, and for the same reasons the consumer does not understand his options or have usable preferences, the reader will not be able to follow the explanation—not without first completing medical school. But here is some indirect evidence.

A brief way of explaining what lies at the bottom of the ongoing attempts to restructure health care provision (not just in the United States, but in every country with which I am sufficiently acquainted) is that institutions are quickly forced to acknowledge the claim I just made: that the recipients of medical services are unable to understand the option space they are faced with, are unable to evaluate the outcomes before and even when they experience them, and simply cannot make decent choices on their own. One way or another, the choices have to be delegated to an agent other than the patient: a general practitioner, an insurance company, an HMO, a government agency, and so on. Now, markets work well when both the costs and benefits of a choice accrue to the agent who makes it; each of the agents on this list of replacements has its own incentive structure; and none of that is a secret. Once the collocation of costs and benefits has been surrendered, market failure is difficult to avoid, and sure enough, the alternative choice mechanisms we have provide perverse incentives to the actors involved. The shortcomings of deferring medical decisions to these substitute agents are evident, and policy makers would not turn to them unless they had no choice. So institutions in one country after another being forced to these expedients serves as the confirmation I am after.

<sup>50</sup>While markets are not alleged to find best solutions, and while the notion of a best solution is perhaps not even well-defined, nonetheless we must be allowed to appeal to disasters: Austrian arguments for markets turn on recognizing the outcomes of central planning as disasters.

and workstations tended to be the exceptions; in those cases, the purchasing decisions were made by professionals with the relevant expertise.)

Those explanations could be understood only by people with a fairly extensive background in computer science and were almost uniformly ignored; in particular and especially, the specialists were unable to convey both the substance of their concerns and their assessment of how weighty competing considerations were. Some of them—for instance, concerns about security—are belatedly starting to gain wider purchase; however, even security issues are still not fully appreciated by nonspecialists, and many other equally pressing issues (e.g., forward- and backward-compatibility) have no uptake at all. Thus the purchasers were guided by preferences, assessments, and evaluations that were inadequate to protect, much less advance, their own priorities, *even from their own point of view*. Here the market did produce a coordinated set of prices and outcomes; while it's hard to estimate the costs of converging on them, what matters just now is the source of the difficulty: when lay consumers, and even business professionals, make purchasing decisions about products whose costs and benefits they are unequipped to assess, markets cannot be expected to generate even reasonably satisfactory outcomes.

Here is perhaps a more familiar case. When part of what consumers are purchasing is competence with and control of standards of assessment—and this will normally be the case in apprenticeship to a professional specialization—they will be unable to identify potential defeaters for the decision ahead of time. Now, consumers choose well when they can decide what they want and what they're willing to pay for it *before* they purchase it. When a transaction is frequently repeated, initial experimentation can be treated as an up-front cost; the main run of transactions can nonetheless satisfy that timeliness-of-assessment requirement. But in the case of professional education, the purchase is typically large enough to be a one-time decision, and the overall results are available for survey by my primary audience: if you are a philosophy professor, you have clients who are unable to assess the quality of the product they are consuming.<sup>51</sup>

---

<sup>51</sup> We have already discussed the form the generic problem takes for professionals (in Appendix A); here, briefly, is how it plays out for the apprentices. They want to learn philosophy, but they don't know what philosophy is, or what it is to do philosophy well. They do have various misimpressions, and part of your job is to disabuse them of their confused preferences and objectives. Your students will control the standards of assessment only once they have come quite far along toward becoming philosophers themselves, and they will be equipped to assess their education only if the education itself was of high quality. Briefly, they will have already consumed the product by the time they are equipped to decide whether to go ahead and do so.

You will see your students turning to two backup expedients. One is incremental assessment (e.g., deciding whether to continue on the basis of how the student liked the last class he took). While there is nothing wrong with that in principle (see pp. 258–259), since the student doesn't control the relevant standards of assessment, he is likely to be relying on his affective responses, and when he is going by how he *liked* the class, he may well turn out to be using an irrelevant index. A recent internal study at my home institution analyzed student evaluations; these turn out not to measure responses along the

Here is a potential defeater for an argument whose conclusion would be the decision to get a philosophy education; my formulation will be perhaps provocative, but because the defeater is available only to someone who already controls the relevant standards of assessment, it makes the formal point I need. While philosophy can be a worthwhile pursuit, there is enormous pressure within the profession first to channel oneself into developing and defending narrow and intellectually frivolous positions and then to settle for arguments that are no better than “gotchas.” (Briefly, the pressure is toward academic activity that really does conform to a layman’s dismissive stereotype of what philosophers do.) The risk of becoming trapped in these activities is real, and if one is uncertain one will be able to resist the pressure, that is a perfectly good reason to abort the career choice. But one can’t properly understand the trap until one has the philosophical maturity to distinguish worthwhile from frivolous philosophizing, and that comes slowly, as one becomes a philosopher. (And as always, when we are considering defeasible inference, that potential defeater is a representative drawn from an indefinitely long list.)

We were illustrating our response to the Hayekian objection, namely, that the inability of a consumer to determine what he wants—and whether he is getting what he wants—is part of the shape that the inability of agents to detect out-of-specialization defeating conditions for an argument takes in a market. We can use it to make a more general and formal point, having to do with the conservation of philosophical work. Expressing a problem in a different idiom only rarely makes it go away. Here we have a problem we don’t know how to solve; clothing it in a different set of practices and procedures does not allow us to do what we didn’t know how to do before.<sup>52</sup> Throughout, I have been

---

many dimensions the questionnaires purport to probe, but rather to be tracking a single feature of a class, probably intuitively best described as “UX” (for “user experience,” the feature for which Apple products are famous).

The other is relying on others’ assessments: student evaluations, published rankings of departments, and so on. Choice on this basis is a wild gamble, in that the consumer can’t tell (i) whether the assessments are even minimally competent and (ii) when there is an idiosyncratic dimension to correct choice, whether the outside assessment tracks it (e.g., if what you need is philosophy done well in a way that suits your philosophical temperament, you have no idea whether the ranking you are using is reasonable for *you*).

The default is not trustworthiness to a naive user. In a market, competitors will try to promote their product by creating and publicizing assessments that recommend it. A naive consumer has no way of discriminating misleading from trustworthy assessment tools. And so when a would-be philosopher starts out by getting exposed to real philosophy, is trained to assess philosophy correctly, and so on—that is *sheer luck*. He has no control over it himself, and in a market, others cannot be counted on to do the job for him. On the gate of Plato’s academy, we are told, was inscribed an injunction to have learned geometry first; on the entrances to our philosophy programs, the appropriate inscription would be: *caveat emptor*.

<sup>52</sup>Is that entirely fair? Hayek (1973, p. 19) argues that our inherited practices and social arrangements are often effective, but in ways that we do not understand. However, his explanation for the

describing the problems we face in terms of the need to draw the conclusions of arguments, and that may have given the discussion an overintellectualized cast. But the problems remain even when the people who have to cope with them do not think of themselves as being in the business of assembling arguments and drawing conclusions.<sup>53</sup>

---

effectiveness is group selection: we were lucky enough in the arrangements we happened to adopt to be one of the surviving societies.

Whatever the results of such a process, we shouldn't expect them to work in novel circumstances. The extreme degree of specialization that we find in contemporary society is very recent; so we have no warranty that the expedients that by chance kept us alive up to this point will continue to work now. And in any case, group selection is a slow and brutal process; it would be much better to understand the problems we face, and preemptively work out solutions to them, than to wait haplessly to find out if we're going to be the lottery winners in the next round of eliminations.

<sup>53</sup>I'd like to thank Heather Douglas, Chris Maloney, Maneesh Modi, Adam Morton, David Schmidtz, and the staff of Academic Analytics, LLC, for helpful conversation. Carla Bagnoli, Teresa Blankmeyer Burke, Sam Fleischacker, Svantje Guinebert, Jamie Hardy, Buket Korkut-Raptis, Jeffrey Seidman, and Jim Tabery gave me comments on an earlier draft, and the Center for the Philosophy of Freedom at the University of Arizona provided support, work space, and valuable feedback over the course of a retreat.



## Practical Reasoning for Serial Hyperspecializers

The theory of practical rationality (that is, the rationality of decision, choice, action, and related matters) dates back at least to Plato and Aristotle. That makes it well over two millennia old, and over that time, we have, to all appearances, built up a substantial body of results: results that allow us to understand what we are doing, when we make up our minds what to do, and which also amount to guidelines that can help us do a better job of it. In my view, however, the body of theory we have developed in this area is unusable. It is, I am going to claim, not so much mistaken as irrelevant, because philosophers have made a very basic error, and an error that, I will suggest, has damaged much other work in moral philosophy. If I am right, at least in the theory of practical reasoning, we are going to have to start over, almost from the very beginning.

### 3.1

Aristotle famously asked, in his *Nicomachean Ethics*, what the human *ergon* was: what it is that people *do*. I am going to try to answer that question, and so I will argue that we ourselves are a solution to a design problem. But I will differ from Aristotle in my account of what the human *ergon* is.

The approach is unusual enough nowadays to require a cautionary remark. Because the current fashion is to explain or replace design-based explanations of, especially, living things with evolutionary explanations,<sup>1</sup> and to accept functional descriptions of an organism only if they can be construed as shorthand for claims about identifiable pressures for that function operating within the history that produced the organism, I am likely to be understood

---

<sup>1</sup>“Especially,” because the approach has been adapted to other areas as well (see, e.g., Boyd and Richerson 1985, Richerson and Boyd 2005).

as claiming that the design solution I will describe is an adaptation for which we have been selected. The biological world, and we ourselves as part of it, are products of natural selection (where the story of natural selection is understood to include the contributions of drift, mutation, etc.), and so of course I want my claim to be compatible with a plausible evolutionary history. But I do not mean to imply that we were selected to be the design solution I will identify. (In fact, I will argue in due course that this hypothesis is quite implausible.) My question is not, How did it happen? but, What is it good for? Compare: eyes solve the problem of seeing, whether or not that explains how they arose.

On the contrary, rather than assuming a sort of reductionism—of functionality to a history of natural selection—I think that the interesting question to ask about natural selection is what design problem *it* solves. The observation I will use to frame my discussion of humans is that natural selection (together with features of living organisms that control ways that taxa reshape themselves in response to selection pressures) is a solution to the problem of producing creatures suitable to available ecological niches.<sup>2</sup> How does it go about that?

Here are three large observations about evolution as it operates in the biological world. First, Darwinian selection produces species that fall along a spectrum, with the very weedy species at one end and the very specialized species at the other. Weedy species (remember: weeds are the plants that you can't keep out of your garden) are the ones that invade niche after ecological niche, and because they travel from one to the other, they're not necessarily particularly tailored to any of them.<sup>3</sup> Specialized species, on the other hand, are often fitted to their places in an ecology with a memorable and jewel-like precision. Think of the migrating bird, spectacularly engineered to exploit resources that are located some thousands of miles apart. When it touches down in the marshlands that are the southernmost terminus of its journey, it turns out to have long stalky legs that are perfectly suited for standing in the marsh, and the scoop-like beak that is perfectly suited for preying on the fish and frogs that it will find there. When it arrives at the northernmost end, it turns out to be perfectly fitted to nest in the crevices of the cliff which it will find

---

<sup>2</sup>Examples of the sort of features I have in mind are sexual reproduction and mechanisms that contribute to the reproductive isolation of a species; that is to say, that many organisms comprise species, in the sense of the term given by Mayr (1984, especially p. 539), is such a feature. Modularity may well be another, and perhaps plasticity itself (West-Eberhard, 2003, pp. 164, 178f; for references to work on the evolution of evolvability, see pp. 182f).

<sup>3</sup>The contrast between weedy and specialized species should not be confused with another familiar and embattled contrast, that between *r*- and *K*-selected species (Stearns, 2004, pp. 105f; for complaints, see pp. 206f). A reminder: not all variants of the niche concept allow for niches to be independent of the organisms that occupy them; to describe organisms as traveling from niche to niche presupposes that sort of independence. For an overview of niche concepts, see Odling-Smee et al. (2003, pp. 37ff); for a note of caution, see Arthur (1987, p. vii).

there.<sup>4</sup> Specialized species take advantage of ecological niches in ways that weedy species normally cannot, but they pay a price in fragility; too often, when the niche disappears, so does the species: as the forests of the Pacific Northwest are cut down, the spotted owl vanishes.

Second, in any species that relies on a nervous system to compute its behavioral outputs, the adaptations that we see are both in hardware and software. The bird's legs and beak are special-purpose hardware designed (or, since it's *natural* selection, "designed"—but I won't repeat this important qualification throughout) for the narrow range in which the species lives. But the bird's cognitive architecture is just as much a part of the engineering solution to the problem posed by the niche. When the seasonal cues trigger the overriding urge to fly north, that urge has an adaptive function: to get the animal from one location, where it has been using resources of one kind, to a second location, where it will avail itself of resources of a different kind. The program is, in cases like these, part of the package, one that is as invariant as the hardware—a fact we mark by calling the bird's impulse to migrate an "instinct."

Third, much of the time, natural selection operates on a use-it-once-and-throw-it-away model. At the level of individual organisms, the process churns out many (often vastly many) near-copies of a design solution, throws away the ones that don't work . . . and throws them *all* away after a lifespan.<sup>5</sup> But it is not just individual organisms that are disposable, and what is interesting to us now is a contrast we see as we move along the specialized-weedy spectrum. Proliferating throwaway species is how sexually reproducing life at the specialized end of the spectrum fills the available ecological spaces: a niche opens up, a specialized species evolves to fit it; the niche goes away, and the species is discarded.<sup>6</sup> On the weedy end of the spectrum, however, while the individual organisms are just as much disposable hardware as ever, planned obsolescence is no longer the most salient feature of the species as a whole. Weedy species are reusable, and a more economical alternative to Nature's all-too-common NASA-style practice of throwing out design solutions after a single use.

Now here's a neat idea that the as-if designer might have: to build a species that was *both* weedy and specialized, that was in fact weedy *by* being specialized. One way to work the trick would be to use less specialized hardware and put the capacity for specialization into the software.

---

<sup>4</sup>For gorgeously shot illustrations, see Debats et al. (2001); to be sure, sometimes the fit of a specialized species is quite awkward, and it survives anyway.

<sup>5</sup>Continuing the list of features started in footnote 2, sexual reproduction is a solution that requires death in order to work efficiently; if older copies of gene mixes stick around, the genetic profile of a population won't adapt as efficiently as if they all die off. Once again, this is an answer to the question, What problem does death solve? not, How did it happen? For recent approaches to the latter question, see, e.g., Stearns (2004, especially pp. 199–202).

<sup>6</sup>The bald way I have just put it might suggest that the niche appears first, which is of course not necessarily the case: often specialized species and their niches coevolve, as in the case of parasites and their hosts.

The weedy-but-specialized species would run different software in different niches, and the movie-perfect fit would be worked, not by the beaks and legs and so on, but by behaviors that changed from niche to niche. How would such a design work?

There is evidently more than one way to answer that question; I am going to focus on two alternatives, one of which is more ambitious than the other. Let's call the less ambitious solution *Piltdown Man*, for reasons I'll explain toward the end of the Chapter. Piltdown Men are born, identify the environment they're in, load a program appropriate to that environment from a database of available strategies, and run the program until they die. Piltdown Man can occupy what would intuitively look like many different locations in many different ecologies, and not the way that mints do: on the seashore, they might be fishermen; in the mountains, they might be yak herders; in the cities, they might be merchants; on the plains, they might be farmers. However, Piltdown Man can only occupy relatively stable niches, because those software libraries have to come from somewhere; and if we don't have a Kubrickian monolith hand-coding them, they will be produced by a process that is either natural selection or that (like meme selection) resembles natural selection in being, by human standards, slow. Since stable niches are a small fraction of all the niches, Piltdown Man will be distributed relatively sparsely through the strategy space.

Piltdown Man is already a departure from the throwaway approach so typically taken by natural selection. Piltdown Men are still individually disposable, in two senses: first of all, of course, they die; second, if the niche drops out from under them, the former Piltdown Man can't become a yak herder or merchant, because his behaviors are fixed by his initial download (to drag in a biological metaphor, by his initial imprinting), and he starves (or ends up in a refugee camp). But the *species* is reusable: when one niche evaporates, new Piltdown Men are produced to occupy other niches. The yak herders die off, but their children become farmers; there's no need to start over and develop a new species, with hardware that is painstakingly (at great cost, and over a long period of time) redesigned for the new environment. Piltdown Man is a much less wasteful, much more efficient approach than natural selection's default strategy.<sup>7</sup>

The more ambitious version of the neat idea pushes it further in a couple of directions. First, you could make the individual creatures reusable rather than disposable. After all, why throw out perfectly good hardware when the program it's running is no longer up-to-date? It's much more efficient to design

---

<sup>7</sup>There are halfway stages to Piltdown Man: think of dogs, which are shaped largely by artificial rather than natural selection. Here, the adaptations are still in hardware and hardwired programs (dachshunds, designed to go down burrows; boxers, designed for bear baiting; hunting dogs; tracking dogs; herding dogs), but the species as a whole is less disposable, because there are those relatively disposable breeds.

such a species to look for new and better software when it's cued that the old software is no longer getting decent results. Let's give this part of the design solution a name: creatures like this are *serial specializers*.<sup>8</sup> Second, instead of relying on a fixed database of programs, you could have the individual creature reprogram itself on the fly, to meet the demands of a new ecological niche. (That is, while these creatures will sometimes search for and adopt off-the-shelf software, when that is the cost-effective option, sometimes, and especially when suitable software is not already available, they will develop their own.) Doing it that way opens up the option of taking advantage of extremely transitory niches: VLSI engineer, comics inker, Cobra gunner, French professor specializing in eighteenth-century poetry, adventure travel agent, director of cinematography . . .

What's more, as the list I just began suggests, it opens up the option of more elaborate adaptations, to niches that turn out to be more exotic and delicately constructed than anything in nature. Suppose that Piltdown Man's software results from a process that, over long periods of time, shapes it to match a stable niche. Over long periods of time, niches will change at the edges, maybe quite rapidly; this means that programs which don't change as rapidly as all that need to be a little bit robust in the face of such fluctuations, that is to say, a little bit weedy: thus, not quite all the way to the specialized end of the weedy-specialized spectrum. (Piltdown Men will be farmers and merchants, but not the peculiar cattle breeders one runs into nowadays, who purchase prize cows at auctions and frozen bull semen by mail, artificially inseminate the cows, and FedEx frozen embryos to be reimplanted in cows elsewhere; and they will not be hedge fund brokers.) However, with on-the-fly reprogramming, that constraint is removed: the specialization can be as filigreed and fragile as the underlying computational resources allow. The improvement over Piltdown Man which we're now considering allows not just specialization but *hyperspecialization*. At this point, we've reached a design solution with which we're intimately familiar (although with important qualifications I'll get to in due course)—from first-person experience.

A moment ago, I compared Piltdown's Man's alternative source of software to the alien monolith in Kubrick's *2001: A Space Odyssey* (1968), and this gives us a way of saying why we shouldn't worry that, for all we know, we're really Piltdown Men. And it also gives us a way of saying why we wouldn't want to be. On the first count, without the ability to do on-the-fly self-reprogramming, the proliferation of specialized forms of life that we see in human societies could only be managed, given the brevity of history, through the intervention of more intelligent outside management. But there are no space aliens or Prometheuses bequeathing us the exotic skills of modern-day cattle breeding

---

<sup>8</sup>For a related set of classifications—generalists, specialists, and polyspecialists—see West-Eberhard (2003, p. 382).

or derivatives trading. On the second count, Piltdown Man seems like a peculiarly *dependent* species. The Norwich terrier is a breed of dog that gives birth by Caesarean section and whose existence is supported by ongoing human intervention; Piltdown Man, in a social world that is reasonably complex by our lights, would be a little like *that*.

If we aren't Piltdown Man, what are we? *We are serial hyperspecializers*. That's the right ecological description of the species *Homo sapiens* and our unusual implementation of the weedy-species strategy. William Tenn once wrote a novel suggesting that humans were a perfect fit for the sort of role occupied by rats and cockroaches,<sup>9</sup> and while people do seem to be able to fill this role when they have to (and historically they often *have* had to), what they do better (and what they do when they have the opportunity) is to mimic—and displace—biodiversity with cultural diversity. Humans change the ways they cope with and fit into their environments, again and again and again, and this has important consequences—consequences for how we should understand our own psychologies, for what rationality must be for creatures like us, and thus for the assessment of much previous philosophy.

### 3.2

Migrating birds come equipped with a set of desires (or anyway, that's how we might as well think of them for now<sup>10</sup>) that match the environment in which they're going to find themselves. The desires are triggered by cues in the environment, but such creatures are not equipped to think about, redesign, or resynthesize their desire set. Our imaginary Piltdown Man, who downloads one behavior-controlling software package over the course of his lifetime, works almost the same way, and the claim I want to put on the table now is that Piltdown Man represents the mainstream philosophical conception of the human being. A survey of the many philosophical views to which this accusation can be made to stick would be a survey of much of the history of philosophy, and of much contemporary analytic philosophy as well. In this Chapter, I am going to confine myself to debates about practical rationality (with one excursion touching on the closely intertwined discussions of substantive moral theory).

The standard model of practical reasoning (or figuring out what to do) is still instrumentalism, the view that your ends or goals or desires are just somehow given to you (hardwired in, or burned in by your upbringing, or prompted by cues in your environment, etc.), that figuring out what to do is

---

<sup>9</sup>Tenn (2001); for an animated film with a very similar view of humanity, see Laloux (1973).

<sup>10</sup>Sterelny (2003, Chapter 5) argues that calling items like these desires or preferences probably attributes too much modularization to the cognitive architecture, and while that's a qualification I'm happy to allow, it doesn't affect the present point.

figuring out how to achieve your ends, and that there's nothing that counts as thinking about what *to* want in the first place, or what your goals *should* be.<sup>11</sup> That is, the default philosophical conception of practical rationality conceives of people as scarcely one step away from creatures like migrating birds.

Instrumentalism is a good theory of practical rationality for Piltdown Man (recall: a creature distinguished from us in the first place by its inability to re-program itself for new environments).<sup>12</sup> But let's stop and ask how creatures designed as serial hyperspecializers *would* have to think about what to do. These creatures solve a design problem, that of occupying novel and ephemeral ecological niches, by producing behavior that is specialized to the niche. Since the niches are paradigmatically novel, the problem cannot be solved by using prestored guides to behavior (such as the urges, triggered by environmental cues, that we noticed in the migrating birds, or, in human beings, stable clusters of memes transmitted culturally). Rather, the behavior-guiding goals have to be computed on the fly. Now, that point is broader than I have just made it sound, so let me adjust it before moving on. An organism's controlling software need not rely exclusively on goals (or desires or ends or preferences); desires or goals are elements of just one of the available control systems, and by no means the most sophisticated of the lot. Accordingly, I will be supplementing talk of goals with more general notions, such as standards, priorities, and guidelines.

If the task has to be performed intelligently, and if we allow that intelligence requires thinking, then it has to be possible for such creatures to rethink their top-level goals, standards, priorities, or guidelines. (That is, in an older philosophical jargon, they have to be capable of deliberation of ends.) The puzzle then is: How can a strategy like that be implemented?<sup>13</sup>

I don't have a complete recipe, but I do have a couple of the ingredients on hand. If one is developing standards, priorities, etc., to govern activity within a niche, and if one wants the standards, etc., to be reasonable guides to action, one had better explore the niche first. A reasonable way of exploring a novel niche is to try things out and see how well they go; but of course in assessing how well they go one cannot invoke the standards one has not yet developed. So we should expect serial hyperspecializers to come with cognitive equipment that can be used for such assessment, and we should expect it to be relatively rough-and-ready (that is, not to presuppose too much about the structure of any niche in particular). One such tool might plausibly be a signal that tells the creature when it's doing better and when it's doing worse; armed with such a

---

<sup>11</sup>For a recent and sophisticated defense of a view in this family, see Vogler (2002).

<sup>12</sup>For supporting argument, see Chapter 4.

<sup>13</sup>Does this follow, once we allow that humans often have much less narrowly specified ends than migrating birds do? See Millgram (2008) for a discussion of what reasoning with and about such ends looks like, and see Millgram (1997, Section 6.7) for argument that it will involve the cognitive techniques I am about to review.

tool, a serial hyperspecializer exploring a novel niche will be able to sort the options it has tried out into an system of rankings—which perhaps can later be refined into a more elaborate system of standards.

And in fact we do find such a cognitive tool in human beings. The most generic form of pleasure—and of its contrary, displeasure<sup>14</sup>—responds to changes in one's well-being and represents, roughly, the first derivative of one's welfare (that is, it indicates the rate of change). If you feel good, things are getting (or have just gotten) better; but after a while, that feeling will fade. If you feel bad, things are getting (or have just gotten) worse; but after a while, modulo malfunction in the signaling system, you will adjust, and the feeling will fade. These signals are an important input into the deliberation of ends: Disappointment prompts people to adjust their ambitions downward, by giving up or scaling back their goals; the elation produced by success prompts them to adopt new and more ambitious goals. In other words, in human beings, goals or ends are adjusted and selected on the basis of the input provided by pleasure and displeasure.<sup>15</sup> Hedonic tone is cognitive equipment that looks to be part of a solution to a problem, that of developing appropriate standards for novel niches.

It is not the only such piece of equipment. When one is exploring a novel niche, one will do much better at it if one's attention is directed to the features that will be important for, among other things, developing a system of standards and priorities appropriate to the niche. So you would expect a serial hyperspecializer to come from the factory equipped with a suitable attention guiding device. Since niches differ from one another, the device must be quite flexible; but it must be able to pick out candidates for exploration with better-than-chance payoffs. And indeed, when you take a look at human beings, they are equipped with a signaling system that does this, namely, interest (and its complement, boredom).<sup>16</sup> Philosophy is a domain in which one audience of this paper will have seen this piece of equipment in action. When you train a philosophy student, you train him to have a sense of what is philosophically interesting; that is, you are training his philosophical nose. Astoundingly enough, it is possible to produce reliable noses (though very hard to produce them reliably), that is, responses of interest and boredom that can be used to guide, and *successfully*, choices between avenues of intellectual exploration. (Astoundingly, because it is not as though the problems of philosophy were the life-or-death demands made on our Pleistocene ancestors.)

---

<sup>14</sup>So-called to distinguish it from pain, the signal that one's body has been damaged (see Pitcher, 1970).

<sup>15</sup>This account of the cognitive role of pleasure and displeasure is developed in Millgram (2005a, Chapter 1); for reasons not to conflate such signals with the goal recommended by hedonists, see especially Section 3, and Millgram (1997, Section 6.4).

<sup>16</sup>This account of boredom and interest is suggested in Millgram (2004a).



On the instrumentalist way of seeing things, success or failure must be spelled out in terms of whether one is achieving already given goals or ends. So notice that on the view I am sketching, the direction of explanation (and of justification) is, for the most part, the other way around. When your hedonic responses tell you that you are doing well, or doing badly, those signals (although they can be theory-laden, and although their use requires intelligent interpretation, and although they can be corrupted, and although they are, once again, trainable) can pull free of the goals and desires you already have—and must do so, at least sometimes, if they are to do their job. When you are getting everything you want, and you still feel like you are failing, that is *telling* you something. Signals of success and failure tell you what your ends and values need to be, and not (modulo the qualifications I just rehearsed) the other way around.

Serial hyperspecializers will be equipped with cognitive devices enabling them to explore novel ecological niches and establish standards, priorities, and goals suitable for guiding behavior in them. I will eventually be in a position to raise the question of why philosophers have not produced a task analysis complete enough to tell us what the full functionality of those devices must be. But we can nonetheless say that we come with what are evidently components of the package built in. That's reassuring, as far as whether the present account is on track: a reality check on the story I'm telling is the ability to identify its elements in our own lives. (Because there is typically more than one way of solving a design problem, pointing out that we are solutions to such and such a design problem, and that so and so would solve the problem, is not enough to show that we are so and so; the reality check is indispensable.) But if we humans are serial hyperspecializers, and these are among the components, then part of practical reasoning, for us, is the deliberation of ends.

### 3.3

Serial hyperspecializers move from niche to niche. So they will need a signal that tells them when to invest more resources in a current activity and when to abandon the activity for another—that is, when to move niches. There can certainly be different indicators of success, but one of them will be competence in the activity. If a method of doing things is working, that's a sign that the strategy of occupying this particular niche is workable; if it's not, that's a sign that it's time to move on. In our own lives (here's that reality check once again), the negative signal is frustration, the feeling that you're banging your head against a wall, that you just can't do it; the positive signal is one aspect or specialized sort of pleasure, an often elated sense that *you can do it*: that is, the cognitive content of one sort of pleasure is competence; a related variety

involves the sense that, as one puts it colloquially, it's all coming together.<sup>17</sup> These complementary signals make up a great deal of what counts as overall hedonic tone. (The early utilitarians were right about this much: what they called "pleasure" and "pain" are enormously important in human life. Their mistake was to think that being important had to be: being a *goal*—instead of being a *signal* used to reformulate goals, as well as other standards, guidelines, and priorities.)

A successful serial hyperspecializer will not just respond directly to signals like these (*keep doing this! stop that!*). It will extract patterns from the signals that it uses in making intelligent decisions about niche switching, and in tailoring its goals as it adjusts to a new or changing niche. From the inside, this is generalizing from experience (philosophers say: "induction"); and because the content of the conclusion isn't just factual, but is about what *to do*, and what *to go after*, it is "practical induction."<sup>18</sup> This sounds fancy, but it's something we're all familiar with from our daily lives (which is, again, the reality check). Maybe you notice that whenever you're photographing people, you're awkward with the models and uncomfortable with your product and, not coincidentally, that the work is making you miserable; but that when you're animating clay figurines, the time seems to fly, you get better and better at it, and because you're getting better, you're enjoying what you do. You decide that what works for you, what you should be devoting your time to, is working with clay rather than photographing human subjects. In short, serial hyperspecializers will reason inductively about what their goals and other priorities should be. They will learn what matters, and what is important to them, by generalizing from experience. If necessity is the mother of invention, practical induction is its midwife.

Now, in the instrumentalist model of practical reasoning, if your ends don't come with a built-in feature that tells you which one is more important, when they conflict, you will perhaps be helplessly paralyzed, or perhaps choose randomly; in any case, you won't be able to make a *rational* decision. Instrumentalists respond by conceiving of desires as having strengths; the alternative—desires that can't be weighed or measured on a uniform scale, that is, *incommensurable* ends or desires—is thought of as a threat to practical rationality.<sup>19</sup>

---

<sup>17</sup>That is, when Vogler (2002, pp. 80–89) reconstructs Aquinas's definition of pleasure, "the unimpeded *operatio* of a habit that is itself in harmony with one's nature," as the sense of a nexus of "agent, action, and circumstance"—the sense that it's all coming together—she is actually describing the content of the signal I have been characterizing. It follows (and here I am taking issue with Vogler) that, when the signaling system has been corrupted, such pleasures can be mistaken (or, as the traditional vocabulary has it, "false").

<sup>18</sup>The terminology is from Millgram (1997).

<sup>19</sup>For an overview of this debate, see Chang (1997).

But let's think about incommensurability from the point of view of serial hyperspecialization. We expect a serial hyperspecializer to come equipped with signals that tell it when it's time to switch niches (and on a smaller scale, when it should reallocate resources among activities pursued within a niche). These signals, we suggested, will tell it something with the approximate content: *This* is working, and *that* is not. To use the signals, the response ought to be, other things being equal, to stop doing *that* (or do less of it) and allocate more resources to doing *this*.

For such signals to be available, and such an allocation decision to be available, there have to be a *this* *and* a *that*. What this means is that serial hyperspecializers will sometimes be (the oxymoronic sound of the phrase notwithstanding) *parallel* hyperspecializers as well. (Only sometimes, however, because while it will sometimes pay for serial hyperspecializers to hedge their bets by pursuing different activities in parallel, dividing up your resources between different activities means that you have less to devote to any one of them.) Let's consider an instance where this sort of strategy is being pursued, and let's take a case where the activities in question are very different from one another: someone who divides her time between medical journalism, avant-garde space installations, and a country home in a small town, to any of which she could devote more or less time and energy.

Evidently, when activities are as different as this, they will involve structurally different standards and modes of assessment. As a medical journalist, for instance, there is a fairly clear hierarchy of clients, and you're doing better as your clients come from higher in the pecking order. As an artist, there's also a (perhaps less clear) hierarchy of professional acknowledgments; for instance, if you're invited to exhibit at the Kassel *Documenta*, you're doing pretty well. Around the country home, your real estate values could be going up or down, and you could be getting along better or worse with your neighbors. But the standards won't normally specify how to make tradeoffs between them. Is a space installation for the municipality a step up from a survey article for such-and-such a trade journal? Neither set of professional standards has a take on this. Is it more important to improve one's standing in the neighborhood by devoting extra time to gardening, or to work on lining up a performance venue in the Congo? Again, the methods of assessment proprietary to the respective activities won't say.<sup>20</sup>

So one reasonable strategy for a hyperspecializer will be to divide its energies between activities that don't share standards and methods of assessment, and let's abbreviate that to the fuzzier, problematic, but more familiar word,

---

<sup>20</sup> And when they *do* say, the answers will often look oddly off-base, because what counts as success by one standard will be given no credit by another. If you're a businessperson, and your measure of success is return on investment, technological elegance is just a distraction; but if you're the researcher, financial success is the distraction.

“values.” It will be a normal side effect of pursuing the parallel hyperspecialization strategy that its values are incommensurable.<sup>21</sup> The hedonic signals that guide reallocation of resources between niches do not require that niche-bound desires or goals or standards be comparable across niches. If you are a serial (and so, often a parallel) hyperspecializer, incommensurability in your values turns out not to be a mark of practical irrationality, or even an obstacle to full practical rationality, but rather the way your evaluative world will look to you, when you are doing your practical deliberation normally and successfully. Evaluative incommensurability is a threat to the sort of rationality suitable for Piltdown Man. Serial hyperspecializers gravitate toward—and come *equipped* for—incommensurability.

### 3.4

I earlier introduced instrumentalism as the standard model of practical rationality, but that needs to be amended: although many philosophers (and not just philosophers) adhere to the letter of the view (only means-end reasoning counts as practical *reasoning*), very few have the nerve to leave it at that, and allow that whatever you happen to want is what you have reason to try to get. (People want all *kinds* of things.) Thus, in contemporary moral theory, the recommended inputs to means-end reasoning are typically made out to be *informed* preferences or desires: not the ones you actually have, but the ones that you *would* have, if you were improved in one way or another—paradigmatically, by being better informed. Faced with a choice between two movies, the right decision is the one you *would* make if you knew more: not just who the actors are, but how the acting and the directing turned out; whether the sentimentality has redeeming social value; whether the sets and cinematography are too slick and overproduced, or rather unobtrusive and convincing; and so on.<sup>22</sup> Informed-desire theory (to keep the prose readable, I won’t keep on mentioning the preference-based variant) is motivated as a way of correcting desires, but not by reasoning; it’s the spoonful of sugar that helps instrumentalism go down. Because this variant on instrumentalism is such a staple of contemporary moral theory, let’s increase the resolution of

---

<sup>21</sup> For a different but compatible bounded-rationality account of the incommensurability of values, see Morton (1991).

<sup>22</sup> Just what it takes to be better (or “fully”) informed is a difficult and contentious question. By way of giving a foretaste of the problems involved in spelling it out, the relevant information, in the movie example, seems to be the information you’d collect by actually seeing the movie; but if you *had* seen the movie, you might well not relish the thought of seeing it (the surprise twist at the ending would no longer be a surprise), and you might not be able to resuscitate the responses that made it such a good idea to see it for the first time. For a discussion of this sort of problem, see Millgram (2005a, p. 53, n.21). For discussion of other difficulties with informed desires, and for pointers to representative views of this sort, see Enoch (2005).

the rendering, and consider, in less survey-like fashion, how this correction technique will do when it is applied by serial hyperspecializers.

The exotic niches which make up a human social ecology can differ arbitrarily from one another, both structurally and in the substance that fills in those structures. The niche-specific systems of standards that serial hyperspecializers develop to navigate the niches they inhabit are tailored to the arbitrarily varying structural features of the different niches, and so they too can differ arbitrarily from one another, both formally and in their content. Competent use of these systems of standards normally requires internalizing them: making them one's own and, typically, losing the ability to explicitly describe one's use of them, in something like the way that someone who has learned to ride a bicycle can no longer explain how he does it, or someone who has learned a language to native or near-native fluency can no longer explain why a particular way of formulating an expression is the right grammatical and stylistic choice. Because the niches are themselves so elaborately articulated, the systems of standards are commensurately articulated and consequently cognitively expensive: people who haven't paid the costs of entering a specialized niche normally don't comprehend the standards that govern its inhabitants. For instance, part of becoming a philosopher is learning how to assess philosophical work, and if you haven't had that sort of training, you don't understand, and aren't able to manage, that sort of assessment.

For an informed-desire theorist, the force of an ought comes down roughly to this: it is or will get you something you would want if you knew more. There are many variations on the ways in which informed-desire theories suggest counterfactually adjusting your desires: if you knew more about the options, or if you knew everything; if you thought about it in a "cool hour"; if you were to undergo "cognitive psychotherapy"; and so on. Let's divide up such appeals to informed desires into those that involve counterfactual adjustments we might reasonably take a shot at implementing and those to which feasibility considerations are not meant to be relevant. In the first category, we find such recommendations as: that you act on the desires you would have if you had more in the way of the relevant facts at your fingertips, and then took a deep breath; this amounts to advice you could (often if not always) take. In the second category, we find such proposals as: your reasons are given by the desires that a highly idealized version of yourself, one who was perfectly rational and who knew everything, would recommend to you; these evidently are provided solely as an in-principle criterion of correctness and not as usable advice. In each of these categories, let's now ask what will come of deploying its version of informed desires: first, to make a decision within a niche a serial hyperspecializer currently occupies and, second, to make a decision in an as-yet-unfamiliar niche.

Given a certain amount of lead time, a serial hyperspecializer will normally have explored a niche it is already in and either borrowed a system of

standards from previous occupants of the niche or developed its own. It will normally adopt goals (i.e., desires) relevant to decisions that are local to a niche by invoking the available standards, and because those standards have been shaped and debugged by ongoing investigation of the niche, such desires will generally (although not always) be a reliable guide to action. Hyperspecializers that are capable of the imaginative techniques described by simulation theory—roughly, taking parts of one’s mind offline, adopting tagged “pretend” beliefs, and discovering, by running a simulation, what desires and other responses one would have to the “pretend” situation—will accordingly find that deploying informed desires is generally a good idea—*when the choice situation is local to the niche*.<sup>23</sup> After all, what they are really doing is exploiting an investment they have already made in tools for navigating that niche. I take it that the effectiveness of such simulations, in a well-explored niche, explains why so many philosophers have found the turn to informed desires so reasonable; in these circumstances, it *is* reasonable.

But now consider what will happen when these sorts of simulation are used to make a decision in an unfamiliar niche, one that a serial hyperspecializer does not yet occupy. The standards invoked by such exercises are the standards it *has*—that is, a system of standards suitable for the niche it is in, and so not normally a system of standards suited for the niche that frames the decision it is facing. When I try to imagine what I would strive for as a singer-songwriter, I find myself vicariously wanting to write more literate and argued lyrics than one hears on the radio, because I am now a philosopher, and the standards I have internalized prioritize literacy and careful argumentation. In the music world, these standards are beside the point, and a competent singer-songwriter hopes, I am told, to manage lyrics that draw on his own experience and let the audience share that experience, lyrics that speak to everyone, that don’t sound like all of one’s *other* lyrics, and that nevertheless don’t try too hard—but not lyrics that are well-argued, and not necessarily lyrics that are particularly literate.<sup>24</sup>

If a serial hyperspecializer attempts to address the problem by cobbling together a system of standards for a niche it has not explored thoroughly enough to understand, the standards it comes up with will normally be wildly inappropriate. And more generally, simulators are attractive devices because they efficiently exploit information and guidelines which are already (tacitly) available. Attempting to tweak simulation techniques to run where the tacit information and guidelines are lacking will produce fragile and badly behaved kludges, rather than smoothly executing and trustworthy procedures.

---

<sup>23</sup>For an overview of simulation theory (although one which advocates replacing the label), see Nichols and Stich (2003, Chapter 2).

<sup>24</sup>I’m grateful to Pepe Chang for being an informant on this topic.

For out-of-niche decisions, then, the appeal to implementable informed desires gives systematically erroneous results. To further explore the appeal to informed desires, let's relax our interest in implementability, and consider purely in-principle criteria for the correctness of the desires or goals that are to trigger one's means-end reasoning. A fair representative, given that our problem cases are out-of-niche decisions, might be: the desires which give you reasons are those you would have as a seasoned and knowledgeable occupant of the relevant niches. For example, a political science professor who is about to reinvent herself as a fiction writer ought to be acting on the desires, goals, or preferences she would have as a long-time and professionally savvy novelist.

Recall that a mental state such as a belief or a desire is *occurrent* when it is right there in your mind (say, when you are thinking of it, or, on some computational views of the mind, when it is explicitly represented in memory). A mental state is *dispositional* when it is ascribed to you in virtue of counterfactuals such as, If I asked you whether the President of the United States has a navel, you would immediately answer that he does. We ascribe to you the dispositional belief that he has a navel, even though you have never considered the question, and even though it is very implausible that it was explicitly represented in your memory. Informed-desire theorists are in the business of appealing to such dispositional mental states—specifically, desires you would have in such-and-such circumstances.

We now need to distinguish two sorts of context in which we might be called upon to ascribe such states. We are comfortable ascribing to someone a belief that his name is Hannibal, even though he is not thinking of it just now. After all, if we asked him, he would respond right away with his name. But we are not at all willing to ascribe to, say, a lectern the belief that its name is Hannibal Lectern, on the grounds that if a sorcerer turned the lectern into a human being, and we then asked it, it would tell us that its name was Hannibal (and we would not *even if* that is the answer it would give).<sup>25</sup> There are probably no sharp lines to be drawn between the *near in* dispositions that support such belief ascriptions and the *far out* dispositions that fail to do so, but the distinction is real—just as mental states ascribed on basis of far out dispositions are not.<sup>26</sup>

Transforming an assistant professor of political science into a fiction writer, or a philosopher into a singer-songwriter, is not quite as extreme as transforming a lectern into a human being, but from our point of view they are much the same sort of thing. To suppose that we can in this way pick out the standards that would guide, for example, the writer's choice of goals and desires in her new niche—the constraints specific to the genre she has not yet chosen, the

---

<sup>25</sup>Lectern name suggested by Jamie Dreier.

<sup>26</sup>For a related difficulty with instrumentalist appeals to informed desires, see Millgram (1996, pp. 208–209).

demands of the idiosyncratic writing voice she has not yet evolved, the authorial agenda she has not yet adopted—is on a par with insisting that the lectern knows its name, albeit dispositionally. In moving from one niche to another, serial hyperspecializers transform themselves too deeply to support this sort of appeal to their dispositions.<sup>27</sup>

Informed-desire theory has looked reasonable because, when it amounts to relying on imaginative simulation within a well-explored niche, one for which you have internalized a usable set of standards, it *is* reasonable. It is a pretty good (although perhaps imperfect) heuristic, one that would work across the board for a creature that never moved niches—that is, for Piltdown Man. But treating informed-desire theory not as a locally useful heuristic, but as one’s principled philosophical account of the correct inputs to means-end reasoning, would make sense only if the niche a person lived in were (entirely and forever) his whole world. That is, informed-desire theory, deployed as the standard supplement to the standard theory of practical rationality, is not suitable for serial hyperspecializers, for whom invoking out-of-niche informed desires means either using the device where it is bound to deliver inappropriate outputs, or engaging in a bit of illegitimate philosophical fantasy. For human beings, informed desires are a decent way of exploiting investments that they have made in learning their way around a specialized niche, but the results produced by a cognitive apparatus tailored to exploit such investments do not travel.

---

<sup>27</sup> In the course of responding to unrelated difficulties with the view, a number of informed-desire theorists have adjusted the view into what now gets called ideal advisor theory: the inputs to your instrumental practical reasoning are not what *you* would want if, say, you knew more, but what an ideal advisor, who does know more, would recommend. Let’s quickly walk through the move tree to see whether the adjustment counts as an improvement.

Either the advice is to be thought of as coming from within a niche one already occupies or from outside it—let’s just say, from a different niche—and let’s divide up that latter option into two cases: the niche has been well-explored by others or it has not. If the advice is actually coming from someone inside your well-explored niche (although actual as opposed to imaginary advisors are not usually what ideal-advisor theorists have in mind), as before, taking it seriously is a reasonable thing to do: you are exploiting an investment made in exploring that niche. (However, see Millgram, 2005a, p. 174, for an evidently overlooked problem with such advice; insufficient attention is given to what really makes someone a good advisor.) If the advice is coming from a never-explored niche, there is no investment to exploit, and the appeal to what an ideal advisor would tell one is, as in the analogous informed-desire case, illegitimate philosophical pretense. If the advice is coming from another already-explored niche, there are two sorts of very deep limitation on what it can do for you. First, recall the professor of political science who is reinventing herself as an author: although advice from other, already experienced authors can be valuable, no one, not even an imaginary self who *has* already made the leap, can tell her what her distinctive authorial voice should sound like. (If she took the voice over from an ideal advisor, even from a somehow-perfected identical twin, it would thereby cease to be *her* voice.) Second, because you do not control the intellectual apparatus of niches you do not yet occupy, you will be unable to understand the opinions that an ideal advisor might have to offer: it will have to be dumbed down for you, in ways that make it far less inferentially valuable than the ideal-advisor view presupposes. For discussion of this latter problem, see Sections 4.7 and 2.3.



## 3.5

Having said that we are serial hyperspecializers, I now need to take some of it back. I just described serial hyperspecialization as a solution to a design problem, and I claimed that we engage in it. But we're by no means optimized for it. For one thing, the hardware platform that the self-modifying software runs on is only just slightly changed from its disposable predecessors: we *die*. We do live longer than many animal species, but the hardware wears out, and still exhibits what in the artifactual world would be perceived as planned obsolescence. People can change careers midlife, but as they get older, they get less flexible, and learning new skills, new languages, and new attitudes gets harder. Even when they can do it, they often stop *liking* it. Some techniques and their associated priorities and guidelines seem to be only acquired in youth: mathematics, for example. And there's a lot of behavior (and associated desires, goals, and ends) that is relatively inflexible. As the theory of natural selection (along with some obvious extra premises) would predict, it often has to do with reproduction, food, and other basic maintenance, like sleeping. Some things are just too important to be left up to the self-modifying software.<sup>28</sup>

That raises a practical question for us: *should* we be serial hyperspecializers? The question has a precedent early on in the history of philosophy. In Aristotle's *Nicomachean Ethics*, he first described the person who excels at a life of public activity. (Think of a wealthy community leader, someone who engages in politics but isn't really a professional politician.) Then he went on to describe another life of which humans are also capable, the life of contemplation. (Think here of a mathematician, someone who spends his time, all of his time, in an office in an ivory tower thinking about hard-to-imagine mathematical objects.) We *can* aim for either sort of life, so the question is, Which (or how much of each) should it be? Aristotle thought that contemplation was a higher and more god-like activity, and so he recommended that. The analogy notwithstanding, I don't propose trying to answer our question in those terms. But we have a very similar question before us. We could work at being better serial hyperspecializers, or, alternatively, we could forego that option and try to become better Piltdown Men. How ought we to think about the choice?

The choice of whether to work on being a better serial hyperspecializer, or a better Piltdown Man, is in one respect less open than it might seem, because we can only entertain it if we are pretty good approximations to being serial hyperspecializers already. Full-fledged Piltdown Men don't ask themselves this sort of question; someone who does is considering becoming a Piltdown

---

<sup>28</sup>That said, there's surprising flexibility even about these. The variations in what people will eat and how they will go about eating it, along with the different ways humans go about reproducing, highlight just how much latitude the self-programming software really has.

emulator, rather than a full-fledged Piltdown Man. Since we *are* entertaining it, we are equipped to engage in practical induction—that is, to learn, from experience, what does well for us and what does badly. And so we should ask: which strategy has worked well in the past? Do the people who adapt to new circumstances do better, or the people who don't? Do the people who try out new things and explore new ways of living do better, or the people who don't? Do the cultures and societies that change and adapt do better, or the cultures that keep the old ways? Once again, these questions are not to be answered in terms of prior, given goals, but by looking to the signals of success and failure that are part of the cognitive equipment of serial hyperspecializers.

At first glance, the answer might seem pretty unequivocal. Examples at the scale of societies—in which failure to adopt the methods and techniques of economic, political, and military competitors, and the systems of standards embedded in them, gives rise to one or another form of catastrophe—tend to divide up into the politically charged and the unfamiliar, and so I will leave these to the reader as an exercise. But for a recent example at the scale of industries, recall when slide rules, not so long ago, vanished off the face of the earth. The people who had worked in slide rule manufacturing divided up into two categories: those who learned how to do something else and those who had dead-end lives.<sup>29</sup>

At second glance, the tradeoffs are more complicated. First, one possibility is that sometimes the serial hyperspecializing strategy does better, and sometimes the Piltdown Man strategy does better. If people exhibit the relevant sort of second-order plasticity, we might expect them to sometimes live one way, and sometimes the other. Which alternative makes most sense will depend largely on one's environment, and especially on how elaborately articulated one's human ecosystem is; if there isn't already a lot in the way of what I've been presenting as the cultural surrogate for biodiversity, there aren't many niches to occupy, and so there isn't much advantage in being prepared to occupy novel niches. Perhaps very poor and war-torn societies today don't reward serial hyperspecializers; it's likely that in our distant past, human societies were also not complex enough to do so; indeed, for most of our evolutionary history, average lifespans simply weren't long enough to permit much in the way of serial hyperspecialization. (For that reason, it is unlikely that we are serial hyperspecializers because ancestors who adopted this strategy were thereby winners of the Darwinian game: the capabilities we are discussing look to be serendipitous, from the point of view of natural

---

<sup>29</sup>You might wonder whether this question is after all trivial: don't serial hyperspecializers do everything that Piltdown Men do, and more? So why isn't serial hyperspecialization the inevitable choice, if you're in a position to make it? That would be to forget the overhead involved in serial hyperspecialization; in describing the cognitive machinery it involves, we are also indirectly enumerating some of its costs, and these are often quite high. (For a brief but suggestive list of costs of developmental plasticity in the natural world more broadly, see West-Eberhard 2003, p. 431.)

selection, rather than an evolutionary adaptation.<sup>30</sup>) Which alternative pays may also depend on the mix of strategies already in play in one's society. For instance, perhaps being Piltdownish is a good strategy in an economy that already contains enough serial hyperspecializers to keep it flexible and afloat. If the dot-commers in 2000, or the Wall Street bankers in 2009, find new opportunities in other industries, you can keep doggedly working at your retail job until you retire.

Second, part of the design solution that human beings implement is the lengthy childhood that permits them to learn not just a system of standards for the environment into which they are born, but methods of learning the systems of standards appropriate for subsequent environments. That requires long-term parenting; parenting is often tedious, frustrating, and variously unpleasant, that is, it requires overriding the signals that tell us to stop what we are doing and do something else.

Those caveats notwithstanding, as far as the big picture goes, my guess is that the first-glance practical-inductive take on whether to invest one's resources in serial hyperspecialization, and whether to endorse the canons of practical-inductive reasoning, has it right. Piltdown Man is boring, and serial hyperspecializers are interesting. Social institutions and lives tailored to Piltdown Man are frustrating; creativity, novelty and originality, intellectual and otherwise, *feel* much better. And to be a Piltdown Man is for changes in one's environment to be nonrecoverable catastrophes. It's a no-brainer.

### 3.6

Theories of practical reasoning are closely tied to corresponding ethical or moral theories.<sup>31</sup> And so while I don't here want to try to delineate a moral or ethical theory that would be suitable for a community of serial hyperspecializers, I do want to indicate, albeit tersely, why the moral theories we have won't do. Each of the theories on the standard menu is naturally taken as according primary importance to one or another possible moral priority; without giving a great deal of argument, I am going to suggest that it would be unreasonable for serial hyperspecializers to accord these priorities the primacy they are given by the standard theories.

Utilitarianism takes the satisfaction of preferences or desires—or, alternatively, hedonic tone (pleasure and the absence of pain)—to be the only thing

---

<sup>30</sup>There is a further reason for not insisting that being a serial hyperspecializer is an adaptation: it would be hard, in principle, to give a legitimate argument for the claim. The decently concrete just-so story that such an argument would be built around requires a definite phenotype to work with, and to be a serial hyperspecializer is precisely to have an *indefinite* phenotype.

<sup>31</sup>This claim is spelled out in Millgram (2005a, Introduction and Chapter 11).

that matters intrinsically. But in the lives of serial hyperspecializers, preferences, desires, and pleasure are cognitive tools; they are no more the source of all value than any instrument can be.<sup>32</sup>

In migrating birds, desires and ends are hardwired, part of a design solution that gets the organism to the next stop on its route. Since the bird will normally be doing well only when it does get to the next stop, desire satisfaction is a good proxy for success. Piltdown Man is only one step away from migrating birds: his strategy for coping with his environment is preprogrammed, and once he steps off the straight and narrow path that it defines or presupposes, he becomes dysfunctional. So for Piltdown Man, also, treating desire or preference satisfaction as success, or a proxy for success, is a reasonable ethical strategy. But because serial hyperspecializers compute desires, preferences, and the like on the fly, what matters is what the desires and preferences (and signaling channels) are there to facilitate: thus, it is always an open question whether the desires or goals one has are *correct*, and always an open question whether the hedonic signals have been corrupted. Accordingly, for serial hyperspecializers, none of these are plausible candidates for being the ultimate repository of value, or even a usable proxy for it.

Kantian moral theory is famously built on three large ideas, and I'll consider just one of them now. The point of having serial hyperspecializers is that they can exploit narrow niches—niches that don't have room for many occupants. That means that, for serial hyperspecializers, the Kantian question, What would happen if everybody did that? (and more carefully, the first formulation of the Categorical Imperative) is simply beside the point.<sup>33</sup> Only the most primitive ecologies contain just one role; hyperspecialization is a strategy suitable for highly articulated ecologies. But because Piltdown Men, living on the plain, are all peasants (and more generally, because they all pick their life strategies from a short menu of fixed options, one which it's reasonable to think of as a single disjunctive strategy), the question, What would happen if everybody did that? is a reasonable test to impose on *them*.<sup>34</sup>

Aristotelian moral theory focuses on the shape of a well-lived life (which Aristotle called *eudaimonia*, or happiness). Aristotle meant to read that shape off of an account of the human "function," that is, off of the functional design

---

<sup>32</sup>Notice that even though some of these tools effectively measure relative success within a niche, and seem to work within many different niches, it does not follow that there is a substantive notion of success that is common to all those niches, and which might amount to a utilitarian's generic goal or end. Here is a relatively local illustration of that point. Athletes may have a sense of how well they're doing at a given sport, and they may appeal to it in deciding which sports to pursue professionally. But sports are very different from one another, and there is nothing substantive that success in the various sports may have in common.

<sup>33</sup>For a quick description of the so-called CI-procedure, see Millgram (2005a, pp. 90f, 141f).

<sup>34</sup>Here and below, the problem is no longer that the agents' desires are treated as fixed; diagnosing the lack of fit between serial hyperspecializers and the theories of practical rationality that underlie Kantian and Aristotelian moral theory will have to be deferred to another occasion.

of the human species. So in that respect his method has a good deal in common with the present approach. But, first, if the account I have been sketching here is correct, Aristotle was mistaken about what design solution human beings (approximately) implement, and so his substantive ethical advice can be at most accidentally right. Second, because, on the present account, much of the shape of a human life, at the level of specificity which Aristotle hoped to capture, is computed on the fly, and properly varies from individual to individual, there is much less in the way of generic but substantive ethical theory to be had than Aristotle thought.<sup>35</sup> And third, where Aristotle was quick to identify successful performance of one's species function with happiness or flourishing, on the present account the connection is anyway much less direct: even if successful niche-hopping is something you had better be able to do, it is hard to believe that human happiness *consists in* the exercise of this capacity.

There is a further difficulty with eudaemonism. Eudaemonism assumes that there is a coherent pattern of activities that will constitute a human life's going well. Aristotle made that assumption because he really did treat ethics as continuous with biology: when you find out how squirrels live their lives (during the summer they bury nuts, and during the winter they retrieve and eat those nuts), you know what it is for a squirrel's life to go well (they succeed in finding the nuts they put down); likewise, when you find out how humans

---

<sup>35</sup>Thompson (2008, Part I) is a recent and very sophisticated attempt in this tradition to delineate species form via "Aristotelian categoricals" or "natural-historical judgements," that is, sentences with the logical form of "Bobcats have two to three cubs in the spring." The problem with using this conceptual apparatus to make sense of ourselves is that, for serial hyperspecializers, the Aristotelian categoricals come and go. "Humans illuminate their dwellings by burning whale blubber" was an observation briefly on a par with "Beavers assure their food supply by building dams." But it is true no longer, having joined the class that includes "Humans prepare documents using typewriters" and that will shortly include "Humans select items from drop-down menus by pointing and clicking." I leave it to the reader to extend the list; what matters for the present point is that this sort of transience extends very far into what is normally the subject of natural history. For instance, "Human females nurse their young" was a natural-historical judgment that might have seemed to belong to the species form, if any did; but there has been substantial movement away from that Aristotelian categorical over the past hundred or so years.

The Aristotelian categoricals that are stably true of humans are mostly of a piece with the ones we have been surveying over the course of this paper, and amount to a description of the ways in which the species form is plastic. (Another way of putting it: the most important Aristotelian categorical about *our* species is that the Aristotelian categoricals true of it change from decade to decade.) I am tempted to conclude that, by the lights of such neo-Aristotelians, we must not after all be a species. In any case, the intellectual apparatus they provide us for reasoning about species has very little usable grip on us.

Thompson himself attempts to peel off the problematic natural-historical judgments by assigning them to a different logical category, having to do with artifacts. The parrying move is unsuccessful, because what matters is not how you categorize those judgments, but what (and how much, or rather, how little) is left over when you delete the ephemeral ones. And, although this is an incidental matter, the reassignment, which turns on the requirement that for such a statement to be true of artifacts, people must know that it is, is inconsistent with the first-person authority that Thompson insists on with respect to our own species form.

live their lives, you know what it is for a human's life to go well. That approach makes sense for squirrels (and for Piltdown Man), because squirrels (and, let's suppose, Piltdown Man's programs) have been around long enough for the rough edges in their form of life to get buffed down; their lives, when they go well, do exhibit coherent patterns of activity. However, if I am right, the forms that human lives take are extremely transitory; it would be unreasonable to suppose that there has been enough time to debug them. And that expectation is confirmed by observation: for the most part, human activities exhibit—once you press on them even a little—deep incoherences. If humans are serial hyperspecializers, and if happiness is taken (as philosophers have traditionally taken it) to involve coherence conditions on one's aims, activities, and so on, then happiness is not a guiding concept appropriate for human beings (although it may well be a suitable guiding concept for Piltdown Man).

The classical moral and ethical theories are impressive intellectual accomplishments, but they are misdirected. Moral philosophers have almost without exception been in the business of designing moral theories for a different species—and they have been quite successful at it. Utilitarianism, Kantian moral theory, and Aristotelian ethics would be highly suitable guides for the communal life of one form or another of Piltdown Man. But they are not at all appropriate guides for us, if we are serial hyperspecializers—or even if we are reasonable approximations to them.

### 3.7

How could so many philosophers have misidentified the very species for which they were philosophizing? The mistake is even worse than it sounds. In producing a theory of practical reasoning that fits Piltdown Man, our human philosophers were after all philosophizing. But philosophy consists largely in the exploration of new intellectual niches, of designing standards for those niches, and setting intellectual goals to pursue in them. Therefore, Piltdown Man does not philosophize.

My own guess is that there have been several contributing factors, and I will mention just two. First, philosophers of the past, even when they were not apologists for traditionalist societies, tended to see much more in the way of stable social roles than we do. Second, and more interestingly, while the philosophical positions constructed around Piltdownish renderings of humanity are various, instrumentalism is perhaps the most central. Now, we tend to forget that instrumental rationality was a hard won achievement. To reconceive human beings as instrumentally rational was a revolutionary and liberating step forward: against sheer rote and custom, it opened up possibilities for criticism, and paths to rapid social reform, that we have since come to take for granted. (Such Enlightenment figures as the Benthamite utilitarians notably

exemplified this frame of mind.) Forcing social institutions to make instrumental sense was a heady and exhilarating enterprise, and an admirable one in its day. A commitment to treating instrumental rationality as the be-all and end-all of practical thinking was, not that long ago, a passionate political commitment.

Perhaps we are not yet finished with the possibilities opened up by instrumental rationality: lip service to the contrary, there is still much less of it around than you would think. Nonetheless, Piltdown Man was a primate species whose best-known property turned out to be *not existing*.<sup>36</sup> The notion that all reasoning about what to do is means-end reasoning, and that you can't reason about what ends to have—that's a theory of rationality suitable not for humans, but for an imaginary species of dwarves, a stunted primate that might have evolved but didn't, and which, if it had evolved, would have had to occupy a very different role in the larger ecology than we do.<sup>37</sup>

---

<sup>36</sup>For the latecomers, Piltdown Man was a hoax involving a human skull, the jaw of a modern ape, and sandpaper; Millar (1972) is a popular account.

<sup>37</sup>I'd like to thank Chrisoula Andreou, Sarah Buss, Pepe Chang, Steve Downes, Sean Kelsey, Mark LeBar, Anya Plutynski, and Connie Rosati for comments on earlier drafts. For helpful conversation, I'm also grateful to Jay Bernstein, Karin Boxer, Havi Carel, Ben Crowe, Christoph Fehige, Gabriele Juvan, Kathrin Koslicki, Chandran Kukathas, Tamar Laddy, Brenda Lyshaug, Peter Momtchiloff, Huw Price, David Schmidtz, Wayne Waxman, and David Wiggins, as well as audiences at a Rosenblatt Free Lunch and a Law School Brown Bag, both at the University of Utah, at Birkbeck College, and at the Bowling Green Conference on Practical Reasoning.

## D'où venons-nous . . . Que sommes nous . . . Où allons-nous?

In retrospect, Bernard Williams was, during the second half of the twentieth century, perhaps our most refined philosopher of common sense. He spent his life exploring a handful of ideas current either among philosophers or the general public, and typically both. The one that will anchor our discussion is that there is a deep difference between matters of fact and evaluations or decisions. A closely related thought, which we will also take up, is that there is an important distinction between science, which is fully objective, and ethics, which is only subjective. A further and connected idea, which will serve as our entry point into the discussion, is that your reasons for action bottom out in whatever it is you happen to care about.

The theses on which Williams concentrated his intellect count, if anything does, as the collective default view about their respective topics, but they have also proven surprisingly hard to formulate satisfactorily. Thus although they have been much criticized over the course of their run of popularity, it has been hard to tell whether the criticism sticks to the ideas themselves, or rather to the clumsy ways they were put. But Williams returned to them repeatedly, carefully recasting those thoughts, and the reasons for them, in ever more nuanced, ever more stripped-down form.<sup>1</sup> With a good philosopher's concern

---

<sup>1</sup>Take Williams's engagement with cultural relativism, the thesis that there are claims which are true for, or inside, a culture, but not for members of, or within, other cultures. It appears as early as his first book (Williams, 1993a, pp. 20–25), which refutes a “vulgar” version of it. He subsequently returns to the topic in his piece on “The Truth in Relativism” (Williams, 1981a, Chapter 11); as its title suggests, it is an attempt to articulate the important thought of which vulgar relativism is a failed expression. Relativism is taken up once again by *Ethics and the Limits of Philosophy* (Williams, 1985, Chapter 9) and is a subtext of *Shame and Necessity* (Williams, 1993b, especially Chapter 5), in which the questions of how deep cultural divides realistically turn out to be, and whether and how cross-cultural ethical assessments are possible, occupy much of Williams's attention. Finally, the topic comes up in *Truth and Truthfulness* (Williams, 2002, pp. 52, 258ff). Cultural relativism will appear now and again at the fringes of our discussion.



for the consistency of his intellectual commitments, he worked insistently at tracing the connections between them, and both negotiating and sharpening the tensions among them.

Philosophy of the analytic variety is evidently at a turning point. If we are to determine where it is to go next, we must first take stock of where we are, and for such stocktaking, we cannot do better than to turn to Williams. His views are the most intelligent and articulate expression of what someone who is reasonably mainstream and reasonably hardheaded can be presumed to think. So we may be fairly sure that when larger criticisms stick to Williams's mature view, they are registering real objections to ideas that philosophers and the public tend to take for granted, rather than superficial difficulties with inept renderings of them.

I am going to recapitulate the evolution of Williams's understanding of a few of these ideas and sketch a constellation into which his mature thought assembled them. The path through the material will be just a bit complicated: because much of our interest is in the distinctions Williams worked to articulate, we will be shuttling back and forth between the subject areas on the contrasting sides of one or another such distinction. For expository purposes, I will organize the first part of the presentation around his criticism of what he called the "morality system," but bear in mind throughout that my interest is not in the relatively familiar arguments against morality: they are sketched merely to display some of the intellectual machinery at work in them. These arguments have generated a large literature which I will not engage here, and where there are differences of opinion between myself and other readers of Williams as to how the arguments are supposed to run, I will not now take up the controversy.

Once the pieces have been put on the table, I will turn to assessment and diagnosis. Those pieces amount, I will suggest, to a rendering of what it is to be a human being, and what sort of world it is that humans are suited to inhabit. I will ask what presuppositions would have to be in place for the creatures Williams has implicitly described to make sense as a design decision. And I will conclude that the unintended lesson of Williams's work is that we have made an astonishing mistake about who we are. The philosophical common sense of the past half-century has been suitable for impossibly simple-minded creatures, creatures competent to live only in impossibly simple environments. Consequently, the descriptive metaphysics and ethics that have been spun out of it are useless to us.

#### 4.1

Williams's early critique of utilitarianism shows why it pays to unload the excess baggage from a vehicle of philosophical thought. Doing so allowed him

to elicit an incoherence in the most straightforward way of spelling out and accommodating one of the widespread ideas that preoccupied him.

One's reasons for action, Williams believed, are a matter of what one cares about. Philosophers have for the most part rendered this thought as instrumentalism: roughly, that one has a reason for action only when the action would be a means to the object of a desire one has. (Slightly more colloquially: the only reason to do something is that it will get you what you want.) Williams put a great deal of thought into streamlining this awkward and rather blunt theory.<sup>2</sup> Especially generous about what could do the work foisted by instrumentalists on narrowly conceived desires, he introduced a generic label for the more broadly conceived class: your "subjective motivational set." Williams habitually and indifferently referred to its elements as "desires" or "projects," but an indication of how much more flexible and inclusive he meant to be was counting emotions and loyalties among them, and we will shortly encounter a further variety of subjective motivation, not normally regarded as desire-like at all, namely, the disposition to apply a certain sort of concept. Williams was equally generous about how, formally, reasons invoking one's motivations could work. They would not have to be what many philosophers think of as means-end reasoning in its strict sense, that is, finding a cause for an object of desire and adopting the intention to bring it about, and he allowed for much more freedom of movement in addressing one's motivations and concerns.<sup>3</sup> The resulting view, which came to be called "internalism," retains the requirement that reasons for action be capable of jointly explaining and justifying the actions in connection with which they are adduced; the substantive but pared-down claim which Williams endorsed was that you could neither explain nor justify if your subjective motivational set did not contain suitable elements.<sup>4</sup>

---

<sup>2</sup>See Williams (2001), which includes a postscript to the original essay, and Williams (1995a, Chapter 3). He has not been alone in thinking that there is something at the core of instrumentalism that escapes the numerous refutations of it; for a very different attempt of this sort, one which follows Anscombe in seeking to depsychologize means-end rationality, see Vogler (2002).

<sup>3</sup>There is a well-known short list at Williams (2001, pp. 80f), which mentions scheduling, coordinating, adjudicating conflicts between ends, being imaginative, and finding "constitutive solutions," that is, what now gets called specificationist deliberation.

<sup>4</sup>Millgram (1996) explores the connections between instrumentalism and internalism, and argues that they are substantially congruent views—which, if Williams was trying to express the thought at the bottom of instrumentalism, is just as it should be.

However, now that I have praised Williams for his deft reformulations of familiar notions and questions, let me gesture at one of those exceptions which prove the rule, and remind one that Williams, too, could make mistakes. Notice an asymmetry in the dichotomy between internal reasons (reasons that depend on the presence of a suitable element in your subjective motivational set) and putative external reasons (those that do not). If the reason depends on a motivation (any motivation at all), it counts as internal; all reasons for action, Williams argues, involve *some* subjective motivation; therefore, all reasons are internal. But why wouldn't it be more helpful to allow for a mixed category of reason in one's taxonomy? Even when there is some tie to a motivation, why can't some of the force of a practical reason nevertheless come from elsewhere? And if there are such cases, why assimilate them to those in which all the force is derived from subjective motivation?

If one's projects and desires are what underwrite all of one's practical reasons, then they must be very important indeed. We ought—and this is another widespread idea which Williams explored, and which, to get ahead of our story, he eventually rejected—to be able to capture and represent their importance to us in a suitable theory. Utilitarianism, the moral theory whose slogan ran, *The greatest good for the greatest number* is perhaps the most straightforward way of doing so. Williams decomposed utilitarianism into its consequentialist structure, on which actions are chosen to produce the highest-ranked globally assessed outcomes, and a substantive view as to what it is in virtue of which outcomes are to be ranked: happiness, subjectively construed, that is, the extent to which the objects of one's subjective motivations are realized.<sup>5</sup>

Williams pointed out that in even-handedly taking everyone's subjective motivations into account, utilitarian agents could easily be put into the position of having to surrender their own projects and desires. Worse, because this eventuality is bound to be anticipated, intelligent utilitarians will come to hedge their emotional investments in their projects long before they are actually required to replace them, in something like the way that employees who are liable to be fired on twenty minutes' notice avoid becoming overinvolved in their jobs. Utilitarians end up lacking "integrity," meaning that they do not act for the reasons they have, but for reasons that they have been bamboozled into thinking they have. When such agents disinvest in their projects, the subjective commitment to them which was evidently the source of their importance is washed away. A world of self-consciously utilitarian agents would be a world in which projects and desires lacked the vigor needed to support the seriousness with which utilitarian theory proposes to attend to them.

Still worse, such agents are unable to acknowledge the peculiarly personal importance of their projects and concerns to themselves. Desires and projects

---

Compare Williams's taxonomic setup and its upshots to the ways we find it useful to think about theoretical rationality. Call an *internal conviction* one that is in any way derived from convictions you already have. An *external conviction* will be one derived entirely from elsewhere—perhaps, one might expect, from observations. However, if you don't believe your observations are any good, you can observe anything you like, and it won't change what you believe. So *theoretical internalism* (analogously, the view that all of your reasons for belief depend on convictions you already have) is true of theoretical reasoning.

But now, isn't the interesting thing about observation, and its effect on belief, that the observations come from outside? Even if this is only possible against a suitable background of already-present conviction, surely much of the force of reasons for changing one's beliefs comes from something other than one's already-present convictions. A set of terms that makes this truism invisible, by making theoretical internalism almost trivially true, is certainly a mistake. Why shouldn't reasons for action likewise derive their force *both* from one's subjective motivations *and* from further sources—for example, from a practical analog of observation? Given the questions Williams means to address by taking up the question of internalism, it must be an error to settle the outcome of the debate by choice of terminology.

<sup>5</sup>Williams (1973a); it is worth keeping in mind how much of Williams's discussion is conducted in the shadow of Sidgwick (1907/1981).

are at the bottom of—for present purposes, pretty much *are*—all of one's practical reasons. So one's reasons to go on living must be projects or desires. Williams noticed that some projects or desires are conditional in content: an old man in a retirement home may want to play golf, but only as long as he's alive anyway; it is not as though he would choose to live longer in order to play more golf. By contrast, he might also want to see his granddaughter graduate and will do his very best to stay alive until she does. Desires or projects that are not conditional on one's being alive anyway (like that desire to attend the graduation) are, in Williams's lexicon, *categorical desires* or *ground projects*. These would be no less vulnerable to being overridden in a utilitarian decision than any other desires or projects; internalism casts them as indispensable ingredients of one's reasons to stay alive; thus, utilitarian moral theory is likely to demand that one give up one's very reasons to go on. Moral theories are, logically, advice, in the sense that they purport to tell you what to do. A good touchstone for the acceptability of advice is whether it undercuts one's stake in one's life: generally, one shouldn't take advice if, were one to take it, one would have no reason to go on living. So utilitarianism gives advice that one should be entirely willing to ignore, which is just to say that it fails as a moral theory.<sup>6</sup>

## 4.2

The incoherence of utilitarianism has to do with the attempt to capture the person-by-person importance of desires and projects (consisting or expressed in being the source of a person's own reasons for action) in an impersonal theory of what would be on the whole best. This first-pass diagnosis invites the response that if utilitarianism is trying to have its cake and eat it too, some other moral theory will do better: say, a Kantian (or, as the terminology went, back when Williams was starting out, a “deontological”) theory.

---

<sup>6</sup>In Williams's usage, ground projects are categorical desires writ large; they're either more important such desires, or enterprises or pursuits driven by and involving categorical desires. Thus, a categorical desire is a little ground project, and a ground project is, or is built around, a big categorical desire; thus there is no deep difference between them, as far as Williams is concerned; and accordingly we can move back and forth between the terms casually.

The argument we just saw renders the observation that utilitarianism can run roughshod over one's stake in one's life this way: one's most weighty or important concerns are what give one reason to go on living. But I doubt that the identification can be sustained, and so I doubt that the rendering has been done correctly. The Stoics believed in caring only about what one can control; they took it that one can control whether one is virtuous, but not the length of one's life; so history records Stoics whose deep commitment to their virtue was nonetheless conditional on their being alive anyway. A more modern version of such a personality might care most deeply about taking life as it comes; might think of death as part of life; and so would *not* make a point of living longer, in addressing his deepest concern. The worry here is that the construction which Williams provides is—like similar constructions in the work of Frankfurt, Velleman and others—a surrogate for an important idea we are trying to use, and one which can be at best extensionally adequate, rather than an explication of the idea itself.

But Williams's characteristic practice of revisiting earlier attempts on a problem led him to broaden his earlier claim: the mistake was moral theory itself, not just the utilitarian flavor of it.<sup>7</sup> His progress toward his eventual understanding of that conclusion started off with attempts to think through another widely accepted idea, that there is a difference between—as some of the usual ways of marking the hard-to-put contrast have it—values and facts, or “ought” and “is,” or practical reasoning and theoretical reasoning (that is, thinking about what to do versus thinking about what the facts are). Williams first took up the question as it was framed by the metaethics of the period, that is, as asking what distinguishes, on the one hand, desires, commands, or statements about what ought to be done (I'll say “oughts” from here on out) from, on the other, beliefs, assertions (exclusive of those oughts), or evidence for belief. Williams's developing view of that distinction passed through three stages.

Among the many formal contrasts he found in his sensitive exploration of it, *agglomeration* came to the forefront early on.<sup>8</sup> Beliefs agglomerate: if the belief that *p* is in order, and the belief that *q* is in order, then the belief that *p* and *q* must be in order—though we'll have to qualify this claim shortly. Desires do not agglomerate: I want to go to Palo Alto today, and I want to go to Berkeley today, but I do not want to go to Palo Alto and to Berkeley today. Imperatives do not agglomerate: when you are instructed by one authority to close the door, and by another to leave it open, each command may be perfectly in order, yet their conjunction is an incoherent and impossible-to-execute proposal for action. In a paper on ethical consistency that became something of an instant classic, Williams argued that oughts do not agglomerate. Although he allowed that this last conclusion was unobvious and his argument for it less than knock-down, a theory of tragedy seemed to follow from it: tragedy consists in having obligations that are jointly unsatisfiable.<sup>9</sup> Williams was impatient with the tendency of analytic ethics to keep “all the important issues . . . off the page . . . and [the] great caution and little imagination . . . used in letting tiny corners of them appear”<sup>10</sup>; that taking failure of agglomeration to be a central formal feature of the practical allowed him to broach an ambitious theory of tragedy would have been, in his eyes, one of the more powerful attractions of his analysis.

The view that seemed to be emerging at this point was that considerations about matters of fact are to be formally distinguished from practical reasons

---

<sup>7</sup>Williams provided a straightforward extension of the argument against utilitarianism to Kantian and, incidentally, other consequentialist theories, which the interested reader can find in Williams (1981a, Chapter 1). Here the individual agent's stake in his desires and projects is treated as an independent premise, rather than a commitment of the theory under attack.

<sup>8</sup>Williams (1973b, Chapters 10–12).

<sup>9</sup>For the reservations, see Williams (1973b, p. 182).

<sup>10</sup>Williams (1993a, p. xviii).

and their relatives in that the former agglomerate and the latter do not.<sup>11</sup> The-ory is responsible to facts—to what is really out there—and so must come out the same from any point of view.<sup>12</sup> Decision and evaluation, and now Williams could invoke his internalism in an explanatory role, are driven by subjective motivation, which comes in packets associated with persons and which allows for a great deal of not-very-coordinated variation among the packets.

Williams's understanding of the distinction went through a next round of modification and complication when he served on the Committee on Obscenity and Film Censorship.<sup>13</sup> In proposing legislation, it is hard to avoid attempting to define its key terms, and in doing so, he observed that concepts such as “pornographic” and “offensive” function as guides for action and that full control of such concepts requires already occupying an appropriate evaluative stance. This seemed to place them squarely on the practical side of the distinction he had been honing. Nevertheless, such concepts figure into beliefs and assertions and are responsible to matters of fact; this seemed to place them on the theoretical side of the distinction. And indeed their applications *behave* like beliefs, in that they agglomerate: if I believe that it's adult material and I believe that it's offensive, I'd better believe that it's offensive adult material. Williams argued that the mix of features exhibited by *thick ethical concepts* could not be accounted for by supposing there to be distinct theoretical and practical thoughts in play: the application of a thick ethical concept does not factor into factual and evaluative components and so seems to straddle the fact-value distinction.<sup>14</sup>

The obstacles which thick ethical concepts posed for the agglomeration-based distinction directed Williams back to a further contrast which had

---

<sup>11</sup> Intentions have come in for much discussion since Williams wrote these early essays (e.g., Bratman, 2001) and they might seem to be an obvious counterexample: they are evidently practical, but the intentions of a single person are required to agglomerate. Notice, however, that intentions do not agglomerate across persons.

Williams's claim in any case requires reasons for action to undergo fairly drastic regimentation. Suppose that one's reason for going to a cafe is that they have a manually controlled espresso machine; Brandom (2001) takes after Anscombe in holding the reason to be just that, and not also that one has a desire for coffee made with such a machine. On such an understanding of practical reasons, one would not find failure to agglomerate to be a likely marker of them.

<sup>12</sup> Williams (1973b, pp. 202–205).

<sup>13</sup> It was “appointed . . . to review the laws concerning obscenity, indecency and violence in publications, displays and entertainment in England and Wales” (Williams, 1981b, p. 1).

<sup>14</sup> The term was being adapted from Geertz (1973). I once complained that Williams had not supported this last claim with satisfactory argumentation (Millgram, 1995). However, I had overlooked the sometimes quite amusing arguments at Williams (1981b, pp. 119–123). Williams's exercise in applied ethics should be read more frequently by philosophers than it is, both because it nicely exhibits methodological differences between applied and theoretical ethics, and because it identifies the concrete instance used by Williams to think through the topic of thick ethical concepts. The thin and moralistic-sounding list he gives at Williams (1985, p. 140) is quite misleading; it becomes much clearer why Williams arrived at the conclusions he did when one redirects one's attention to concepts like “obscene.”

long been of interest to him. That our subjective motivations figure into our applications of our concepts is a special case of those contributions to concepts which make them more local, more subjective, and more tied to the idiosyncrasies of our constitutions. Think of how our color concepts reflect the idiosyncrasies of the human visual apparatus and are therefore local to human beings; if birds and bees could speak, they would have different color vocabularies; if mole rats could speak, they would have no color vocabulary at all. The disposition to apply a thick ethical concept is a special case of such contributions; it can be counted as an element of one's subjective motivational set; such concepts are typically shared within cultures, which is why beliefs that embed them agglomerate, but within single cultures and not, generally, across different cultures.

Now, our local and constitution-laden descriptions differ among themselves as to whether they are amenable to a certain sort of reformulation. In certain cases, we can, Williams supposes, “form a conception of the world which *contains* [ourselves and our] representations” and which will then permit the indirect agglomeration of what were formerly only locally agglomerable applications of concepts.<sup>15</sup> For instance, I may start out with the belief that litmus paper turns red in acid, and space aliens might start out with a belief that I am not in a position to so much as state, because it contains vocabulary tied to the deliverances of the aliens' peculiar sensory organs; these cannot be conjoined into a claim in our vocabulary or in theirs. Once our account of pH has become that of contemporary chemistry, we can describe what happens to litmus paper without having to deploy our secondary-quality vocabulary for color; once our color science is good enough, we can explain the response of our visual system; our transformed account could thus in principle be understood by intelligent space aliens, creatures who do not share our secondary-quality vocabulary. Our chromatic and other descriptions can thus be transformed so as to render them (and now we are giving content to the hazy words that figured in one of our popular *ur*-thoughts) more objective and less subjective. We introduce the Absolute Conception of Reality (or, as Williams sometimes also put it, when he was raising his eyebrows at the moral version of the ambition, the Point of View of the Universe) as the limit in which all knowledge amenable to such transformation has been rendered *completely* objective. Knowledge inflected by our constitutions need not agglomerate with the knowledge of alien communities, but knowledge so transformed as to constitute part of the Absolute Conception *does* agglomerate with any other knowledge so transformed: when we encounter the space aliens, our theoretical physics will join with theirs—again, this is Williams's understanding of science, and not one I am endorsing myself—to form a single body of doctrine.

---

<sup>15</sup>Williams (1978, pp. 64f); Ismael (2007) provides useful discussion of both what this process realistically looks like and what, realistically, are the benefits to be had from it.

It is a guiding aim of the scientific enterprise to produce knowledge that forms part of the Absolute Conception, and Williams takes it that this is not an unrealistic or misguided aspiration. But ethical knowledge—to a first approximation, knowledge in which the application of thick ethical concepts centrally figures—does not survive the reformulation we have described. When our subjective contribution to the application of a thick ethical concept (that is, in the first place, the role of our subjective motivational set) is made explicit, our ability to deploy the concept is undercut, and knowledge evaporates, in something like the way that the funniness evaporates when you explain the joke: reflection, Williams summarized his claim, destroys ethical knowledge.<sup>16</sup>

The deep distinction, then, that had been initially supposed to run between what to believe and what to do, or between fact and value, seemed instead to be the distinction between science and ethics. Scientific knowledge agglomerates, once suitably transformed; ethical knowledge does not. The enterprise of moral theory is defective in that it treats ethics as though it were science; that is, as though ethical considerations and conclusions could be gotten to agglomerate, whether they were anchored to different cultural backgrounds, or to different persons, or to different commitments copresent in a single person (recall Williams's explanation of tragedy), or to different times in a single person's life (the case exploited by Williams's famous discussion of moral luck).<sup>17</sup>

---

<sup>16</sup>For Williams's supporting argument, see Williams (2006b, pp. 289, 294). That famous slogan has been taken to mean that nihilism is inevitable in a reflective culture such as our own: pretty soon, all our thick ethical concepts, and so all our ethical knowledge, will be destroyed. Later on, Williams hedged, insisting that he had meant that our ethical knowledge *can* be destroyed, but not that it *must* be (1995b, p. 238). Whether the reason I am about to give was the point of the hedge or not, it does not follow. Even though ongoing reflection is busily whittling away at our ethical knowledge, we need not run out of workable thick ethical concepts if they are being introduced faster than they are being destroyed. That is evidently our own situation; think of such recent accretions to our repertoire as "eggshell plaintiff," "big organic," and "unputdownable". . . or slightly less recent but still undissolved instances such as "kitsch," "camp," and "politically correct." Still, such concepts are unlikely to serve in adjudicating disagreement as effectively as their vanished predecessors; since they come and go, they are much less likely to be widely shared.

<sup>17</sup>This was a further argument against the "morality system," and I will briefly sketch it so as to bring out the role played in it by the ideas we have introduced. The distinctive formal commitment of the morality system, reflected in its requirement that what we are bound to do can be systematized into something like a scientific theory, is that its oughts and its assessments more generally must agglomerate, in something like the modified sense we have just seen: like knowledge deploying our color vocabulary, they can be recast into a form that is conjoinable and consistent. Williams worked up thought experiments meant to elicit from his readers pairs of vicarious moral judgments, one in anticipation of a course of action, and the other, in retrospect and with the advantage of hindsight. The standards of assessment invoked in those judgments differ deeply. To take Williams's famous discussion of Gauguin as an instance, it was Gauguin's body of work that created the artistic standards by which it could be judged successful enough to justify, perhaps, the step of leaving his family for the sake of his art; obviously those standards could not have been appealed to before the course of action had been embarked upon. The cases are chosen with the intent of forestalling transformations, analogous to those that turn theoretical knowledge into components of the Absolute Conception, that make them out to express what is at bottom a single, consistent set of standards. If that intent succeeds, then (and this is the point of his argument) moral judgments that we find intelligible and compelling turn out to violate the agglomeration requirement to which the morality system is committed.



The morality system's distinctive commitment is to this very idea, roughly that there is available in principle a point of view of the universe on what it is right (permitted, and so on) to do. That is why morality is a mistake.

The problem with utilitarianism had at the outset been narrowly construed as an internal tension: that of trying to capture the importance to persons of their reasons for action, from the standpoint of no person at all. There were thus two ways of resolving the conflict. Having explored and rejected the attempt to retain moral theory by deemphasizing the importance of subjective motivation, Williams found that the fatal objection to utilitarianism was just that our thought about—to put it as neutrally as possible—what one should do, or how one should live, cannot be configured as anything like a scientific theory. Here we are seeing the benefits of paring the ideas in play down to their most skeletal form.

Williams was, again, our most refined philosopher, in that his *modus operandi* and signature style was progressive refinement. Opting for the alternative resolution, he began to revise and reformulate what had been traveling under the heading of “morality,” sloughing off its unsustainable commitments to agglomeration and systematic theory. “Ethics” came to be Williams's term of art for what he hoped would come of that reformulation, and if you are wondering what it would look like, *Shame and Necessity* is almost certainly meant as his demonstration of ethical thought cleansed of moral theory. But I am going to leapfrog it, and proceed directly to Williams's final investigation of the factual materials for which, he was convinced, theory remains an appropriate medium.

### 4.3

When distinctions like those between fact and value, or science and ethics, crop up, philosophers are all too often tempted to explain them metaphysically: in this case, by appealing to the notion that science is about the facts, and ethics isn't. Williams's early essays had flirted with such metaphysical explanations, and later discussions made much of the Point of View of the Universe or the Absolute Conception of Reality. Critical responses have unsurprisingly tended to focus on the merits of this sort of explanation.<sup>18</sup>

However, the complications which thick ethical concepts forced on Williams's rendering of the distinction made it clear that what is demanded is not (or, not just) a metaphysical explanation of it. The principled difference between science and ethics appears only when the attempt is made to reformulate knowledge as part of the Absolute Conception. Those attempts are expensive (a point we will return to shortly). So the pressing question is practical: why, of all the metaphysically available distinctions we could draw and emphasize, is *this* the one (or, *a* one) we would prefer to use? Accordingly,

---

<sup>18</sup>For example, Putnam (1990b); for a brief reply, see Williams (2006a, pp. 184–187).

when bouts of multiple myeloma forced him to choose the one project he would finish, Williams settled on *Truth and Truthfulness*: a book devoted to explaining, in terms of basic structures of human social life rather than those of metaphysics, the role and importance of truth.<sup>19</sup> The choice was well-motivated; the account of truth is meant to serve as the keystone of his lifelong reconstruction of common sense.

Several years previously, Edward Craig had developed some of Williams's early remarks into an account of knowledge.<sup>20</sup> Craig proposed sidestepping the exercises in conceptual analysis that still dominate epistemology by asking what function ascriptions of knowledge serve; he addressed his question by imagining a society that lacked the concept (a society in an epistemological "state of nature") and considering whether and why they would want to introduce it. The need he identified was a generic and transmissible certificate for information. To say "I know that *p*" conveys roughly that *p* is good enough to go on; that I can tell you so without having to ask precisely to what use you are going to put *p*; and that you can tell the next person to come along what I have just told you. (Of course, the first-person use of such a certificate is not necessarily the primary one.) Any society lacking such a certificate would have to invent one, and when they did, it would have roughly the contours familiar from our concept of knowledge.<sup>21</sup>

Now, Craig's treatment perhaps inadvertently highlights a choice that epistemologists have made, for the most part unawares: to provide a theory of the most generic epistemic certificate, rather than the many other more specialized certificates in circulation. Some examples of the latter: not all knowledge is *news* (which implies recent provenance and salience); publication in academic journals provides a very large variety of different sorts of epistemic certification; degree-granting institutions provide a very large number of distinct forms of epistemic certification to individuals (rather than to packets of information); labels on products may inform you that they contain chemicals "known to the State of California to cause cancer, birth defects, or other reproductive harm." The choice historically made by epistemology as a field is reasonable if the generic certification is especially important, but not otherwise.

---

<sup>19</sup>For Williams's gloss on this bit of shorthand, see Williams (2002, pp. 6–7). For a survey of themes in the book that I'm here leaving to one side, see Fleischacker (2004).

<sup>20</sup>Williams (1973b, pp. 145–147); around the time that Craig (1990) was published, Williams was known to begin lectures on unrelated subjects by urging his audiences to read it.

<sup>21</sup>Many devices could serve while varying a great deal in the details, and this is why, on Craig's account, an exact conceptual analysis of "knowledge" is an unpromising project. That implicit prediction is confirmed by Weinberg et al. (2001); the Chinese version of the certificate differs from the Western with respect to Gettier intuitions. (For a review of the Gettier literature, see Shope 1983.) Here is another way to see the consequences of Craig's account. Philosophers have been exercised by the question of what challenges we must be able to meet if we correctly claim to know something; they have tended to set the bar very high indeed. The standards a certification with a real social function must meet are negotiable, and normally the certificate is useless if they are set unrealistically high.

Williams self-consciously modeled his treatment of truth on Craig's account of knowledge. The function of the concept of truth was to be elicited from a state-of-nature argument, one that showed why, if we did not already possess and use the concept, we would have to invent or adopt it.<sup>22</sup> Human beings are social creatures that pool information.<sup>23</sup> Williams focused his discussion on the virtues that need to be in place if information is to serve as a shared resource: sincerity, the disposition not to withhold information from the pool, and not to intentionally add fraudulent information, and accuracy, the disposition not to add to the pool sloppy or shoddy information. What those virtues are meant to assure is the quality and the range of the contents of the common pool. Quality control of this sort requires an appropriate assessment concept. Truth is the feature of (and "true" the label for) suitable contributions to the pool.

"True" is the *generic* label for information appropriately added to the pool. There are perhaps fewer standardized nonepistemic certifications that stand to "true" as the more specialized epistemic certifications stand to knowledge. (Call these *specialized alethic concepts*.) That is because, in actual use, epistemic certificates tend to shoulder the real communicative work and so end up doing duty for the alethic ones. (Here's the sort of thing I mean. When you tell someone that something is true, normally you mean to assure him that it is—that is, to give him the generic certification that says he can go with it, on your say-so. But then you might as well just tell him that you know it.) There are, however, exceptions: the censor's *nihil obstat* confines itself to matters of concern to the Church. And ad hoc alethic certificates, with variable content, are common, as when you tell someone that an approximation is true enough (accompanied by an implicit or explicit gesture at the circumstances in which it is), or that a claim is true for present purposes, or that it's good enough for government work. Choosing to focus on the generic certification will be reasonable if the generic certification is especially important, but not otherwise. To anticipate, its centrality or otherwise will be determined by the use made of the contents of the pool: how people go about withdrawing or accessing the information in it.

---

<sup>22</sup>Its function, not its definition: Williams (2002, p. 61) followed his sometime colleague Donald Davidson in holding that truth neither needed nor bore definition. Williams invokes Craig very early in his discussion (p. 21), and again later on implicitly (p. 79); in doing so he runs together—ill-advisedly, in my view—state-of-nature arguments and the rather different exercise of Nietzschean genealogy.

<sup>23</sup>This a statement to be heard in roughly the register of: bees are hive creatures that store honey. But only roughly; Williams (1995a, p. 102) devoted a great deal of worry to what he called the Representation Problem, that is, the problem of how to talk about the phenotypic traits of a species that is essentially cultural.

I should emphasize that "information" is every bit as problematic a concept as "truth" and "knowledge" and is unlikely to be independently intelligible. That is, the train of thought we are following is not one of those attempts to explicate a difficult concept by substituting an easy one for it.

## 4.4

A few observations about this third stage of Williams's thinking about the distinction he meant to be refining, before we move on to the next round of argument. It had looked as though Williams was committed to there being just *one* deep distinction after which both philosophers and ordinary speakers were groping, and the real question was exactly what content it had. At this point, however, the earlier version of the distinction is back in play, side by side with what had looked like its successor. I am not now going to pursue the question of how that happened, and so we have, on the one hand, the earlier contrast between beliefs and assertions and, variously, imperatives, oughts, desires, and the like; the generic label, "true," accompanies the use of this contrast. On the other, we have the contrast between truths that can be rendered fully objective and those which can't, typically because they involve applications of thick ethical concepts. (And possibly the latter category should be understood to include as well those other expressions of motivation that aren't so much as true at all.) Let's think of the pool of more impressive information on the one side of *this* contrast as marked by another label, "objectively true." Because the points I mean to make about them are analogous, I will consider the distinctions in parallel. We have just seen a social explanation advanced for using the former; such an explanation would be equally in place for using the distinction between objective and merely subjective truth (whether or not Williams himself ever got around to giving it).

There is a delicate point here: To say that "true" is the (or a) generic alethic certificate is to say that we do not, in using it, have to specify which pool of information is in question, or who the certificate is meant for, or what its ranges of acceptable use are. However, that we do not specify which pool of information the certificate is for does not entail that there is only one such pool, but only that the work of picking out the relevant pool is done tacitly. For example, and this was one of the prompts to cultural relativism, different cultures may employ different ethical vocabularies. The different languages may effectively segregate their ethical observations into distinct, nonagglomerating pools, even though both cultures nonetheless use the generic tag, "true" (or rather, the words for "true" in the respective languages).

Now, putting a generic certificate into circulation implies that you can just go ahead and use it, without knowing anything *else*. Philosophers nowadays are much enamored of contextualist approaches to one problem or another, so notice that a generic certificate does not give you what you need to apply an item of information in a context-sensitive manner. The certificate's failing to tell you whether its use is restricted to particular contexts is equivalent to its not telling you when it is and isn't permissible to agglomerate that information with other items of information: restrictions can be represented as restrictions on agglomeration. And notice that it doesn't tell you whether *you* have, as it

were, the right security clearance for the information. When agglomeration is restricted, additional detail is required, to enable more nuanced data management. So a presumption built into use of a generic alethic certificate is that any items in the pool of information can be agglomerated.

Much information starts out its life with many restrictions on its use. Trivially, statements containing indexicals like “now” or “I” or “here” must be processed into a different form if they are to be used later on, or by someone else, or elsewhere. More importantly, much information is generated as approximations, idealizations, and their less-than-accurate kin, and whether these are usable or not depend on what aims or concerns one has. (There is no such thing as an approximation’s being good enough, plain and simple; to use an approximation competently, one must be able to answer the question, good enough *for what?*) Typically they are explicitly or implicitly accompanied by more specialized alethic certificates, such as “true enough, when  $v \ll c$ ,” or “acceptable when the volume is far enough away from the vessel wall.” Since what I mean by calling an alethic certificate “generic” is that it does not specify conditions of use, both “true” and “objectively true” will count for present purposes as generic alethic certificates. So, perhaps counterintuitively, we are thinking about more than one generic certificate.

If expiration dates, owners, and framing concerns are not specified on a generic certificate, that must mean that they are not needed (as a general rule, anyway). So the exclusive (or almost-exclusive) use of a generic alethic certificate implies that available information is being *transformed* into a non-restrictedly or indiscriminately agglomerating form, as it is being contributed to the pool. Thus exclusive focus on generic alethic certificates presupposes that the costs of making information indiscriminately agglomerable are being paid up front. How high those costs are depends, of course, on which pool the information is being added to: the merely true, or the objectively true.<sup>24</sup> Either way, they’re generally quite high—a point that seems to have escaped Williams’s attention because he was preoccupied by the dangers of wishful thinking and self-deception.

A policy of assuming high costs up front had better come with decent reasons, amounting to an argument that a pay-as-you-go policy would not be more efficient. Why not use specialized alethic certificates instead, and

---

<sup>24</sup>The Absolute Conception of Reality was a limit concept introduced as, roughly, the body of information we obtain by converting all information amenable to such conversion into a form that renders it maximally agglomerable with any other similarly processed information. Our second distinction thus presupposes or aspires to a pool of information that can be shared by absolutely anyone, including intelligent space aliens. It is obvious that science directed toward this goal will require a very large “investigative investment”—a phrase used at Williams (2002, p. 124).

This observation tells us that we really do need an explanation of very much the sort we are edging up on for generic truth: under what social circumstances will the distinction between fully objective truth and everything else be one for which we have a real use?

perform context-to-context conversions on a demand-driven basis? Why wouldn't it be more cost-effective to partition the contents of the pool of information into ranges on which a single user might plausibly draw (where the partitions are chosen to minimize the costs of conversion while enabling agglomeration within each range)? The alternative to a pay-as-you-go system is one in which indiscriminate usage is made of the information in the pool. So near-exclusive use of a generic alethic certificate practically presupposes widespread and indiscriminate use of the information in the pool.

Drawing distinctions around one or the other of these generic alethic certificates, and treating them as deep and important, makes sense if there is exclusive (or near-exclusive) use of those certificates. The overhead is perhaps more daunting in converting information into the format demanded by the Absolute Conception. However, the costs incurred in making information belonging to either pool indiscriminately usable is reasonably assumed only if information resources really are held in common: only if (pretty much) anyone could realistically help himself to any of the items in the pool.

Williams takes the significant contrasts to be information versus motivation, and motivation-inflected information versus fully objective information. Information agglomerates (within the local generic pool, or within the more ambitious generic body of information sharable among scientists of all species). Motivation fails to agglomerate (even when motivations are shared by groups, as when a culture shares the use of a thick ethical concept; such motivation-inflected information will not generally agglomerate across group boundaries). Indeed, it is this variation in motivation that makes the virtues of accuracy and sincerity into the urgent issues that Williams saw them to be: in human social life, individuals and groups are often at odds with each other, and thus our use of the distinctions marked by the generic alethic certificates is being explained as a consequence of our belonging to a species whose members compete uneasily with one another, but who nonetheless pool information.

## 4.5

The time has come for considering what design justification there might be for creatures for which Williams's account would be satisfactory.<sup>25</sup>

The elements of a subjective motivational set function as inputs to choice and decision, that is, as something on the order of guidelines or priorities.

---

<sup>25</sup>The heuristic has a precedent in Grice (1975, especially on pp. 37ff); he calls it "creature construction." At this stage in our collective intellectual history, I should not have to provide the following disclaimer, but here it is anyway: the thought experiment we are about to undertake does not require or imply that we are the products of a design exercise. A further disclaimer: neither does it purport to provide the elements of an adaptationist evolutionary explanation.

These do not generally agglomerate, and in fact the distinctive *logical* features of the different sides of Williams's contrasts are accounted for by the care taken to assure agglomerability on one side of the contrast, and not on the other.<sup>26</sup> Agglomeration and the possibility of correction are procedurally tied to one another: if A and B don't agglomerate, pressures for consistency (and so, what consistency amounts to) come to something very different than if they do. Formally, even if A is inconsistent with B, in that it wouldn't make sense to advance "A and B," that fails to show that there is anything wrong with either A on its own or B on its own (or with the reasons for them).

Thus the incorrigibility of elements of a subjective motivational set is equivalent to their not being subject to an agglomeration principle. That motivations fail to agglomerate is exhibited in the most striking logical feature of internalism (and of its cruder relative, instrumentalism), namely, that one's bottom-line desires and projects are incorrigible. You want what you want, and someone who insists that you are wrong to do so, when mistakes about such things as how to get what you want are not at issue, is just *bluffing*.

Subjective motivations can change, in all manner of ways, but they cannot be corrected, and this means that nothing could count as the rational investigation, on the part of such a creature, as to whether its bottom-line guidelines and priorities were correct.<sup>27</sup> Since the creatures do not correct their own motivations, the design strategy is reasonable only if they do not need to; in other words, only if, for the most part, the designer can equip them with motivations (or ensure that they pick up motivations from their surroundings) that will not need correction. That in turn is feasible only if the designer can anticipate the practical problems his creatures will face, and only if the guidelines his creature would need to negotiate them are sufficiently compact to be stored and accessed. Given plausible cognitive constraints on processing, memory, and so on, that in turn requires that the environment the creature is anticipated to face be both stable and simple.

Coloring in the line drawing, we see that Williams's alethic state of nature is something on the order of a tourist-brochure version of a village in the hills of Provence, where life goes on as it has since time immemorial. The villagers work their plots of land, growing the same grains and vegetables they always

---

<sup>26</sup>In accepting the labels "information" and "motivation" for the contrasting classes, we are allowing them whatever features enable the differing uses made of them; I am leaving whatever metaphysical questions these may raise for other occasions and, in particular, whether and why agglomeration can be assured on one side, but not on the other.

<sup>27</sup>That "bottom-line" requires reiterated emphasis: internalists attempt to accommodate what look like practices of correction by expanding the range of corrigible but derivative desires. Or again, they typically insist that you may be mistaken about what your desires are; here the correction is not of your bottom-line desires, but of your opinions about them. A further but related issue: some self-declared internalists have taken pressures for motivational coherence to allow for correction proper of bottom-line desires; because Williams thought otherwise, I won't consider this possibility further just now.

have; they herd their sheep and goats; they bake rustic bread and knit rustic clothes; they hunt rabbits and deer; they build houses out of the local stone; they marry and raise children; when they get old, they sit outside the village pub and drink pastis; they play boules in the park; eventually, they die, and are buried in the cemetery behind the church.

The internalist design solution is satisfactory for this form of life. The designer knows that his peasants will have to work the fields, so when it comes time to own a field and work it, they come to have a desire to do so. They need to be made to reproduce, and thus are built so that, when they get old enough, they will want to have children, or anyway want to do things that as a predictable side effect produce children. Not all of a subjective motivational set need be hardwired, of course; a disposition to mimic others, and to learn and adopt one's elders' thick ethical concepts, will keep the games of boules going and the pastis flowing. Because life in the mythical village never changes, there is no need to delegate to the peasants themselves the task of investigating what their motivations ought to be, and no need to equip them to correct their motivations; thus, there is no need to complicate their cognitive or normative systems with the gadgetry that would take. A Dinah Shore song called "Doin' What Comes Natur'ly" nicely captures the internalist's implicit faith in such programming.<sup>28</sup>

We live in a complicated world. Because it is complex and unstable, it repeatedly confronts us with practical choices that we cannot (and that an imaginary designer could not) anticipate; no set of preformulated guidelines could carry us successfully through the challenges we face. Even if an omniscient designer were somehow able to foresee all of the vicissitudes into which we might be thrust, their number and variety is too great to allow the case-by-case advice he might want to bequeath to us to be compressed into small enough packages—that is, packages capable of being stored and accessed by our small-finite cognitive systems. (Data compression formats of the sort I have in mind include utility functions and Aristotelian virtues.) Internalism could not be a satisfactory design for creatures that have to live in *our* world.<sup>29</sup>

---

<sup>28</sup>Shore (1991).

<sup>29</sup>This brief account should not on its own persuade an internalist, who will wonder why a course of action, or a life, should count as a failure if it is not a failure by the lights of one's own desires. He will also wonder why a designer cannot equip his creature with suitably generic ends: social status, perhaps, or even just plain happiness. For one way of filling in the argument, see Millgram (1997, Chapters 3–5, and Section 6.7).

We can now redescribe the contrast between morality and ethics. The internalist solution presupposes an environment stable enough to permit programming agents up front, with reasonable performance in the field assured. Unfortunately, it is obvious that the subjective motivations required to make societies run smoothly are not programmed into us. The "morality system" appears when the conclusion is drawn that the built-in motivations must be overridden by a set of rules. Such systems of rules are generally formulated to satisfy an agglomerability constraint; normally, they can be represented by a moral theory. What the morality theorist (or moral legislator) shares with the internalist is the assumption that the environment is stable and that we can know enough about it up front to



## 4.6

Creatures that pool their information (whether all of it or just the high-value part they allocate to their science), and pay the costs of converting it into a form that makes it generically agglomerable, are a sensible design decision if, realistically, each of them can withdraw and deploy arbitrary items of information from the shared pool. That requires, among other things, that all (or almost all) information in the pool be comprehensible by any (or almost any) of the creatures that might use it. And that in turn requires an environment simple enough to keep the representation of information uniform enough to allow across-the-board comprehension, given the cognitive limitations of the creatures in question.

Although there may be difficulties in getting the suspicious, lazy, and quarreling villagers to contribute accurate information to their shared body of knowledge, the obstacles to deploying and to agglomerating arbitrary information retrieved from the pool are temporary and surmountable; as far as the model goes, they are noise and can be ignored. While it may be, for instance, too late for a villager to use a bit of gossip he has heard, there are no obstacles on the order of: in principle he could not understand it. And that is true even though there is division of labor, and thus division of intellectual labor, within the village; the farmers tend to know things that the housewives do not, and vice versa, but each can understand what he or she is told by the other. In such an environment, assuming the up-front costs involved in a generic pool of information is a good strategy.

But, once again, our own circumstances are nothing like those of the peasants in Williams's state of nature; our forms of life are consequently nothing like theirs; consequently, the design solution that made sense for our mythical peasants would not be workable for us. And consequently, just as the practical rationality and ethics or morality that come with that design solution could not be, for us, a correct ethics or rationality, the descriptive metaphysics that comes with it could not be a correct metaphysics for our species.

A good way to get a sense for how complex our environment actually is would be to survey the strategies we are forced to use in representing and navigating it. I won't undertake that sort of survey here, but as a stopgap, consider how we academics and intellectuals divide up the task of understanding, defining subject matters small enough to be mastered by individuals, or by groups of individuals with shared training. We develop discipline-specific

---

design a satisfactory system of rules. That those who produce moral theories do not expect their rules to need subsequent correction is clear enough from the history and practice of moral theory: neither Kantians nor utilitarians nor Aristotelians allocate theoretical resources to anticipating the circumstances in which—and the cognitive mechanisms by which—their own theories will be shown to have been superseded.

forms of representation: not just specialized vocabulary and notation, but distinctive representational techniques, tailored to discipline-specific modes of reasoning and argumentation. In order to avoid being swamped by the complexity of maps even of particular subject matters, we deploy approximations, idealizations, and many other related devices.

If the world is complicated enough for these coping strategies to be necessary, then we have what we need to say why the image of humanity that Williams uses to explain the formal features of truth and science cannot be an image of us. Because we are intellectually specialized, the information we store is not in fact generically accessible: if you are not a specialist yourself, you cannot read the journal articles.

If we cannot expect across-the-board preprocessing that would permit the extensive use of one or the other of our generic alethic certificates, we might imagine that information is agglomerated *within* the narrow confines of a specialized subject matter. But once we have parcelled out the information to smaller groups of users, what we in fact find within scientific and other disciplines is only sporadic agglomeration of information, due to the already-remarked widespread use of approximation and idealization. As we have seen, approximations and idealizations must be tagged with conditions of use, and typically with goals or other concerns, and cannot be assumed to agglomerate. Notice further that representations of these sorts can be used only when one has sufficient control of both thick ethical concepts proprietary to a specialization and the goals that are internal to the disciplinary enterprise. Within a small group of specialists, the local skills and additional training required to manage the more complicated certificates and the hedged representations they accompany turn out to be affordable.

Now, if this is what science is really like, something has gone wrong in Williams's rendering of the science-ethics distinction. Recall that Williams marked the distinction by claiming that science, but not ethics, supports (anyhow, postprocessing) agglomeration—that is, that the results of science can eventually be shared with just anyone. But the idea that we can share information with just anyone is most plausible precisely when we are not considering science.<sup>30</sup>

---

<sup>30</sup>That is of course not to deny that there is a distinction, and no doubt a very interesting one, between the two; if there were not, we wouldn't be able to discriminate them. But Williams has made what is, by his own lights, a mistake in how he tries to say where the line is. "Scientific" is, formally, a thick ethical concept; to deploy it correctly, one must occupy the appropriate evaluative standpoint, one amounting to a nuanced appreciation for science done right. The point is made—if not quite in those words—by Putnam (1981, pp. 132–136); Williams was, if you think about it, attempting to factor "scientific" into its factual and evaluative components and to give a value-free definition of the freestanding cognitive or factual component—the very factoring and definition he had elsewhere insisted was not possible.

## 4.7

If I am right, the generic alethic certificates on which Williams lavished his attention could not have the near-monopoly on usage that his account presupposes. But then, what allowed Williams (and has allowed others) to imagine that they did? I'm going to sketch a candidate explanation, one which will have the added bonus of explaining the scientism that worried Williams, but to which—if I am hearing his emphasis on the Absolute Conception of Reality rightly—he fell prey himself.<sup>31</sup> The mark of scientism, as I will construe the disposition, is that of placing at the center of one's philosophy an emphatic insistence on what scientific theory and practice are like, but one that doesn't have much to do with the activities or products of actual science. (A crude but familiar representative might be physicalist metaphysics, with its confidence that everything supervenes on, for instance, the values of fundamental physical quantities at space-time points.)

Consider how information exported from within a discipline to outside users must be formatted. Elaborate qualifications on when it can be applied, and with what it may be agglomerated, must be for the most part stripped off: nonspecialists will not have the training necessary for this sort of information management. That is, the sort of information passed across disciplinary boundaries is, almost always, a watered-down version meant for popular consumption, and not the complexly qualified formulation that carries the most authority *within* the discipline. The impression that we operate largely with generic alethic certificates is, I am hypothesizing, a byproduct of the fact that, outside one's own specialization, one operates with *exported* information, which does bear simplified alethic certificates. (And because we do not see past the border of an exporting discipline, we do not see how vastly greater are the number of occasions on which less-simplified certificates are being relied upon.) But these simplified certificates do not have the force of the generic "true." Rather, they mark the exported information as "true-enough-for-you."<sup>32</sup>

---

<sup>31</sup>Williams (2006a, pp. 182ff).

<sup>32</sup>See Ravetz (1979, pp. 13, 103–104, 200–201) for some description of information that has been watered down for export.

We now have a way to say how cultural relativism misses the boat, namely, in its preoccupation with the first- and third-personal "true-for-us" and "true-for-them" that are its bread and butter. The second-personal form we have just introduced is far more important in the lives of members of a species like our own.

Confusion about the workings of "true-enough-for-you" certificates frequently plays out as confusion in public policy debates. For example, when creationists insist that the theory of evolution is only a *theory*, that is their way of expressing the observation that the results exported from various scientific specialties are accompanied, within those specialties, by complicated, qualified, and nuanced alethic and epistemic certificates. They take this to be a matter for complaint, because they assume that the uncomplicated certificates are and ought to be routine inside a scientific specialty and because they assume that the unqualified certificates are of higher quality than the qualified ones. Or again, work on

Allow that we are (these days, anyway) as intellectually specialized as I have claimed. What would be an appropriate cross-disciplinary information management strategy? In general, those who are equipped to criticize the results within a specialized field are those who are also in the field; critical attitudes are appropriate on their part. For those outside the field, the correct attitude is simply to accept the pronouncements of respectable representatives of the field as authoritative. Philosophers, standing outside of one scientific enterprise or another, are the consumers of the information such enterprises export, and we have seen that it is normally presented as much more cut-and-dried than it is, and as much simpler than it is. It is often enough presented imagistically, and it is all too natural for such consumers to turn an image of science into a metaphysics.

That is to misapprehend the point of what is after all a correct information-management strategy. The scientific philosophers' mistake is two-fold. First, they are misinterpreting the authority they correctly accord to a foreign discipline as warranted by the secure status of its results.<sup>33</sup> The justification for the deference toward authority required by the strategy is not the high quality of the information, but the incompetence of its consumers. If you are on the outside, such and such is authoritative, but if you are on the inside, it's the dubious and likely-to-be-superseded opinion of someone else's research group. Second, information tagged for export generally is massaged into a form which just about anyone can consume. So it looks (from the outside) like a science is delivering generically sharable information: information that you can agglomerate with any other information (or anyhow, any other information like *it*). But whether or not scientists are after the Absolute Conception of Reality, those aspirations are not what explain the exported information being in this way shareable.

Scientism in philosophy is a predictable side effect of failing to recognize how creatures like ourselves compartmentalize their information pool. Most of the alethic certificates in circulation do not come with elaborate warnings about their circumstances of use; that gives us the impression that we are

---

climate change is, among the specialists, managed using sophisticated and not-at-all-generic alethic certificates: there are many different models, each with its own management issues. Politicians who would like to disregard the prospect of climate change talk as though this meant that the scientists are not sure about what they are doing, or about their results, and we should not be, either. The politicians understand that the nongeneric alethic certificates are what underwrite the simpler certificates accompanying results exported to the general public; they mistakenly take the qualified certificates to be shakier currency than properly issued unqualified certificates; thus they take the unqualified certificates accompanying predictions of global warming to be thereby discredited.

<sup>33</sup>This is what gives discussion of the so-called pessimistic meta-induction its nervous edge. The cause of the misinterpretation is nicely rendered by Ravetz (1979, p. 103): "informal knowledge of the higher elements of the craft of scientific inquiry . . . is so different in character from that embodied in the published results, and is transmitted through a different channel, [that] it is not capable of the same universality of diffusion, nor of the same closeness of control of quality." (I have reversed the order of parts of the quotation.)

deploying a generic certificate that merely marks items belonging to a generic (or to a generic scientific) information pool. But when it does, we are failing to appreciate how our information management handles specialization.

## 4.8

We cannot conduct ourselves in the simple-minded manner of our imaginary peasants. Instead, we partition our environment into what we might as well think of as niches. We specialize, and indeed we *hyperspecialize* to the different niches we occupy. And it is a remarkable feature of human beings that our first specialization is not necessarily our last: an individual may over the course of his life move from niche to niche, respecializing to each of them. A good description of the strategy might be this: human beings are serial hyperspecializers. If that is the strategy we implement, the observations we have made fall into place as its consequences.<sup>34</sup>

Because we cannot know up front what the guidelines and priorities suitable for navigating a particular niche might be, when we enter a novel niche, we have to *learn*, in a very demanding sense of the notion, what the appropriate action-guiding standards are. That means that the cognitive elements that guide action in creatures like ourselves must be correctable, and that we can ourselves attempt to correct them, which in turn means that internalism must be false for our species.

Because we construct specialized representational systems for the different niches we occupy, and because reasoning and inference within a niche tend to rely heavily on approximation, idealization, and related techniques, information cannot be relied upon to agglomerate—neither across niches, nor within a niche.

Finally, although philosophers tend to argue about distinctions as though they were arguing about matters of truth and falsity, a distinction is neither true nor false. Quine famously argued that the analytic/synthetic distinction was impossible to draw, and Williams attempted much the same with the fact/value distinction. But the point of arguing that a distinction is hard or

---

<sup>34</sup>This is a good occasion to attempt to forestall a handful of misunderstandings. First, my occasional choice of vocabulary notwithstanding, the claims I am making about human beings are not to be mistaken for the exports of some scientific subspecialty; I hope to avoid the solecisms of scientism myself. Second, my descriptions of the surrounding novelty and complexity that account for our information management strategies should be understood as true enough for present purposes, but not as an entirely accurate rendition of how matters are backstage, when viewed from what Williams called the Point of View of the Universe. Third, my claims should not be expected to agglomerate indiscriminately with further and similar claims, and here's a way to see why. There have been many characterizations of "man" that are true as far as they go—for example, that man is the rational animal, and that man is the animal that laughs; but people laugh a great deal when they are being silly rather than rational. My description of humans as serial hyperspecializers is, like these other descriptions, partial and idealized.

even impossible to draw is that it is, as a *practical* matter, badly chosen. If a distinction is well or badly chosen, it can be so for a variety of other reasons as well. The real question is normally not whether the distinction is, metaphysically, *there*, but how much attention it should get. If we are serial hyperspecializers, the distinctions which will help us understand our lives and practices are not those that Williams spent so much time and effort articulating. And so I have been arguing that Williams's error was to devote his attention almost exclusively to the wrong distinctions, distinctions that matter to us scarcely at all.

#### 4.9

University presses do not like to pay for four-color cover art. I suspect that Williams made the case for using a Gauguin on the jacket of *Moral Luck* because the title of that painting—which I have taken as the title for this Chapter—asked the questions that Williams found he was trying to answer: Where do we come from? What are we? Where are we going?<sup>35</sup>

The questions were well-chosen; as the title of his next collection put it, what philosophy is in the business of is making sense of humanity. But if my argument to this point has been on target, Williams's answers were deeply mistaken, and I have been claiming that the answers were not his alone: one more time, Williams was our most refined philosopher of common sense. Analytic philosophy has done something that is quite peculiar: instead of making sense of humanity, we have been philosophizing for the inhabitants of a romantic fantasy of traditional peasant life—or, bearing in mind Williams's choice of illustration, a European fantasy of life on a South Pacific Island—and not for the sort of creatures any of us are, with the lives any of us have. (And when I say “any of us,” I mean, not real peasants either: farmers are not simple-minded in the way that the presuppositions of common sense require.) It would not be overstated to say that we professional philosophers have misidentified the very species for which we have been philosophizing.

In being critical of what I am claiming are central theses of contemporary philosophical common sense, and in recapitulating where that criticism has brought us, let us not forget that we are standing on the shoulders of a giant. It is easy enough to object to the picture of humanity implicit in internalism, in the preoccupation with the fact/value and science/ethics distinctions, and in other ideas belonging to the closely connected cluster of theses that Williams explored over the course of his career. But it is easy only because his ever more refined reformulations of them are so clear that we can finally see what it is that they really express.

---

<sup>35</sup>In reproducing Gauguin's title, I have regularized his orthography.

Analytic philosophy is visibly in a high-entropy condition; it needs a new agenda. Thinking through Williams has presented us with one. At its top are a descriptive metaphysics and an ethics—not necessarily theories, but rather a careful and uncluttered rendering of a number of interconnected ideas, of just the sort that Williams hoped to provide—suitable for *us*: not for impossibly simple-minded almost-people capable of living only in an impossibly simple world, but metaphysics, ethics, and an understanding of rationality for serial hyperspecializers.<sup>36</sup>

---

<sup>36</sup>For helpful conversation, I'm grateful to Amy Johnson and Gloria Park and, for comments on earlier drafts, to Chrisoula Andreou, Sarah Buss, Dan Callcut, Alice Crary, Heather Douglas, Sam Fleischacker, Mark LeBar, Alasdair MacIntyre, Toni Nicoletti, Valerie Tiberius, and an audience at the University of California, San Diego.

## Millian Metaethics

In remarks in passing in his essay on Auguste Comte, John Stuart Mill suggests that—as we would nowadays put it—utilitarianism is the only substantive moral theory that lives up to the standards of a positive metaethics. Mill's remarks are too abbreviated for speculation as to why he thought that to be more than speculative. So although I will get around to speculating, I propose in the first place to use them as an entry point into a discussion of what a positive metaethics could be.

Once we have sketched some of the ideas that Mill appropriated from positivism, I will introduce the *Puzzle of Positive Metaethics*. I will suggest (but not argue for) a reading of Mill's remarks that explains how he perhaps thought he had solved it, and then turn to the theory of practical reasoning to show how the solution it provides to the Puzzle allows us to exit the sterile back and forth between ever more sophisticated versions of noncognitivism, naturalism, and Moorean supernaturalism that make up the lingering history of twentieth-century metaethics. By the lights of Mill and Comte, the parties to that debate are, as I will explain, *superstitious*.

### 5.1

Toward the end of his first essay on Comte's life and work, Mill remarks that

theological and metaphysical conceptions [applied] . . . to the rule of duty, and conduct in life . . . [were] based, either on a divine will, or on abstract mental conceptions, which, by an illusion of the rational faculty, were invested with objective validity.

When these opinions began to be out of date, a rival theory presented itself to take their place . . . to [which] the term metaphysical,



in M. Comte's sense, cannot justly be applied. All theories in which the ultimate standard of institutions and rules of action was the happiness of mankind, and observation and experience the guides . . . are entitled to the name Positive.<sup>1</sup>

Taking the happiness of mankind as the standard for political and personal decision is the familiar doctrine still taught from Mill's short classic, *Utilitarianism*. However, although Comte has been quite influential at various times and places (and even has a museum dedicated to him in Paris, the Maison d'Auguste Comte), in the English-speaking world his positivism came to be overshadowed and displaced by logical positivism, a movement with a similar-sounding name, but very different ideas and forms of expression. In order to determine what claim Mill is advancing on behalf of his own ethics, we first need to reintroduce the terminology he is adopting from Comte.

As they mature, Comte held, sciences progress through three stages, the *theological*, the *metaphysical*, and the *positive*. In the theological stage of a science, facts are explained by invoking deities or other supernatural beings. For example, in the theological stage of physics, objects move because gods push them around; and in the theological stage of biology, human beings are alive because God has breathed life into them, or invested them with an immaterial soul. In the metaphysical stage of a science, the gods, spirits, and so on are no longer in play, and their role is taken by forces; in the metaphysical stage of physics, objects move because forces push them, and in the metaphysical stage of biology, life is explained by a vital force (VIII:929n). Forces, however, are just sanitized superstition; the *élan vital* is just a theologian's soul in disguise, and a physical force is merely a somewhat less colorful god or spirit. In a fully mature science, forces have been expunged; all that remains are patterns. In the sciences, these patterns allow one to make predictions on the basis of observations of parts of the pattern. When positive physics describes the orbits of the planets, it allows you to predict where they will be in the future; when positive biology picks out patterns in the structure and activities of organisms, it allows you to predict how a particular organism will, in the normal course of things, develop and behave. Although Mill found many of Comte's views to be wrongheaded (and some, even ridiculous), he accepted this analysis of progress in science as an important insight, and in his own *magnum opus* attempted to extend it to logic. The *System of Logic* is Mill's dramatic and insufficiently appreciated attempt to move the science of logic from its metaphysical stage—in which it presents itself as the study of logical necessity, a force exerted perhaps on reasoners, or perhaps by the premises of arguments

---

<sup>1</sup> References to Mill's work by volume and page in the standard edition of Mill's works (Mill, 1967–1989); “Auguste Comte and Positivism” is to be found at X:263–368; the quote is from pp. 298f.

toward their conclusions—to its positive stage, at which we see logic to be merely a science of patterns.<sup>2</sup>

A quick terminological parenthesis: To keep the exposition uncluttered, I'm going to use "ought" as my representative of and stand-in for the family of terms that are the markers of ethical or moral or practical subject matter—I mean terms such as, for instance, "good," "should," "right," or "duty."<sup>3</sup> What is more, I'm going to construe them broadly rather than narrowly: as figuring not only into precepts covering the treatment of our fellow human beings, but into all sorts of pronouncements about what to do, what not to do, assessments whose primary function is to determine what is to be done, and so on. Finally, because use-mention pickiness turns into visual clutter after a while, I'm going to take the quotes off the oughts, even when I am talking about the term "ought," and I'm going to let context distinguish between mentions of the term or notion and assertions or directives whose main operator is an ought.

The announcement that with utilitarianism ethical or moral theory has entered its positive stage accordingly tells us that moral theory too is going through a similar process of maturation. Mill means, then, first that moral theory at one point went through a theological stage, in which the ethical requirements—the strictures that laid down what one ought and ought not to do—were accounted for as the dictates of God, or of the gods.

Mill further means that, to the extent that moral philosophy has emerged from the theological phase of its development, it is for the most part trapped in its metaphysical stage. Mill of course had in mind his own time, but we can be quite sure that he would take turn-of-the-millennium moral philosophy not to differ significantly in this respect. In its metaphysical stage, moral and ethical theory delineates the forms, operations, and nature of a peculiar but compelling force. Registered by oughts and their ilk, this force is referred to in the contemporary literature as "normativity." But, and Elizabeth Anscombe has made us familiar with this line of criticism, the alleged moral ought is merely a divine command with the backstory that had once made it intelligible taken away; in the old days, when people believed in God, moral oughts were a superstition; now that people no longer believe in God, moral oughts

---

<sup>2</sup>For the full story, see Millgram (2009b). It is quite unusual for Mill to allow humor into his writing, but he makes an exception when the time comes to mention certain features of Comte's religion of humanity (X:343f). However, we should not allow Mill's evident amusement to obscure his respect for Comte's accomplishments. Bain (1882, p. 63) reports Mill as stating that Comte "makes some mistakes, but on the whole I think [the *Cours de Philosophie Positive*] very near the grandest work of this age."

<sup>3</sup>I'm not, of course, suggesting that only such terms can be so used. There has been a good deal of discussion contrasting "thin" terms like these with "thick ethical concepts"—see Millgram (1995) for an overview of some of it—and one can obviously discuss moral matters without using a distinctive vocabulary at all.

I'm putting to one side the "epistemic" versions of "ought" and its relatives—that is, the usages that appear in such exchanges as this one: "Have you seen my umbrella?—No, but it ought to be in your backpack." I'll touch on these below, in footnote 43.

are an unintelligible superstition.<sup>4</sup> When moral philosophy emerges from its metaphysical teething stage, the illusion of this peculiar force will vanish, and ethics as well will simply render patterns.

Metaethics as we know it is in the business of explicating the force of the moral ought and its relatives. When Mill tells us that utilitarianism is positive, he is claiming that it no longer postulates an intangible force analogous to the forces of physics, at that science's metaphysical stage, or to the logical necessity that coerces the processes of inference, in the metaphysical stage of logic, or to the vital force that putatively animates living things, in the metaphysical stage of biology. Positivism about ethics, restated in our own vocabulary, seems at first glance to be the position that metaethics has no subject matter.<sup>5</sup>

## 5.2

In the exemplary domains we have reviewed, positivism commits us to deleting necessity operators. Where the operation of a force allowed a metaphysical-stage physicist to conclude that an unsupported object *has* to fall with an acceleration of 32 ft/sec<sup>2</sup>, the positive-stage physicist concludes that it *does* fall with an acceleration of 32 ft/sec<sup>2</sup>. Where logical necessity allowed the metaphysical-stage philosopher of logic to determine that the conclusion of a sound and valid argument *must* be true, the positive-stage logician concludes merely that it *is* true.<sup>6</sup> All of that seems like something a reasonable philosopher can swallow: since you never *saw* the “necessity,” but only the truth, the appearances (adapting a turn of phrase taken over by another positivist philosopher) have been saved.<sup>7</sup>

Analogously deleting the deontic necessity operator from a precept or guideline is, by contrast, quite evidently unacceptable. What is left of “You ought to do it,” after we delete the “ought to,” is just that you do it. But it is an uncontroversial observation that people do not do everything they ought, and do things they ought not to.

---

<sup>4</sup>See Anscombe (1997); as a religious Catholic, Anscombe was not herself suggesting that divine prescriptions are to be dismissed, but that is the way her point has been taken by her secular readers.

<sup>5</sup>This view is to be contrasted with the similar sounding but importantly different suggestion advanced in Crary (2007), that ethics has no subject matter. Michael Huemer has suggested to me that positivism about ethics can be characterized as noncognitivism in meta-metaethics; that way of putting it may be helpful to some readers, but we will presently see why, to a positivist, it will count as an unfriendly redescription.

<sup>6</sup>That can be heard the wrong way: the clear-headed metaphysical-stage philosopher intends the *must* to stick to the entailment, not to the conclusion itself.

Mill himself dismisses deductively valid arguments as misinterpretations of a mnemonic device; *his* positive-stage logician determines the conclusion of a correctly performed inductive inference to be true, but does not insist that it *has* to be true (VII:186–193). Millgram (2005a, p. 81, n. 18) is a brief explanation, in contemporary terms, of Mill's reconception of deduction.

<sup>7</sup>Duhem (1985).

The Puzzle of Positive Metaethics, then, is that if metaethics has no subject matter, we can't tell "is" and "ought" apart. In one way or another, there have to be, after all, oughts and something to say about them. Let's designate whatever that turns out to be as the subject matter of positive metaethics. I am going to advance a positivist way of construing and answering the question of what the force of an ought is. I will not claim that my proposal is unique: it is meant to show how positive metaethics *might* proceed.<sup>8</sup>

Let's begin by distinguishing central from penumbral instances of an ought. Suppose someone tells you that you ought to do something, that you ask why, and he provides a response on the order of "just because" or "no whys, you *just* ought to." In that case, you're likely to think that he's bullying you; recall the context of one extreme and memorable pronouncement of this sort, *Hier ist kein Warum*.<sup>9</sup> Or you might take him to be patronizing you, or to be mouthing off; when you think any of those things, you're rejecting the thought that you ought indeed to do it. This tells us that "*just* oughts" are not where we should begin. Perhaps there really are such oughts; it's too early in our investigation to say. But it's not too early to decide that we'll figure out what to say about them after we've figured out what to say about the oughts that come supported by reasons.

How can we approach this latter sort of ought? There is a great deal of literature on reasons that is oddly oblivious of the fact that we philosophers have a method for spelling out reasons, namely, representing them as arguments. (This is the distinctively philosophical use of the term "argument," opposed both to popular usage in which it is more or less a synonym for "quarrel" and to many other disciplinary uses—e.g., that of the poets, in which the "argument" of a poem is, roughly, its *drift*.) To give an argument for *p* is simply to lay out reasons for *p* (generally only some of them), in a way that makes both the component considerations and the relations between them visible. Analytic philosophers, especially, have normally undergone explicit training in taxonomizing and dissecting arguments; while thinking about reasons, we would be foolish not to exploit that investment.

In trying to figure out what a positivist can make of an "ought," we're confining ourselves for now to claims about what someone ought to do that are the conclusions of arguments. Practical reasoning is reasoning about what to

---

<sup>8</sup>Carla Bagnoli has convinced me that constructivism, understood as a metaethical position, could count—once one has allowed that construction procedures are themselves to be the products of construction—as positive in the sense I am spelling out. However, a recent collection (Lenman and Shemmer, 2012) shows that many of those who self-identify as constructivists are continuing to practice old-school metaethics. I will briefly compare the view I am developing to positive constructivism in footnote 32, below.

<sup>9</sup>Levi (1996, p. 29); Williams (2001, p. 87) is exhibiting this reaction when he presents the appeal to irrelevant would-be reasons for action as "bluff." (Irrelevant, he thinks, because they do not motivate their audience, but we don't need that diagnosis for the present point.)

do, as opposed to reasoning about how the facts stand. So an argument about what someone ought to do is practical reasoning. Accordingly, the oughts we are now considering are supported by practical reasoning.

Here's some foreshadowing: Contrast the ways in which ordinary people and metaphysical-stage moral philosophers are going to construe the question, "What's the *force* of the claim that I ought to do such and such?" Metaphysical-stage metaethicists look for an analysis of the force (rather as though it were a sort-of-magnetic field accompanying an imperative, or a special sort of motivational pressure): perhaps it asserts a complicated natural fact; perhaps it derives from a desire or motivation; maybe it states an irreducibly normative fact, one involving some non-natural property of some object. Ordinary citizens, on the other hand, will take the question as a request for the reasons supporting the claim, and so as a demand for supporting argumentation; if no arguments are forthcoming, they will dismiss the demand as having "no force."

Positivist moral philosophers, I want to suggest, should follow the ordinary citizens and construe the claim that one ought to do such and such as (in the central cases) implying that there is a good argument for doing such and such. (In less central cases, there is room for the thought that something on a par with an argument, in ways we can for the moment leave open, is waiting in the wings. We will take up a couple of classes of these cases shortly.) If someone asks you what the force of the claim that you ought to do such and such amounts to, the default answer is to give him an argument for doing such and such (or reasons that could be unpacked into an argument), and when you have done *that*, you *have given* the force of the ought.<sup>10</sup>

---

<sup>10</sup> Analogous remarks are in place vis-à-vis the proper account of terms like "may," "right," and so on; see, for instance, Schmidtz (2009, especially pp. 75, 94f), on "required" and "right." However, the ways in which these terms reflect patterns in argumentation differ.

Scanlon (1998, pp. 96ff) proposes a "buck-passing" account of "good" that is in some ways in the spirit of the present approach. A way of stating it that emphasizes the common ground might run like so: something is good when it has features that make appropriate one or more of a range of responses, where the entailment is understood to be defeasible, and where the appropriate responses may include both attitudes and actions. The proposal has generated a mushrooming literature: see Liao (2010) and Stratton-Lake and Hooker (2006) for a sample.

I'll leave to one side the question of just what the right way of reading Scanlon's account is, in favor of pointing out two ways in which the position the literature discusses—whether it's Scanlon's or not—differs from the positive view. First, the position is not construed as a rejection of old-school metaethics; for instance, there's discussion of the objection that buck-passing defers the reason-giving force of "good" to thick ethical concepts (i.e., concepts which mix the factual and the evaluative) and it is taken for granted that one's account of *these* must be realist, or noncognitivist, or whatever. Second, it treats "good" as a property to be analyzed, rather than an intellectual device whose function has to be clarified; I discuss this contrast further in Chapter 6.

Although the buck-passing account of "good" is in various ways continuous with the account I am proposing of oughts, I am uncertain whether to endorse its nuts and bolts: I am not myself confident that I know how to use the concept "good," but I suspect that one of its homes, maybe its primary home, is satisficing or some relative thereof. (I mean this: when the waitress comes back to refill my coffee for

## 5.3

Before considering further what, in view of the Puzzle, positive metaethics can be, I want to take time out to explain why a self-aware positivism will contrast itself not only with the descendants of Moorean views (on which “good” signifies a non-natural property, “ought” points toward an irreducibly normative fact which functions as a divine command for the godless, and so on), but with noncognitivist and naturalist views as well. These latter today make up the mainstream of contemporary metaethics, and they present themselves as resolutely antimetaphysical.

Let me quickly remind you what these are. *Noncognitivist* views have become progressively more complicated, as they have been reformulated to preempt a long series of objections to them, and recent examples are hard to present quickly. Nonetheless, the earliest members of the tradition give the flavor of the position: emotivism had it that putative moral judgments were merely expressions of the speaker’s emotions, rather than genuine assertions; prescriptivism had it that they were concealed commands; projectivism, that they were projections of the speaker’s emotional responses onto the world.<sup>11</sup>

Turning now to *naturalist* theories: these purport to identify the referents of terms like “good” within the natural world, and to manage something of a piece with that, when, as in the case of “ought,” the terms don’t present themselves as referring expressions. It’s hard to tell just what natural objects or properties are being identified as goodness and so forth by squinting at the available formulations of these theories, but it’s easy to give a toy example: a theory on which being good is being made of heavy cream would count as a naturalistic theory, one on which ice cream turns out to be good, the topping of a strawberry shortcake turns out to be good, and so on.<sup>12</sup>

Now, if positivism is the stage to which a theory proceeds as it emerges from its metaphysical stage, why are these metaethical positions not themselves positive? Consider the postures that might be adopted in a debate about the existence of ghosts. A realist theorist who argues that ghosts are supernatural

---

the fourth time, and I say, “No thanks, I’m good,” *that* is usage that should be taken as central: a way of marking that we have had enough, or that sufficiently many items on a checklist have been gotten to; thus, we need the contrasting “better” and “best,” whereas there are no analogous superlatives of “ought.”) And I don’t see this side of “good” highlighted in any of the treatments of buck-passing I have encountered.

<sup>11</sup> Ayer (1951, Chapter 6), Stevenson (1944), Urmson (1968, especially Chapter 2, pp. 48, 64), Hare (1961), Mackie (1977, Chapter 1). For the most recent (“expressivist”) iteration of noncognitivism, see Blackburn (1998), Gibbard (1990, 2003), and Richard (2008, Chapter 3).

<sup>12</sup> This toy theory would have the merit of providing, at any rate by some people’s lights, a counterinstance to Moore’s Open Question Argument: if it’s made out of heavy cream, how *could* whether it’s good be an open question? For examples of such theories, see Jackson (1998, Chapter 6), Boyd (1988), and Brink (1989).

beings composed of ectoplasm obviously believes in ghosts. But a naturalist theorist, someone who argues, in book after scholarly book, that ghosts are the shaking of leaves in the wind on a dark night *also*, in his own backhanded way, believes in ghosts. (If you don't believe in ghosts, you don't try to say what they *are*.) And a noncognitivist about ghosts, say, a theorist who argues, again in book after scholarly book, that ghosts are merely the projection of one's fear of ghosts onto the branches shaking in the dark, believes in ghosts in rather the same way as does the naturalist. Worse, both the naturalist and the noncognitivist are obviously *afraid* of ghosts: you do not write book after scholarly book unless there is a real fear to assuage. The positivist way with ghosts is short: enough with the campfire stories already. If you have a great deal to say about what ghosts are, you're not a positivist.<sup>13</sup>

A positivist thinking about moral theory is in the metaethics game in *this* sense: he's willing to give you an account of the force of an "ought." But his treatment is orthogonal to the way of framing the topic that is shared by the twentieth-century debate, in something like the way that a number theorist's answer to "What is a number?" is orthogonal to those provided by twentieth-century philosophy of mathematics. (Where the philosopher of mathematics starts in on, say, modal structuralism, the mathematician gives you the Peano axioms, or, rather, directions to an elementary exposition of them.) The positivist treats the request for an account of the force as shorthand for: Back up your insistence that you ought to do such and such with an argument, or a decent surrogate for one. The "force" isn't a *something* that takes supernaturalist or naturalist or even noncognitivist analysis.<sup>14</sup>

---

<sup>13</sup>Is error theory (the view that practical judgments are, one and all, mistaken) a distinct and antimetaphysical position? Is it perhaps a form of positivism itself?

Error theory differs from positivism in its insistence that something has been lost. It is taken to entail that you have no reason to do anything, because you have no *metaphysical* reason; that is, error theory amounts to nihilism about reasons. In a much reprinted chapter of Mackie's *Ethics* (Mackie, 1977), he tells his readers that moral judgments are uniformly false. (This was also our example of projectivism, so maybe error theory isn't a distinct position after all.) In subsequent chapters, he goes on to present the particular moral theory that he endorses (as it happens, a version of utilitarianism), and this strikes most of his readers as a lurch. When a positivist informs you that there are no invisible glows, and then goes on to tell you what you have reason to do, there is no lurch. Thus, if error theory is antimetaphysical, it's in a different sense than positivism's.

<sup>14</sup>Is positivism naturalistic by the naturalists' own lights? I've never met anyone able to give me a satisfactory explanation of the term "naturalism"—Stoljar (2010) is a useful first pass over the difficulties in attaching a thesis worth defending to such a label—and I've long suspected that it takes an emotivist analysis, that is, that it boils down to, "Hurrah for science!" (For second thoughts, see Section 8.3 below.) But for those who think that naturalism does have cognitive content, and thus that the question is well-posed, I have a further question. Naturalistic philosophers exhibit enthusiasm for only the sorts of item that they expect to end up in science textbooks of the future; so, do you anticipate that *arguments* will turn up—not as part of the presentation, but as the purported objects of a science—in the science textbooks of the future? That's not meant merely as a rhetorical question; *logic* might be a candidate for such a science.

## 5.4

A positivist looks to replace forces with patterns. The distinction between what one ought to do and what one does that will resolve the Puzzle of Positive Metaethics requires a pattern found not in what one is now doing, or is presently motivated to do, but somewhere else.

Mill was an instrumentalist, that is, someone who thought that all practical arguments consist in means-end reasoning, in showing that a proposed action is a way of attaining something one desires. But he was also in the business of telling people that they ought to do one thing or another—even when they didn't want to and when no end of theirs would be served by doing what they ought.<sup>15</sup> I'm going to say how I think that Mill thought he had solved the Puzzle of Positive Metaethics, but once again I don't think that we're in a position to nail down any particular reading of the texts on this point.

In a very well-known passage in *Utilitarianism*, Mill wrote that

Of two pleasures, if there be one to which all or almost all who have experience of both give a decided preference . . . that is the more desirable pleasure.<sup>16</sup>

In the literature nowadays, the assessment method being introduced is called the Decided Preference Criterion.

Asking other people what they think of the choices you face is not itself practical reasoning; rather, the results of the poll are inputs to your practical reasoning. The divergence between what you're inclined to do as things are, and what the survey results advise you to do, is discernable (sometimes only potentially). If you are considering seeing two rather different movies this evening, and someone recommends that you change your mind about which one to go to, you can ask what the force of that recommendation is. The Millian response to the request for the force of this sort of ought is, as you no doubt expect, that most of the critics (or perhaps most of the lay audiences) who have seen them both prefer this one. And this reply isn't unreasonable; that's why we ask people for such opinions.

---

<sup>15</sup> Vogler (2001, Chapter 6) helpfully emphasizes this last point.

<sup>16</sup> X:211; shortly thereafter, he adds:

From this verdict of the only competent judges, I apprehend there can be no appeal. On a question which is the best worth having of two pleasures, or which of two modes of existence is the most grateful to the feelings . . . the judgment of those who are qualified by knowledge of both, or, if they differ, that of the majority among them, must be admitted as final. (X:213)

Mill is in the middle of telling you that you can use the preferences of other, more experienced people to establish, in contemporary terminology now, lexical preference orderings. He subsequently uses these preference rankings to argue for such important theses as this one: liberty and certain sorts of personal security must be given absolute priority over other goods in the design of political institutions.



The traditional metaethicist will be wondering where the normativity has gotten to: *Why* should you pay any attention to what other people think or prefer? What's the *force* of the recommendation that people ought to follow *that* policy? And this is a genuinely hard question, because there are many reasons against treating the Decided Preference Criterion as a uniformly good policy. For one thing, suppose that I am faced with the prospects of watching *Alien vs. Predator* or *Rambo IX*; as far as I am concerned, merely having *seen* both of those movies disqualifies a volunteer informant from being a decent source of advice.<sup>17</sup>

However, it is clear what must be, by Mill's lights, the principled positivist answer to the question. What recommends the Decided Preference Criterion is that people who have tried it both ways—who have both listened to the voices of the experienced, and, on other occasions, hardened their hearts and refused to listen—now prefer to take account of and do their means-end reasoning on the basis of the deliverances of those experienced judges.

As far as I know, no one has done a survey to show that this last claim is true, and no one has produced the sort of social-science argument that Mill thought it methodologically imperative to substitute for such surveys.<sup>18</sup> I have by this point shifted into the gray area between a reconstruction of what Mill said and what he should have said, and my own sense is that Mill never pushed his arguments on this topic out to the end. So even though it's midstream, I will in a moment change stalking horses and turn back to what I was earlier calling the central oughts. Before switching, let's just recall that the role which Mill allocated to the preferences of other, more knowledgeable people is typically, in contemporary moral theory, occupied by one's own informed preferences or desires: not those anyone actually has, but the ones that you *would* have, if you were improved in one way or another—paradigmatically, by being better informed.

Now that we've taken a look at Mill's appeal to the preferences of the experienced, and reminded ourselves of the contemporary literature on counterfactually informed preferences and desires, we see that there are at least a couple of categories of penumbral oughts: not the evidently central oughts that are supported by practical reasoning, but not "*just* oughts" either. This is a good occasion to make three points. First, on the basis of my earlier remarks, you might have been anticipating an account of oughts on which they

---

<sup>17</sup>The move is a variant of the Conditional Fallacy, first laid out in Shope (1978). For another thing, doing something on the basis of others' recommendations is often very different from doing it unrecommended. Before the Grand Tour became mandatory for fashionable Europeans, travelers had gone to Italy on their own and encountered Roman antiquities which they found deeply moving. Those who followed the guidebooks written on the basis of the original travelers' enthusiastic recommendations had a very different and far less valuable experience, that of following a beaten path from one must-see to the next.

<sup>18</sup>For a recap of his reasons, see Millgram (2009c, Section 3).

are solely quantifiers over arguments. (That is, “You ought to do such and such” is correctly analyzed as the assertion that there are good arguments for doing such and such.) But heuristic devices like the appeal to other experienced people or to counterfactually improved desires or preferences show that that can’t be all there is to it.<sup>19</sup>

Second, we have seen one relatively straightforward way that the positivist project might pan out in metaethics. There are no ectoplasmic forces, only patterns: in Mill’s case, patterns in other people’s preferences.

Third, we can finally assess Mill’s explicit claim that utilitarianism is positive (and his implicit claim that *only* utilitarianism is positive). A positivist’s oughts point to patterns—in argument, and, for Mill, in other people’s preferences. Mill was an instrumentalist, that is, his official view was that all practical argumentation is means-end reasoning; means-end arguments point you to your ends, that is, to the desires you are attempting to satisfy. The Decided Preference Criterion points you to preferences corrected by the preferences of others. (Mill doesn’t distinguish preferences from desires in anything like the way we do.) But now, utility (or happiness) is Mill’s label for the satisfaction of suitably corrected desires and preferences. Millian utility is accordingly a kind of compendium of the possible drivers of practical reasoning, and so a substantive moral theory that tells you to pursue (and even to maximize) utility amounts to a theory that merely tells you to do what you ought to do, that is, to do what there are arguments for doing: and what could be more positive than

---

<sup>19</sup>There are other reasons that treating oughts as quantifiers over arguments is more complicated than it sounds. Mostly (though perhaps not always) what is meant by an ought is that there is a *good* argument for doing it. And often an argument is defeasible—that is, is a good argument only when there are no defeating arguments in the sheaf. So when you indicate that an argument is good, you’re not normally talking about just *that* argument. However, notice that we can expect anyway nonphilosophers not to have tightened up their views about whether a decisive argument is required, or rather merely an argument that establishes its conclusion *prima facie* or *pro tanto*; this range of options will be relevant when the time comes to consider whether ordinary people can live up to the commitments involved in advancing a claim about what someone ought to do. A more startling option—that you can have good (i.e., decisive) arguments for doing each of two incompatible actions—might allow us to accommodate Williams’s view of tragedy, on which it arises when you ought to take each of two incompatible courses of action (Williams, 1973b, p. 173).

Notice also that treating oughts as quantifiers over arguments allows a positive metaethics to make sense of the debate as to whether oughts are objective or subjective or contextual or relativist; see Kolodny and MacFarlane (2010) for recent discussion of the options. On this way of thinking about oughts, whether they are, say, objective or subjective will be a matter of whether we allow in the domain of quantification only arguments whose premises are accepted by their consumer, or rather arguments with true premises (more generally, premises having whatever counts as the thumbs-up designation for practical argumentation), regardless of whether they are accepted by their consumer.

Likewise, topic-restricted oughts are sometimes contrasted with “all things considered” oughts; I expect these contrasts can also be construed as restrictions on the domain of quantification. That you legally ought to do so-and-so implies that there are good legal arguments for doing so-and-so; that you prudentially ought to do so-and-so, that there’s a good prudential argument. None of these suggestions address the substantive question of what to decide when several arguments, perhaps of different types, point in different directions (in an older vocabulary, when different types of oughts conflict).

that? However, Mill's understanding of utilitarianism as positive evidently depends on his background view of the mechanics of practical rationality.<sup>20</sup> If we are more open-minded than he was about what practical inference can be, we will be unlikely to accept that positivism commits you to utilitarianism.

The penumbral oughts we've identified point us not to arguments, but to, as it turns out, shortcuts. For both types of penumbral ought, it's clear enough, first, where the patterns that a positive metaethics relies on are to be found. Second, it's clear enough for the Millian version (and although I won't argue it here, the point goes as well for the contemporary alternative) that shortcuts are important, but they have their limitations, and it's essential to be aware of what they are. The Millian strategy of relying on the preferences of others is sometimes an effective heuristic, but sometimes it's not a good idea at all. For instance, when you do rely on somebody else's preferences, you're free riding on his cognitive investments; not everybody can always be a free rider, and whether the strategy is a reasonable one depends on the quality of those investments, from your point of view. You need to ask yourself whether you are merely harvesting other people's brute responses, responses that may have very little to do with what makes sense for you, or whether these are the results of careful and experience-driven deliberation, results to which you should be giving a good deal of credit.<sup>21</sup>

Mill's use of his Decided Preference Criterion amounts to an example of a positive metaethics. As I will shortly explain, I think that the view of practical reasoning that underwrites it is unsatisfactory, but the technique is nonetheless likely to be a (penumbral) part of a correct positive metaethics.

## 5.5

Let's return now to those central, argument-supported oughts and to the prospects and methods of positive metaethics. Is positive metaethics merely a bit of belated Enlightenment housecleaning, or does it have substantive content and a promising agenda? Consider some positivist responses to a series of objections on the part of a traditional metaphysician of morality.

"The job of metaethics isn't to say what this or that particular ought amounts to, but rather to explain to us what oughts and their relatives amount to *in general*. What you have labelled positivism is really just quietism, a way of sidestepping the philosophical and metaphysical task of accounting for normativity—for the all-important difference between what *is* done, and what *should be* done."

---

<sup>20</sup>To fill in steps of this very terse recap, see Millgram (2005a, Chapter 2, but see also pp. 13–16 for caveats).

<sup>21</sup>For discussion of when counterfactually informed preferences are a useful heuristic, see Chapter 2 in this book.

It is just here that the turn to practical reasoning makes positivism a philosophically empowering doctrine. In the central cases, an ought amounts to a gesture at a practical argument. It is the job of the theory of practical reasoning to survey and analyze the different forms that practical arguments can take, and here are a couple of examples.<sup>22</sup> Some practical arguments are instrumental: they tell you to do such and such as a way of attaining a goal. Other practical arguments are inductive: they tell you that the lesson of your practical experience is that such and such matters or is important. A systematic theoretical treatment of the different sorts of reasons—that is, the different forms or modes of argumentation—that can be brought to bear on the question of what to do amounts to a general philosophical account of oughts and their relatives.<sup>23</sup>

In my own view, it is in great part the inattention to the varied forms of practical argumentation that made the metaethical back and forth over the last century contribute so very little to our philosophical understanding of ethics. Positive metaethics means no longer just spinning our wheels, because it is redirecting our attention to the real materials of practical thought: in the first place, the different forms of practical argumentation that might be deployed in the service of a claim that you ought (or ought not) to do something.

That inattention also explains why positivism has been so long overlooked by metaethics. An astonishing number of philosophers have assumed—“assumed,” because there is almost no actual argument for the view—that practical reasons are exclusively instrumental, that is, means-end. The instrumentalist rendering of practical rationality normally takes one of two forms. On the majority view, ends are set by desires, and all reasons for action boil down to: “*I wannit!*” With such an impoverished and in the strictist sense childish view of practical reasons, there isn’t a lot in the way of practical argumentation to work with.<sup>24</sup> Where you’ve gotten rid of the practical argumentation, there’s nothing left for the force of an ought to be but a sort of magnetic field. And where you haven’t, practical argumentation that bottoms out in a desire—in “*I wannit!*”—leaves you wondering what there is to a desire that makes an appeal to one count as an *argument*. The upshot is that once you’ve gotten rid of all but instrumental practical argumentation, there’s nothing for you to theorize about but the peculiar sort-of-magnetic

---

<sup>22</sup>For an overview of a number of them, see Millgram (2001).

<sup>23</sup>We inherit Aristotle’s logical hylo-morphism, on which all the work done by argument is done in virtue of its *form*. If you believe that not all successful argumentation can be understood on the hylo-morphic model, then you should also think that, however useful a taxonomy of the forms of practical argument turns out to be, it will not cover all of the territory; and then there will be the question of how a positivist is to make sense of the unformalizable remainder, which I will defer to another occasion.

<sup>24</sup>I’m not the first to have noticed this problem with instrumentalism. Dewey (2008a, p. 208) remarks, “Only a child in the degree of his immaturity thinks to settle the question of desirability by reiterated proclamation: ‘I want it, I want it, I want it.’”

field induced by a desire—or alternatively, the peculiar sort-of-magnetic field induced without a desire.<sup>25</sup>

“But.” (The traditional metaphysician of morality isn’t done yet.) “When you produce, or even only gesture at, the arguments for doing something, you’re implying that one *ought* to accept their conclusions and act on them. What about *that* normativity? What’s the force of *those* arguments? You haven’t answered the real question, only pushed it back a stage. And if you haven’t, your view still is just another philosophical quietism.”

Because John Stuart Mill did his very best, in his philosophy of logic and science, to adhere to the strictures of the British Empiricist tradition in which he was raised, and thus also, where they overlapped, to the strictures of positivism, he can serve as our guide at this point. Humean skepticism about causation was out of place, he thought, because there are inductive arguments that such and such causes are followed by such and such effects. And where Hume pressed the question of what force could underwrite such an inductive argument, if not causal necessity, Mill responded by gesturing at a grand inductive argument, to the effect that inductions have been observed to work in the past, and so they will continue to work in the future. This is not the place to explain why Mill thought the position he was spelling out to be other than viciously circular.<sup>26</sup> What I want to emphasize now is that Mill is showing us how, from the positive point of view, the right response to the demand for the force of some class of arguments is normally a further argument, to the effect that you should take arguments in that class seriously—that is, accept their conclusions.

Let’s use Mill’s treatment of the problem of induction as a model for a positivist response to the traditional metaethicist’s second objection. The force of practical arguments of one or another sort is to be made out by giving a further argument (a practical argument, no doubt) to the effect that one should deploy and act on arguments of that sort. Such meta-arguments might take the form of practical inductions: we have learned from experience that such and such a sort of practical argument is a good idea. They might take the form

---

<sup>25</sup>Smith (1987) is a well-known discussion of the former sort. Williams (1995a, Chapter 3) is an attempt to work with the very minimal materials provided by instrumentalism, and one which exhibits the distortions to which the restriction gives rise.

The minority view of instrumental rationality holds that such reasons for action simply point out that the action is part of a larger action which you are already in the course of performing, as when someone asks you why you are getting in the car, and you answer that you are going downtown. (Vogler 2002 and Thompson 2008, Part 2, are two recent—but not exactly instrumentalist—variations on the Anscombian view; for the qualifications, see Millgram 2006b, 2009e; Andreou 2006 is fully instrumentalist, but frames the conception of instrumental rationality in terms of intentions.) On this sort of view, instrumental reasons for action boil down to: “*This* is what I’m doing!” Here also there isn’t a lot in the way of practical argumentation to work with; and why does the expression of one’s dogged persistence—since the bottom line is an “*I am* doing it!”—count as an argument at all?

<sup>26</sup>For that part of the story, see Millgram (2009b).

of transcendental deductions: deploying practical arguments of such and such a sort is a necessary precondition of agency, or of figuring out what to do. They might take the form of specificationist arguments: when we engage in the practical task of specifying more concretely the abstractly described goal of figuring out what to do, what we come up with is argumentation of such and such a form. Or one might try to give them a Kantian form: a demonstration that you cannot will the universalization of the maxim, “I will not reason using such arguments.”<sup>27</sup> These meta-arguments are, however, unlikely to consist in demonstrations that deploying arguments of such and such a form will help us attain our goals. The target argument type has to be understood as legitimate, and “I *wannit!*” is just about never a good way of establishing *legitimacy*. This is a further manner in which instrumentalism has impeded the progress of positive metaethics.

One more objection from, and response to, a traditional metaethicist. “Look, it’s all very well to complain that my conversation partners and I have been arguing about ghosts, but we were trying (for an entire century!) to give an account of what oughts *mean*. You owe us that account; it’s no good to tell us that oughts are the conclusions of practical arguments if we don’t know what oughts are telling us, or if they’re even telling us anything!”<sup>28</sup>

---

<sup>27</sup>For practical induction, see Millgram (1997); for a sample transcendental argument with this sort of conclusion, see Millgram (1997, Chapters 3–4). A transcendental argument for a rather different and Kantian conclusion about practical rationality can be found in Korsgaard (2009). For an overview of specificationism, see Millgram (2008); for exposition and critical discussion of the Kantian CI-procedure, see Millgram (2005a, Chapter 3).

<sup>28</sup>However, not all traditional metaethicists focus on content analyses, and in particular, some expressivists have had second thoughts about it (e.g., Blackburn, 1998, p. 85).

A great deal of contemporary metaethics consists in debate over what sorts of mental items make up the trains of thought in which practical reasoning consists. Surely a positivist owes answers to these questions, and when he tries to supply them, won’t his position turn out to rest on metaethics of the traditional sort? For instance, if he holds that the trains of thought consist of beliefs and desires, won’t the argument for that claim be that we can’t make sense of mental states that represent moral facts? And won’t that argument presuppose a refutation of moral realism or, anyway, take moral realism on its own terms, as an alternative to be refuted? (I’m grateful to Benjamin Kiesewetter for this objection.)

For what it’s worth, I don’t think we have a way of characterizing the psychological stages of a stretch of practical reasoning, independently of and prior to figuring out how arguments of its type go. Recall how the logical positivists (not the positivists, now) thought that you could make one and the same philosophical claim in startlingly different vocabularies. For instance, if you said that there were such things as sense data, you would be speaking in the “material mode”; if you said that our language contained sensation terms, you would be saying the same thing, but in the “formal mode.” Most philosophical talk of the “mental states” that figure into practical reasoning is what we might as well call the *psychological mode* of expression of views about what counts as a legitimate form of argument; talk about values and whether or not they are real is the *material mode* of expression of those same views. We read what we’re going to say about the “mental states” off the slots that a particular form of argument (or the particular forms that intersect at a slot) provide. And this is what you’d expect of a positive metaethics: the investigation of forms of thought that appear in practical deliberation will be conducted by investigating what legitimate patterns of practical deliberation are.

Certainly if some deliberative move cannot be cognitively implemented, that may constrain our theory of deliberation: if your theory of deliberation calls for a Halting Problem oracle, or an NP-completeness oracle, then it’s irrelevant to human decision making, and no doubt there are many

In my own view, the semantic analysis of ought and its relatives comes last, both in the order of priority and in the order of explanation. You know, philosophers are often historians of philosophy as well, and when they are, they learn from their professional experience that the best way to figure out what an obscure claim or term means is to see how the dead philosopher one is reading uses it in argument. Since a conclusion of an argument must mean something that follows (or would have seemed to follow) from its premises, seeing what arguments the dead philosopher gives for it normally tells you what it means; seeing how that conclusion is used in further ongoing argumentation likewise tells you what it means. Here's a method of semantic analysis that *works*: why abandon it when we come to moral subject matter, and practical subject matter generally? The semantics-first approach has had quite long enough to convince us that it's going somewhere, and what we see now is a century's worth of progressively higher-entropy reformulations of a handful of very tired proposals. It's time to cut this short, and try it some other way—the positive way.<sup>29</sup>

The positive metaethicist, then, can respond to the request for a general philosophical account of the force of an ought, in the first place with a taxonomy of the different varieties of practical argumentation which might support it. Second, he may also respond with a taxonomy of heuristics and shortcuts, such as informed desires, or surveys of other people's preferences, which can be used to supplement practical arguments. Third, he may respond with further practical arguments, to the effect that one ought to deploy and act on the varieties of practical argument adduced in his first response, and to the effect that one ought to use heuristics and shortcuts enumerated in his second

---

similar restrictions. But such implementability constraints have nothing in common with the pronouncements about what our mental states are that get made in the register of traditional metaethics.

<sup>29</sup>However, here's an objection to register and to take seriously: we philosophers are much more expert and have a great deal more control of arguments that are, roughly, truth-preserving than we have with practical argumentation, that is, argumentation whose ultimate conclusion is about what to do.

Wedgwood (2007) shares a roughly Brandomian approach to understanding oughts, by which I mean that he attempts to make sense of oughts in terms of the inferential commitments involved in advancing one; this is a good occasion to mark some of the ways in which our views differ. First, and borrowing vocabulary from Millgram (1997), Wedgwood focuses on forward-directed inferential commitments, where I have been pointing towards backward-directed inferential commitments. That is, he is concerned with what the ought commits you to doing next (namely, acting), where I am directing our attention to the arguments that support the decision to act. (Belated acknowledgment: when I published *Practical Induction*, I had not yet read Brandom 1994, which should get the credit for introducing inferential commitments into the current debate.) In a response to an objection which he attributes to Foot, Wedgwood (2007, p. 106) shies away from deploying backward directed inferential commitments and when he attempts, at another point, to tie different senses of "ought" to different types of reasoning, he picks out types of reasoning by the characteristic mental states in which they terminate, rather than by differences in the inference patterns that lead up to those states (Wedgwood, 2007, pp. 120ff; see especially p. 124). And second, of course, Wedgwood goes on to address the traditional metaethical questions, rather than do the positive thing, which is to dismiss them.

response.<sup>30</sup> If the traditional metaethicist replies by asking for an additional and very different sort of account of the “normativity” of an ought, one that won’t accept a practical argument as a response, he really has lapsed back into superstition: he is someone who believes in ghosts and is insisting that you tell him what a ghost *is*.<sup>31</sup>

## 5.6

The turn to the practical argumentation that is invoked by central rather than penumbral oughts is a solution to the Puzzle of Positive Metaethics: the account of the force of an ought is to be found, not by purporting to analyze ghosts, but by investigating the forms of practical reasoning—that is, patterns which can be taken by practical argumentation. It is also a promising research agenda: where obsessively reanalyzing the nonexistent ghost of an ought will improve neither our philosophical understanding of our choicemaking nor the choices themselves, an investigation of the patterns of argument legitimately used to support choicemaking will plausibly do both.

---

<sup>30</sup> Are we done? Notice that we have not yet taken up those “*just* oughts.” Nor for that matter have we considered putative oughts supported by particularist arguments, that is, arguments whose force does not survive a shift of context. (For an overview of the particularism debate, see Hooker and Little 2000.)

If positivism is committed to making sense of oughts by anchoring them in patterns, then a positive account of particularist reasons must also appeal to patterns. One familiar characterization of particularism has it that reasons of this type can’t be understood simply as the applications of general rules. If we allow the characterization, a positive account of particularist reasons would turn on appropriate patterns that do not conform to general rules. Whether there are such patterns and, if so, how we work with them are questions I don’t want to start in on right now.

I will turn to those “*just* oughts” in Chapter 6. For now, let me parry a Moorean reason for moral philosophers being willing to accept “*just* oughts” as a sort of bottom line. Our conviction that one ought to do such and such is often visibly more secure than any of the arguments for so doing; philosophers too easily assume that this sort of security must be due to the foundational role played by their moral intuitions. But if oughts amount to—again, in the central cases—quantifiers over arguments, then a practical conclusion may be supported by many arguments, in which case the conclusion can be expected to be more secure than any one of those arguments, or any step in one of those arguments. (Compare Wimsatt 2007 on multiple derivations in the sciences.)

<sup>31</sup> Blackburn (1998, p. 295) complains about Dworkin’s rather different attempt to turn away from metaethics (Dworkin, 1996) that noncognitivist treatments “are not on the face of it *moralizing*”; whatever Dworkin may think, there is obviously a difference between first-order moral theory and what Blackburn and his interlocutors are doing, and therefore it is illegitimate to insist “that there is no such thing as metaethical thought at all.” As it stands, Blackburn’s objection to rejecting metaethics in favor of “internal, first-order moral questions” begs the question: someone who thinks that what metaethicists have been doing is philosophically worthless will not regard their alleged accomplishments as evidence that there is a task there to be accomplished. However, notice that the distinction that Blackburn is appealing to is saved by the present account; there *is* a difference between first-order argumentation and second-order argumentation that takes as its subject matter the efficacy and point of one or another type of first-order argumentation. The point overlooked by Blackburn is that although the latter is not necessarily moralizing, it is *formally* of a piece with what we do when we moralize.



But before I take my leave, I want to do another pass over one of our traditional metaethicist's complaints. "Suppose that we accept, for the sake of the discussion, that the central oughts are to be explicated in terms of patterns of argumentation. Still, that argumentation cannot be merely those arguments that someone has actually advanced in public, nor even the arguments that someone has gotten around to thinking up. Rather, he must have in mind the *possible* arguments, and the relevant notion of possibility is itself normative: it counts as a possible argument if you *ought* to draw the conclusion! The oughts haven't been discharged, and if positive metaethics isn't viciously circular, it presupposes old-school realist metaethics, or one of its old-school alternatives."

And he might continue: "It's not the traditional metaethicist who is superstitious, but *you*, the positivist. Imagine someone who thinks that prepending a ritual inscription to an ought makes it be the case that you indeed ought to do it. And imagine that when you ask him why his ritually required orthography and arcane calligraphy have that effect, his response is to produce more oddly spelled and peculiarly written inscriptions. Isn't this just *magic*? What difference could it make to then be told that the ritual form of the inscription is numbered steps, some of which are labeled as 'premises,' and others of which are followed by incantations such as 'universal instantiation'? What difference could it make that the inscriptions contain such exotic glyphs as ' $\wedge$ ' or ' $\supset$ ' or ' $\exists$ '? That is, what difference could it make if we don't know, ahead of time, what makes those inscriptions into an *argument*?"<sup>32</sup>

---

<sup>32</sup>The objection is adapted from Hussain and Shah (2006); see also Hussain and Shah (2013). They take their foil to be constructivism, and I need to explain the differences between constructivism and the view I have been developing.

A glance at an anthology I have already mentioned (Lenman and Shemmer, 2012) will suggest that "constructivism" has in recent years come to mean almost all things to all people; there is remarkably little focus on the important ideas that Rawls introduced a quarter of a century ago, and a great deal of recycling of the usual suspects: coherence, Aristotelian phronesis, and so on. So we first have to say what *we* mean by it. Consider the status of being the winner (or loser) of a game of chess; this status emerges from executing a "construction procedure"—in this case, from playing the game out to its end—and there is obviously not much room for the sort of debate over the metaphysical status of someone's being the winner that typified old-school metaethics—that is, over whether you are the winner in virtue of possessing a non-natural property, or in virtue of attitudes of approval, and so on. (Whoever won the game is the winner, and that's that.) To be a constructivist about oughts would be to take oughts to emerge from an analogous construction procedure. (The clearest discussion is still Rawls 1989, where the technique in question is the Kantian "CI-procedure.")

A construction procedure should be uniform and well-defined: not Calvinball—"never the same . . . you'rre maaakinnngggg iiit uuuuup aaas youuu gooo" (Watterson, 1996, p. 101)—but something you can step through. If you take the outcome to be determined by how the procedure *would* turn out, as opposed to how it *did*, you will need an *especially* well-specified procedure, one that supports counterfactuals of this kind. Such procedures must be uniform: roughly, you do it the same way every time, for everybody, in the way that, whenever people play chess, the rules, the starting position on the board, and so on are the same. And it's only a procedure if, after you walk through it, it's *over* (and it's clear *when* it is). Now, recall that existentially quantifying over arguments doesn't commit you to being able to produce them, and note that "finding arguments" isn't a procedure, or anything like one: we're not

There is a real problem here, but we need to get it clearly in focus before we can consider how a positivist is going to address it. The problem is not circularity: while it is true that on a positivist account, an ought may amount to insisting that there is a good argument for doing such and such—that is, an argument whose conclusion you *ought* to draw—the contents of those two oughts are not identical: the first one commits its user to there being good arguments for doing such and such, whereas the second one commits its user to there being good arguments for drawing the conclusions of the arguments picked out by the first ought. The arguments being existentially quantified over need not be the same arguments, and it is not viciously circular to appeal to *further* arguments.<sup>33</sup>

The problem is also not that you had better be able to delineate the oughts without giving arguments; that would amount to a variation on the back-and-forth in Lewis Carroll's famous parable of Achilles and the Tortoise.<sup>34</sup> Spelling that shorthand out a bit, the traditional metaethicist had better not be adopting *this* posture: presented with an argument (*any* argument), he consistently responds, "So far, that's just an *inscription*; I'll accept it as an argument only after you show me—give me an argument—that it is one." If he does adopt such a posture, he will not accept any response a positivist might muster, but he also will never be able to nail down the sort of traditional theory of normativity that he hopes for and insists on (because surely we are supposed to

---

in a position to point to the arguments we *would* find, if we went looking. Different people on different occasions will go about hunting for arguments in very different ways. We don't yet have so much as a catalog of all the forms a practical argument can take; so there's no way to marshal all of the practical arguments there are in any systematic way. And procedures *terminate*, but when you're assembling arguments for and against a claim or a decision, there's nowhere definite that you have to stop.

<sup>33</sup>You might think that the conclusion of a practical argument is not that you're going to do it (a decision), but that you *ought* to do it; in which case, how can an ought tell you that there's an argument for doing it? You *might* think that, but using oughts to tag the conclusions of practical arguments is like announcing the conclusion of a bit of theoretical reasoning you have just given to be true: just as this sort of merely redundant use of "true" is allowable but dispensable, so the use of ought to mark the conclusion of a practical argument that you have just given is allowable, but (normally) dispensable.

Here's what that "normally" marks. You can draw a practical conclusion but not act on it; that's the observation that generates the age-old philosophical puzzle of *akrasia*, or weakness of will. So practical conclusions can't always be decisions or intentions or (as Aristotle famously suggested) the action itself. An *akratic* is best construed as drawing a practical conclusion, but not one that's as *final* as that; an ought—which the *akratic* can use to acknowledge that there's a conclusive argument for doing it—can serve as such a conclusion.

It doesn't follow, however, that the conclusion of *every* practical argument must already contain an ought. (The weak-willed need fancier ways to think their waffling conclusions than more straightforward folk, and that shouldn't be surprising.) It's also worth reminding ourselves that there are likely to be other forms that not-yet-final practical conclusions can take. I've picked oughts as a representative metaethical device, and focused the present investigation on them, but evaluative concepts—metaethicists love to discuss "good"—can also be brought to bear in evaluative judgments that seem often to function as not-yet-final practical conclusions, i.e., steps arrived at in the course of a practical argument that are not yet the final determination to act.

<sup>34</sup>Carroll (1895); Carroll was the pen name of the Oxford logician Charles Dodgson.

accept whatever the right metaethical theory is on the basis of an argument). So we must presuppose that we are able to *give* arguments.

Of any argument that you *do* give, the demand that you show it to be an argument may well be in place; that demand gets addressed by a further argument, if you can think of one. And that's fine, because we're now allowing that we're able to argue. However, that *each* such demand can be so addressed doesn't entail that there's *one* response that addresses all those demands together: the demand to exhibit what makes *anything* an argument is off the mark. The thought that first of all we need to know what "normativity" is involves the unsustainable assumption that there is *one* answer to the question; what makes something an argument whose conclusion you should draw? (Actually, of course, it rests on an even broader assumption—that there is *one* answer to the question; What makes an action something you should do?) Arguments differ, and what makes it be the case that you ought to draw their conclusions also differs, rather dramatically, from type to type.<sup>35</sup>

---

<sup>35</sup> An updated version of the objection we are considering, found in Hussain (2012), presents a variation that we are now in a position to disentangle. (I'm going to change out his talk of construction procedures with the relevant notion here, that of argumentation.) One strand is the worry that there is likely to be a symmetry between good arguments that such and such is a good argument and bad arguments that what is in fact a bad argument is a good argument. The other strand is that, if all we have is regresses of arguments that such and such is an argument, "at the most fundamental level there is nothing to make it the case that one normative claim is correct rather than another" (Hussain, 2012, p. 193).

Hussain seems to think that presenting these worries side by side makes them more pressing, but I don't see that we should take them that way. On the one hand, it is a familiar point that there are such symmetries, and you have to live with them whether you are a positivist, or a constructivist, or you have some other old-school metaethical view: as Salmon (1974, p. 56) long ago reminded us, if you are willing to affirm the consequent, it is easy to argue that affirming the consequent is a valid mode of inference. Insisting that there is something that *makes modus ponens* valid, and that *makes* affirming the consequent invalid, is perhaps an expression of conviction and of frustration, but it is not a way of telling which is which: only a good argument will do that. ("Really, *really*, *REALLY*!" is never a good response to a skeptical worry.) And because a good argument is, in the first place, something whose conclusion you *ought* to draw, and because oughts are, in the first place, existential quantifiers over arguments, it does seem to me that the closest thing we have to something that makes so and so an argument is: another good argument. (However, because these arguments can be expected to be substantively as different as you like, from occasion to occasion, we should not expect there to be any *one* thing that makes *any* one "normative" claim rather than another correct.)

On the other hand (and this is warmup for ideas I'll develop less imagistically in the main text), the notion that at the most fundamental level we need a brute metaphysical fact to make a "normative" claim true is on a par with the thought that drives confused lay economists to insist on the gold standard: that there has to be something that's really, intrinsically valuable—*gold*—to make banknotes into money (into something of value), as opposed to worthless paper. Thus, old-school metaethicists appear to positivists as rather like monetary metaphysicians who think we need, and try to supply, a theory of the intrinsic value of gold.

In fact, banknotes are money largely because other people will accept them as money, and those people are willing to do so because they expect that further people, in their turn, will accept the bills as money. Is this a regress? How could anyone know that, in a million years, American dollars will still be accepted as currency? On the contrary, a long track record of defunct currencies tells us that, in much less time than a million years, the dollar will no longer be in use. And that does not stop anyone

Continuing to leave penumbral oughts to one side, allow that telling someone he ought to do something is committing oneself to there being a good argument for doing it. Allow that to be a good argument is to have a conclusion that one ought to draw. And allow that it follows that in telling someone he ought to do something, one further implies that there is an argument to the effect that you ought to draw the conclusion of the initial argument. What will that second-order argument look like?

*Sometimes* the further argument, that you should draw the conclusions of an initial argument, appropriately consists in spelling out the initial argument more clearly. However, if that is all you ever do, your further arguments turn into redundant table-thumping.<sup>36</sup> If you have an instrumentalist's impoverished sense of what forms of argumentation are available, and so if you give practical arguments that are all of the same form, you are likely to feel that a general account of the "normativity" of a practical argument is a reasonable request, and even that a general account of "normativity," plain and simple, is a reasonable request. After all, those reasons all seem the same to you, and why shouldn't there be a homogeneous account of a homogeneous phenomenon? But in this case, the charge of vicious circularity, and the charge that your arguments are no more than magic inscriptions, is entirely in order.

Those further arguments, to the effect that the arguments gestured at by your initial ought really *are* arguments, must be substantive additional contributions: by and large, they must be *different* arguments. So far, so good; my experience is that, when you go looking, there are often interesting arguments to be had about what counts as a legitimate form of practical argumentation. But there is a regress pending, and now we are getting to the real problem at the bottom of the objection.

The regress is problematic because it seems to involve a commitment on the part even of ordinary users of ordinary oughts to the availability of an endless number of very complicated arguments about what counts as a good argument for being a good argument. How can they be so sure? That is, the concern at the heart of the objection is that the view I have been developing makes out almost anyone who deploys an ought to be full of hot air and talking through his hat. It is this concern that most charitably explains the turn to superstition

---

from spending their dollars (even though if we knew that next week the dollar would be defunct, everyone would stop accepting dollars as payment right now). Our requirement that a currency be a going concern, while apparently involving the expectation that the currency will continue to be in use indefinitely, does not commit anyone to insisting that it will *always* be in use. When it comes to money, regresses of value don't have to be stopped by bimetalist doctrine; and when they are not stopped, they don't turn out to be full-fledged infinite regresses. Why shouldn't the regress of arguments we are now considering behave like *this*?

<sup>36</sup> Compare Robert Nozick complaining about the sort of back-and-forth that tends to characterize discussions of Newcomb's Problem: "it will not do . . . to just repeat one of the arguments, loudly and slowly" (Nozick, 1997, p. 48).

in metaethics that so bothered Mill.<sup>37</sup> On the one hand, our survey of the various oughts tells us that they mark a course of action as supported; in the central case, supported by argument. On the other hand, the upshot of the regress we're considering is that, even when you're in a position to spell out the first round of arguments, you're never in a position to spell out all the argumentative or other support that the ought tells you—once you squint at it a bit—has to be in place. In circumstances like these, philosophers easily get confused and are prone to reify demands and commitments into facts: here, a fact which *inexplicably* provides the impossible-to-spell-out support. Such a philosophical discussion too often turns into the attempted analysis of the mysterious fact, the one that might as well be an invisible glow.

But the misstep is motivated by a mistake—albeit a very natural one—that of assuming that the scope of a commitment can generally be read off its overt content. Here are a couple of related applications of the distinction I'm making, just to give you a sense of how I mean it. If you believe that *p*, you are taking *p* to be true, and then you are inferentially committed to the entailments of *p*. However, if you advance *p* as a simplification or an approximation

---

<sup>37</sup>It's unsurprising that, when it comes to arguments, philosophers look for intrinsic features of the argument—most prominently, its form—that will certify its correctness. This variant of the turn to superstition consists in imagining a substrate whose possibilities of configuration will determine what the possible (that is, correct) arguments are. A variant of this way of thinking: you want a notion of “possible argument” that includes even the awful arguments, so that you can come up with a way to distinguish the bad ones from the good ones. In that case you also need a medium whose combinatorial possibilities will give you all of those “arguments,” rather in the way that the alphabet determines what books are to be found in Borges's “Library of Babel” (Borges, 1999).

But there is no *single* medium in which all arguments, practical or otherwise, can be exhibited. I am, on and off, convinced that the intense interest taken in propositions by twentieth-century philosophy of language was due in large part to thinking that there must be *one* medium in which *all* arguments can be realized. Propositions were abstract objects with the look and feel of metaphysics in the old-school and pejorative sense: they behaved like what architects call gleepsite and engineers call unobtanium. Like the impossible product advertised in Tom Waits's track, “Step Right Up” (Waits, 1976)—which slices and dices, entertains visiting relatives, walks your dog, disinfects and sanitizes, delivers the pizza, mows the lawn, finds you a job, makes excuses for unwanted lipstick on your collar, and much else besides—propositions had too many jobs to do. As per usual, the search for the impossible object arose from mistaking requirements (in this case, requirements that change from dialectical occasion to dialectical occasion) for a *thing*.

In fact, there are, as far as I can tell, indefinitely many kinds of argument, and different media are suitable for different forms of argumentation. Some arguments really can be thought of as patterns in (say) counterfactual mental states and, in fact, to the extent that they depend on unformalizable judgment calls, really have to be construed that way. Some arguments (especially the deductive and formalizable ones) can be represented as idealizations of inscriptions. Some arguments require diagrams, or, as in Plato, dialog (complete with tone of voice). There are indefinitely many media in which different sorts of argument might be realized, and each has its pros and cons: features which suit a medium for one sort of argument rather than another. The media in which arguments are conducted are not fixed; we invent and fabricate new ones. For instance, when Cora Diamond complains about a very crude form of cookie-cutter argumentation prevalent in moral philosophy, she lauds the alternative as “anything but argument”—the title of a paper in Diamond (1991). In fact, she is developing a new and very interesting form of philosophical argumentation, and a novel prose medium in which to conduct it.

or an idealization, then you don't take *p* to be true, but rather true *enough*; you don't believe it; and your competence in deploying this sort of partial truth is in large part a matter of knowing when *not* to draw apparently entailed conclusions.<sup>38</sup> In this case, you can't just read what you're inferentially committed to off that propositional content, *p*. Or again, Margaret Bowman has argued that the point of having many of our long-term plans is not to execute them; if we are at all self-aware, we expect to abandon most such plans somewhere in the middle, and we will not regard that as any sort of failure. (Little boys and girls plan to become firemen and ballerinas when they grow up, but usually they don't, and that's just fine; we're *all* like little boys and girls in that respect.) The cognitive function of such plans is to guide choice in the here and now, not to bring one to attain the goals at the farther reaches of the plans. When you go into a job interview or an oral examination—though these are shorter-term examples of the sort of thing Bowman has in mind—you are well-advised to work up an agenda, that is, a plan for how the conversation will go, in advance. You may fully expect to be derailed in the course of the conversation; in fact, if the interview or exam goes well, it will not look anything like the plan you made for it. Nonetheless, the guidance provided by such an agenda, even and especially as you cope with unexpected swerves, can be invaluable.<sup>39</sup> When it comes both to partial truths and to Bowman's "aspirations," the explicit content (the goals specified by the plan, the sentence used to express the true-enough approximation) outruns the commitments of the user.

If I am right, oughts exhibit a similar sort of divergence of commitment from content. The regress we were worrying about is a way of spelling out what the content of a central ought amounts to. But the commitments of non-PhD-bearing laypersons (and even of reasonably alert philosophers who specialize in practical rationality) give out long before the end of the regress: my own experience is that if you get those laypersons to embark on the regress, and then ask how they know the arguments won't run out at the *n*th stage, they just *shrug*. They are willing to assume a commitment to there being support, for some indefinite stretch of the way out; they hadn't immodestly meant to assure anyone that the arguments would turn up *all* the way out.

When it comes to practical matters, the real question is normally not one of fact, but rather about what to do.<sup>40</sup> Here we're dealing with a practice, one of saying what you have reason to do. Experience with the practice of argument might lead philosophers, especially, to develop a tentative but reasonable confidence in their arguments, accompanied by a characteristic

---

<sup>38</sup>For further description, examples, and supporting argument, see Millgram (2009a).

<sup>39</sup>Bowman (2012).

<sup>40</sup>I'm adapting the following move from a manuscript by Sarah Buss, "Against the Quest for Normativity." I'm very grateful to her both for allowing me to read it and for conversation; of course, she's not responsible if I've misrepresented her ideas.

phenomenology.<sup>41</sup> Perhaps the phenomenology sometimes leads people to overreach, and make a habit of insisting, dogmatically, that you ought to do this or that: I mean, in a manner that commits them to a panoply of supporting arguments in whose existence they really shouldn't have their sort of confidence at all. But in that event they *shouldn't* insist, and many of us don't. The reasonably thoughtful rather exhibit a distinctive sort of tentativeness in their pronouncements about what one ought to do; this consists in a willingness to retract their judgment when it turns out that the arguments for it are not really very good arguments after all, together with a lively sense that turning out that way isn't impossible or unprecedented.<sup>42</sup>

## 5.7

Heuristics, shortcuts, and perhaps other exceptions to one side, oughts point us to arguments.<sup>43</sup> And that means that once we put the superstitious obsession with the ghost of an ought to one side, the core of metaethics—positive metaethics—is, in a likely and promising incarnation, the theory of practical reasoning. Metaethics as the theory of practical reasoning *is* much more promising than old-school, superstitious metaethics. An alleged analysis of supernatural magnetic fields can't improve your thinking or your ability to get around in the world. But positive metaethics can improve your reasoning

---

<sup>41</sup>As with much phenomenology, the phenomenology of argument is most naturally described imagistically. Someone who encounters a rope descending out of the fog can come to be pretty sure that it is tied to something, even if he cannot see the top of the rope. If it doesn't fall of its own accord, if he can tug at it and it stays more or less where it was, if he can climb it, and so on, he will eventually come around to the working assumption that it's not about to just come down. (And that's likely to be true whether he thinks it's attached at the top to something that's not a rope, or that it's a skyhook: rope all the way up.) The experience of working with arguments is a lot like that; some of them come down when you give them a tug, but a lot of them, even when they move a bit, stay suspended. After enough pulling and tugging, you become confident that a lot of them, anyway, aren't going anywhere, in something like the way that—reverting to the comparison of note 35—your experience with a currency gives you confidence in it.

<sup>42</sup>There is a further respect in which our commitments may be more muted than the regress as we characterized it presupposes. Arguments can't generally be assessed in a vacuum; accordingly, a more contoured rendering of what you are committed to by treating an argument (or a penumbral substitute for one) as satisfactory is that there's a further argument for doing so—provided that there is an *occasion* for such an argument, one that gives content to a substantive demand that the further argument might meet. If this is correct, the regress of commitments has been modulated into a form that is much more reasonable to take on, because you may be pretty confident that the arguments you are invoking are unproblematic and that occasions for providing further defenses of them are unlikely to arise.

<sup>43</sup>What about those "epistemic" oughts? "The umbrella ought to be in your pack" *might* mean that there's a case to be made that it's in your pack. However, that's a good deal weaker than there being a good argument that it's in your pack, which would entail that the umbrella *is* in your pack. And that wouldn't dovetail with the way in which the "epistemic" ought is used to express one's reservations.

by forcing you to figure out which forms of (putative) reasoning work—and work best for what applications.<sup>44</sup> Millian metaethics has a great deal to recommend it, not least that it might in this manner *make us more intelligent*. (Millian metaethics makes you smarter!) With this sort of payoff in prospect, we definitely *ought* to be pursuing the power of positive thinking.<sup>45</sup>

---

<sup>44</sup>We can now return briefly to the charge of “quietism.” Usage in this debate has little to do with the literal and religious sense of the word, and seems to label three related claims: (1) that “there is, in some sense, no way of getting outside normative thought to explain it”; (2) that “no metaethical theories are possible”; and (3) that it is pointless to attempt to “give an account of what it is to think a normative thought” (Hussain and Shah, 2006, pp. 268f). Since the view on offer here is that giving an account of what it is to think *a* normative thought (not all of them at once) is perfectly possible, but doesn’t require stepping outside all normative thought, and that this is what a successful metaethics consists in, these claims needn’t travel together, and the very use of the one word, “quietism,” is evidently question begging. Hussain and Shah frequently repeat the phrase, “the failure to distinguish normative from metaethical questions,” but it likewise begs the question against positive metaethics to assume there is a distinction to be made.

Let’s now consider the *sound* of the word, which suggests that someone is disposed to sit there in a blissful stupor and leave everything that needs fixing in its dilapidated state. A “quietist” philosophical view would be one that assuaged your worries about “normativity,” and so allowed you to go on in your deliberative and moral practice just as before. That sounds like old-school twentieth-century metaethics. Whereas positive metaethics gives you an incentive to improve your understanding of practical argumentation, and even to invent new kinds of practical argument—i.e., it encourages you to do things differently—and so positive metaethics is precisely not a form of quietism.

<sup>45</sup>I’d like to thank Krister Bykvist, John MacFarlane, Shaun Nichols, Mark Timmons, Derek Parfit, Susanne Mantel, Erasmus Mayr, and Georgetown University’s Reasons Reading Group for conversation on these topics; Chrisoula Andreou, Carla Bagnoli, Sarah Buss, Alice Crary, David Enoch, Christoph Fehige, and Mariam Thalos for comments on earlier drafts; and audiences at the University of Utah, Ben-Gurion University, the University of Genoa, the University of the Saarland, the University of Chicago, and Oxford University. I’m also grateful to All Souls College and the Bogliasco Foundation for residential fellowships, to the University of Utah for travel support, and to the Arizona Freedom Center for hospitality.



## Why Do We Think There Are Things We Ought to Do?

Twentieth-century analytic philosophy was characterized by a semantic approach to its problems, by which I mean that you would typically approach a philosophical problem by trying to give a theory of the meaning of some class of expressions. To do that, you would, again typically, rewrite those expressions in a notation that was adapted from, or an extension of, mathematical logic. This way of proceeding was the “analysis” that gave analytic philosophy its name.

These techniques have had a spectacular hundred-year run, but I think it’s time to try something new. My intent here is to exhibit an alternative approach, and to suggest that it should sometimes supplement and often replace semantic analysis. The alternative way of proceeding involves giving a design characterization of an intellectual device, along with an argument meant to exhibit the work such a device does within a larger cognitive system and intellectual environment; just to have a name for it, I’ll call these *cognitive-function analyses*. The problem I will use as my testbed is a staple of the metaethics of the previous century, the correct understanding of the practical or action-directing “ought” or “should.”

When the demands being made on them change, devices quite properly change as well. I will in due course suggest that the oughts which have served us throughout recorded history are in the process of being exapted to a role that allocates additional functionality to them, and that these newer oughts are coming to be somewhat differently contoured. Once you start looking at the intellectual devices as design solutions, the natural next step is to consider how they get—and can be further—*redesigned*.

That’s an ambitious endpoint to be aiming toward, and I’ll need to take the long way around to get there. I’ll begin with a just-so story meant to explain a very raw sort of ought, one that has been treated as a moral bottom line by philosophers of a certain stripe. After reminding readers how those raw oughts

have been construed, I will deepen the just-so story by adding to it a sketch of an engineering problem, meant to display what the oughts in the just-so story are doing for us. Taking it as a baseline, I'll subsequently broaden the account out to accommodate oughts of other kinds. At that point I will explain why we should not regard the account as a semantic analysis.

Then I'll consider how the device we will be examining is starting to change. I'll suggest that the unprecedented depth and pervasiveness of specialization that has transformed our lives over the past couple of centuries imposes stiffer demands than formerly on the oughts delivered up by experts. I'll briefly discuss how our older oughts have been adjusted as they have come to occupy their new role, and how they might be adjusted further. With the extended illustration complete, I'll conclude with an overview of the cognitive-function approach and its prospects.

To keep the prose readable, and the typesetting easy on the eyes, I'll be dropping the quotation marks around "ought," and I'll let context determine when what is intended is the term and when it is a directive or assertion whose main operator is expressed by "ought" or "should." Finally, I'll let my oughts—all of them practical—cover the class of determinations that one ought to or should do one thing or another, whether or not the intent of the practical determination is specifically moral.<sup>1</sup>

## 6.1

On the distant planet of Asimov, the end of the world comes once a generation.<sup>2</sup> The survivors are children, the oldest of whom have only just learned to read, and in the cataclysm all of the paper, parchment, and of course electronic media are destroyed. The soon-to-disappear elders, call them Generation X, regard it as a matter of some urgency to leave what they can of their accumulated practical wisdom to their successors, call them Generation Y. The only sufficiently permanent medium is stone, and so, for their children's benefit and edification, Generation X carves into giant stellae a relatively short list of the things that one ought to do. The medium enforces brevity: whatever supporting argumentation there was is of course omitted, and with it, the qualifications and hedges that accompanied both the strictures themselves and the reasoning that led up to them.

---

<sup>1</sup>By "practical," I mean to mark the contrast with an epistemic ought, that is, the "ought" of "He was only five minutes away, so he ought to be there by now." I don't think that I'm in a position to say what counts as specifically moral; the reading instruction is just telling you to treat directives like "You ought to take the next bus" as in the mix, and not to suppose that we're confining ourselves to those on the order of "You shouldn't steal pencils from the blind."

<sup>2</sup>Sci-fi buffs will recognize the story I am telling to be an adaptation of Asimov (2001).

Generation Y starts off its career with a list of oughts whose terseness ensures that it will directly cover only a small part of the range of decisions its members will have to make. Thus the real function of the list, almost always, is to supply premises for practical argumentation. As Generation Y gains experience with the deontic repertoire, the limitations of the dicta composing it gradually become apparent; the more perceptive users notice that some of the advice was in various ways mistaken. Sometimes it pans out well most of the time, but not always; sometimes situations arise in which the guidance proves inconsistent; sometimes the members of Generation Y decide, on the basis of their experience with drawing such conclusions and acting on them, that the inherited advice was blind to a way things can matter. They revise their list of usable premises, supplement it (and their downstream inferences) with hedges and qualifications . . . and eventually the time comes to leave their own list of oughts to Generation Z.

Now, recall one of the metaethical frames for the oughts that have traditionally been of most interest to moral philosophers. Again and again over the history of moral theory, we find them adopting a way of thinking on which there are actions you *just* ought to do, or sorts of things that *just* are good—not because there are arguments in favor of doing them, or reasons supporting the evaluation, but as a matter of brute, metaphysical fact. Because those assessments are held to be unsupported by argument, but made nonetheless, they must be available without argument, and so such moral philosophers take themselves to be appealing to their moral intuitions—a class of evaluative judgments on which they can *just* rely. Typically but not always, those judgments exhibit a marked lack of nuance: when a theorist of this persuasion tells you that (and here’s a real and often-adduced instance, one that conveys a sense of just how crude these assessments are) torturing babies is morally wrong, he can usually be gotten to insist that torturing babies is *always* wrong, no exceptions allowed. To have a name for this slant on moral and other assessment, let’s call it the *just-really view*: because they’re *just* facts (really!), you *just* know them (really!), and there are *just* no exceptions (really!).<sup>3</sup>

A stance of this sort is bound to prompt objections to each of its components: that there could be no such metaphysically brute evaluative facts; that there is no justification for relying on assessments that you can’t justify; that there is no limit to the level of nuance and sensitivity to the specifics of a situation that may be required of a competent action-guiding assessment.<sup>4</sup>

---

<sup>3</sup>My gesture of inclusion is vague around the edges, and not every philosopher who buys into one component of the just-really view buys into all of them. There is no point in providing a lengthy list of targets, but analytic philosophers will recognize a subsidiary tradition stretching from G. E. Moore to the later Derek Parfit. Huemer (2005) is a spirited and recent defence of the first two components.

<sup>4</sup>The most widely cited instance of the first two complaints is Mackie (1977, Chapter 1), probably because his response is as baldly put as the stance it is addressing; Jonathan Dancy (1993) has become identified with the last of the objections.

And when the objections sound unanswerable, and their target seems to be the theory of evaluative and action-guiding assessment that is correct if any is, nihilism can appear inevitable.<sup>5</sup> But by this point we should be wondering whether something has gone wrong in the take on moral philosophy that started off the slide to nihilism in the first place.

Return to the Asimoveans: what will their moral phenomenology be like? The young—of, let's say, Generation Z—are raised on a series of dicta that are presented to them unsupported by arguments; they encounter them on stone tablets that might as well have come down to them, like the Decalogue, from a divine dictation exercise. When a young Asimovean deploys such a dictum as a premise in his practical argumentation, it turns up as something that he *just* accepts. When he is asked to justify it, he is unable to provide any arguments for it—even though it was inscribed on the Karnak-like edifice only because it *was* the conclusion of a (very long!) series of arguments—and he can do no more than reiterate it. So he finds himself insisting that this is *just really* how it is, that it is a *fact*. On the one hand, his confidence in his position is not unmerited; on the other hand, he is unable to explain those merits. So the first component of the just-really view appears here as the mode of expression of someone who is at a loss for words, but who does not realize *why* he is.

The young members of Generation Z receive their moral imperatives unqualified. The correct explanation has to do with the way that the hedges that originally accompanied those dicta depended on the arguments that supported them: when the arguments are lost, the qualifications are lost, too. But unless a Gen-Zer is very thoughtful indeed, the lack of hedges in his inherited practical wisdom will turn up in his moral phenomenology as the lack of nuance that we remarked on as characteristic of the just-really view. The exceptions never made it onto the stone tablets, and so the young insist that the dicta apply with no exceptions. Already we can see how it will come about that their moral philosophers thump their tables and insist that torturing babies is morally wrong, *just* morally wrong, absolutely wrong, no exceptions allowed.<sup>6</sup>

As we have told the story so far, the Gen-Zers know where their moral imperatives come from: they are inscriptions on enormous stone tablets. When they are asked how they know that this or that is what they are to do, they end up pointing back to the tablets, and when asked why they rely on the tablets, they insist that gigantic stone tablets *are* the kind of thing on which, in practical and moral matters, you can rely. Their philosophers find themselves arguing about the epistemology of stone tablets, and one of the respectable positions to occupy turns out to be “moral tabletism”—the view that inscriptions on stone tablets are *just* the kind of thing on which you may rely.

---

<sup>5</sup>There is by now a lengthy catalog of attempts to respond to the objections on their own terms; Crary (2007) is a recent example.

<sup>6</sup>Am I suggesting that it *is* sometimes alright to torture babies? I'm not squeamish, but I don't think so; in footnote 16 I'll explain why not.

But let's imagine that even the enormous stone tablets disintegrate (perhaps they're made of sandstone, and the weather is bad even between cataclysms), normally just as the children are growing up. The adults they become will have internalized what the tablets said, but will no longer remember the inscriptions themselves. And so when they are asked how they know this or that is what they are to do, they fall back on saying that they *just* know. And when they are interrogated as to why they rely on these dicta, they fall back on saying that, when it comes to pronouncements like these, you can *just* rely on them. Once he's lost track of where his practical premises come from, a philosopher who needs an epistemology is likely to think he has no alternative but to treat convictions of this sort as self-warranting: this is where talk of moral intuitions comes from.<sup>7</sup>

The Asimoveans—the less thoughtful of them, anyhow—will *find* themselves with the just-really view. I am about to suggest that the Asimoveans are a highly simplified, idealized model of the way we wipe our own cognitive slates clean when the burdens of inference become unmanageable, and that we should not be surprised when our moral philosophers come up with the just-really view. But in that case, we should not take the just-really view at face value: in the Asimoveans' case, *every element* of the just-really view is *just* a mistake.

It is not merely that we are exempted from taking time out to refute a position that is arrived at on the basis of the sorts of fallback table thumping we have just described. Let's remind ourselves of the fancier names that PhD-bearing philosophers have for the components of the just-really view: *moral realism*, *moral intuitionism*, and *absolutism*. Those fancier labels obscure the extent to which each component is merely a peculiar sort of emphasis, normally expressed with a sound one might make if one were trying to lay an egg—that is, not a content or a theory or a *view* at all. You can't refute an *emphasis*, and it's a mistake to make doing so into a high-priority action item. For this reason, the standard labels are too misleading to be tolerated.<sup>8</sup>

---

<sup>7</sup>It turns out that the idea that moral intuitions are a side effect of forgetting one's reasons has been invented before—and, perhaps all-too-appropriately, given the thesis itself, forgotten: in 1768, Abraham Tucker, writing under the pseudonym “Edward Search,” called the process “translation” (Albee, 1957, pp. 136f).

<sup>8</sup>What the so-called redundancy theory of truth gets right is that when someone asks why you think that *p*, and you reply, “Because *p* is *true*!” you're just repeating yourself. Now, that observation is equally on target if you say instead, “Because it's a *fact* that *p*!” There's every bit as much truth in a redundancy theory of facts as there is in a redundancy theory of truth. That's why it seems to me that the version of moral realism which claims that our moral beliefs reflect the *facts* is actually a matter of emphasis conveyed through reiteration (rather than, say, italics), and not a theory, properly so-called.

But the moral realist component of the just-really view has also been couched more indirectly, as, variously, a claim that moral views are beliefs, as opposed to emotions or other noncognitive attitudes, and that the objects of those beliefs are propositions. When the philosophers' technical vocabulary is paraphrased back into honest English, that latter boils down to a reaffirmation that what was said was

## 6.2

I'm going to take a shot at framing the just-really view as a series of side effects of a bounded-rationality solution to a problem generated by a pervasive formal feature of practical argumentation. Allow me a working hypothesis, that the central (though not the only) use of a practical ought is as an existential quantifier over arguments for doing something: to say that you ought to do something is, again in this central use, to say that there's a good argument for doing it—and if you think that you can have good arguments both pro and con, make that a *winning* argument.<sup>9</sup> If that's right, what do the arguments being gestured at—as philosophers would say it, the arguments in the domain of quantification—look like?

An argument form is *defeasible* when an argument of that form can be in order as far as its assembly and structure go, as well as having correct premises, so that its conclusion would normally also be correct—yet nonetheless be aborted by considerations that crop up in one or another occasion of application.<sup>10</sup> Deductively valid arguments guarantee the truth of their conclusions, conditional on the truth of their premises; a defeasible inference form does

---

the sort of thing that might be true or false (and is in fact *true*)—so, once again, emphasis. But there is room for speculation as to how one would arrive at *that* way of reiterating one's practical claim.

Alasdair MacIntyre has pointed out that as a matter of historical occasion, emotivism arises when students of intuitionists correctly characterize the posturings of their teachers, but mistake that for a theory of moral discourse in general (MacIntyre, 1997, pp. 11–18). When the students complain that the teachers are just adopting an attitude, or striking a pose, it's all too natural to respond that what is being expressed is not just an emotion or a concealed imperative, but a belief.

However, that response misinterprets the force of the complaint—and does so even if the misinterpretation was shared by the emotivists and their successors. A sports fan's "The Jazz are the *best*" is properly characterized as merely the expression of an attitude in just this way, whereas the careful and argued evaluation of a student's paper is not. Now, as far as types of propositional attitude go, and as far as formal objects of the respective attitudes go, both are on a par; whatever your theory of evaluations is, both are evaluations. Construing the charge that what we have is merely the expression of an attitude as a claim about what type of propositional attitude is in play is a convenient misreading of what is at stake; and attempting to rebut such a claim with a counterclaim to the effect that one is expressing a belief whose object is a proposition is a red herring.

<sup>9</sup>The proposal is developed further in Chapter 5. Now, even though the range of the claim is restricted to the "central" oughts, I have found that the existential-quantifier proposal attracts objections to the effect that there is a mismatch between what oughts there are and what arguments are available. Usually the drift of these objections is that there are fewer arguments than oughts: as I just remarked, it's thought that on many occasions you *just* ought to do something, no argument required. However, notice that the mismatch could go the other way: there could be a greater number of successful practical arguments than allowable oughts, and Robert Nozick, to whom arguments came much more easily than to others, can serve as an example of someone who might well have thought so: see Nozick (1981, pp. 2ff) and Millgram (2004b, p. 334). I will return to the complaints in note 35, below.

Broome (2013) also uses the designation "central ought." His use and my own are entirely different; please don't get confused.

<sup>10</sup>The near-synonym in the world of artificial intelligence is "nonmonotonic"; for brief discussion of the logic of nonmonotonic inference, see Brandom (2001, pp. 472f).

not.<sup>11</sup> For example, brand loyalty is often based on an inductive inference: the performance of the product up to this point has been reliable, and so I expect it to be reliable in the future. The inference is often perfectly in order, but if you learn that the manufacturer has just been bought by a larger firm, you would do well to think twice; it is standard operating procedure for many large firms to treat brand reputation as a short-term resource they can run into the ground. You could obviously go on as long as you like thinking up defeating conditions for this particular inference, and that is characteristic of defeasible inferences: they come with, not exactly infinitely many defeating conditions, but *indefinitely* many.

Just about any practical inference form that you would have real occasion to deploy is defeasible. Means-end inference is defeasible: if you need more cash, and this ATM right here is a way to get the cash you need, straightway (as Aristotle would say) you insert your bank card, and that practical conclusion is normally fine—but not if there’s a card catcher or skimmer on the device, and not if there’s a shady character lurking around the site, and not if the fees for using this particular ATM are unreasonable, and not if you get that urgent phone call from a hospital emergency room, and not if you’re participating in a consumer boycott of the bank . . . and, as is characteristic, this list can be continued indefinitely.

Practical induction is defeasible: Perhaps in order to satisfy academic requirements, and later, because your employer requires it, you learn one language after another. You discover, to your surprise, that there are real pleasures to be had from the mastery of a language; figuring out its tricks and quirks and modes of expression is surprisingly rewarding. And at some point you realize that acquiring those languages is making you, in small but important ways, a more thoughtful, *better* person. You had had no prior interest in this way of finding things engaging, nor in this sort of self-improvement, but you generalize from your observations to a practical conclusion: learning new languages is desirable, because these benefits matter.<sup>12</sup> Although your practical induction is perfectly in order, it would be aborted were any one of indefinitely many further considerations to be brought into play: learning more languages is not a good idea if there’s a ceiling on how many you can keep straight, and you’re pressing up against it; and not if the overhead involved in keeping them up starts to become unsustainable; and not if the thoughtfulness turns into the dissolution of ethical knowledge that Bernard Williams worried about so poignantly in his *Ethics*

---

<sup>11</sup>A generalization of the notion of deductive validity suitable for practical reasoning would construe it as guaranteeing that the conclusions of an argument take a designated value (not necessarily truth, which isn’t the dimension along which we standardly assess intentions and choices) when the premises take that value.

<sup>12</sup>For more description of the form of inference, and an argument that practical induction is legitimate, because indispensable, see Millgram (1997, Chapters 3–5).

and the Limits of Philosophy (in which you become alienated from your native ethical vocabulary, and all the others too); and not if your delight in the idiosyncrasies of different languages makes you tedious to your friends . . . and so on.<sup>13</sup>

Specificationist inference is defeasible. Johannes Kriesche set out to make good art, which is, however, not a definite enough goal toward which to take steps; first he had to specify his objective concretely enough to anchor chains of means-end reasoning. This deliberative process took him decades, during which time he found his way to a medium and a subject matter: canvases covered with paraffin, through which the viewer could make out monochrome images of German gas stations at night. In a country whose businesses still almost all shut down at the end of the day, the images show islands of illumination in dark, abandoned landscapes, and Kriesche's "light temples" remind their viewers of Hopper's *Nighthawks*.

Although his specificationist deliberation proved remarkably successful, his train of thought could have been defeated by any number of considerations. Had someone else already settled on that subject matter and medium, the inference would have been defeated; had gas stations turned out to evoke inappropriate associations, as perhaps they would have if, say, a politician had themed an election campaign around them, the choice would have been an error; had the fragility and heat-sensitivity of the paraffin proved an insuperable obstacle to having the work transported and exhibited, Kriesche would have had to rethink his artistic self-definition; likewise had he simply found himself uninterested in gas stations . . . and so on.<sup>14</sup>

There are further forms of practical reasoning, and I won't review them all here. As far as I can tell, with occasional exceptions in unusual circumstances of application, arguments that take any of these forms prove to be defeasible. I hope on a future occasion to take up the task of explaining why this is so, but for the moment I am going to treat it as an observation.<sup>15</sup> And now we can return from our characterization of practical argumentation and connect it with the hypothesis I introduced a few steps back, as to what an ought (centrally) indicates. When practical oughts are quantifiers over arguments for

---

<sup>13</sup>Williams (1985, pp. 167ff); see Section 4.2 in the present book for a brief summary of the last-but-one problem.

<sup>14</sup>For a survey of the specificationism debate, see Millgram (2008).

<sup>15</sup>Kant's categorical imperatives may look like the obvious exception: if the "CI-procedure" rules out a maxim that tells you to lie, it *always, exceptionlessly* rules it out. But Kantians devote a remarkable amount of casuistical effort to making room for defeating conditions in the particular cases where these have come to their attention: what if there's a murderer at the door—can you lie *then*? (See Schapiro 2006 for a recent and typical entry in this discussion.) It's not an accident that one contemporary Kantian, Christine Korsgaard, finds herself forced to make room for defeasibility in her official rendition of the theory (as "provisional universalizability"; for a critical overview, see Millgram 2011a). A close reading of Herman (1993) will further suggest that for Kantians questions of defeasibility tend to get hidden inside the question of what maxims to submit for testing.



doing something, with exceptions that are too sparsely distributed to matter, those quantifiers range over defeasible arguments.<sup>16</sup>

### 6.3

Defeasible inference patterns come with extra cognitive overhead. When you deploy a deductive inference pattern, it suffices to plug the premises of your argument into the slots provided by the pattern to draw the conclusion. (Of course, if you find you can't buy the conclusion, you may end up running the argument backwards and discarding one or more premises.) But when the inference is defeasible, you have to *pay attention* to disqualifying possibilities; you have to be constantly on alert. This can be quite demanding: there is no finite list or formula to check your circumstances against, and you cannot memorize the exceptions. Normally, recognizing a defeating condition involves understanding and bringing to bear the *point* of what you're doing, and the sorts of intellectual activity this involves can't be performed on automatic pilot.<sup>17</sup> Nobody knows how we manage the defeasibility conditions of our inferences—it's one of the great philosophical mysteries—but it is nevertheless a truism that the more attention you have to pay and the more alert you have to be, the greater the cognitive load.

We said that the *ceteris paribus* clause built into a practical inference can normally be triggered by indefinitely many defeating conditions; so it is not exactly that two such inferences involve *more* defeating conditions than one.<sup>18</sup> Still: Maybe Hume was right when he insisted that you can't get an ought from an is, and maybe he wasn't.<sup>19</sup> But his claim wouldn't have been at all plausible

---

<sup>16</sup>Return to those tortured babies. If I'm right, oughts are, in the most central cases, quantifiers over arguments, and it's actually very, very easy to give good arguments for not torturing babies. Each of those arguments is practical; accordingly it will have defeaters. Moral philosophers who have noticed potential defeaters—though they tend to think of them as defeating conditions of the dictum, rather than of the argument for it—often attempt to preempt them by adding extra clauses: for example, what is immoral is torturing babies *for fun*. When they do so, they are making a *logical* mistake; as we've seen, defeaters don't run out, and there's no point in trying to fold a list of them into one's practical conclusion.

Nevertheless, we can often be quite confident in the shared conclusion of many convergent practical arguments, as in this case.

<sup>17</sup>Recall one of the earlier illustrations: you become a better person when you learn another language because, you realize, being able to distinguish your thought from its mode of expression in a particular language frees you from moral blind spots. The ethical paralysis that Williams describes undercuts that point of learning further languages, and *that's* why it's a defeating condition. The relevant feature of the illustration is that drawing connections of this kind isn't a formulaic matter. For further illustrations and discussion of the claim, see Millgram (2005a, Chapter 10).

<sup>18</sup>Although "indefinitely many" is not the same as "infinitely many," we can be reminded of an elementary sum in transfinite arithmetic:  $\aleph_0 + \aleph_0 = \aleph_0$ .

<sup>19</sup>Hume (1888/1978, p. 469).

if arguments ending in an ought didn't often enough have oughts among their premises. Suppose practical oughts *are* (again, in their central use) quantifiers over arguments: what that would mean is that the conclusions of practical arguments frequently serve as the premises of further arguments; and the conclusions of those further practical arguments may serve as premises in yet further arguments to still further practical conclusions. The existential quantifier tells us that there are arguments for the premise it marks; it does not itself imply that anyone actually has produced those arguments, but normally those premises are so marked because they themselves were arrived at through arguments that were actually produced, either by the person enunciating the ought or by someone else.

When a long train of practical reasoning is assembled out of such stretches of defeasible inference, it requires its users to keep track of the defeasibility conditions of all of those stretches of argument: any of them can abort the final conclusion of that longer train of thought. Typically, different sorts of concern are relevant to determining defeating conditions of the different stretches of inference. And the more attention you have to pay to different kinds of concern, the *more work* it is.

So suppose you are constructing an argument for a practical conclusion, and suppose that it draws on premises that are marked as oughts, because they were themselves arrived at by argument. That further argumentation in turn contains steps that are supported by previous argumentation, and the lines drawn between the various arguments are obviously merely arbitrary. Thus when you are constructing your own argument, you are in fact in the middle of what is likely to be a very long argument indeed. Your confidence in the conclusion of your argument properly depends on your control of the defeasibility conditions of the very long argument. In typical cases, different stretches of this very long argument will have very different sorts of potential defeaters: issues to which you need to remain alert, which may be as qualitatively different as you like. In most cases, that's just too much to ask of anyone.

Perhaps there are ways to tamp down that daunting assessment. One could opt for principles that are allowed simply to override defeaters, or one can simply decide to ignore whole classes of defeaters up front, as when I tell myself that in *Realpolitik*, morality doesn't count.<sup>20</sup> Again, sometimes it's not all that difficult to deal with the consequences of a missed defeater. If you insist on having a defeasible argument mimic a deductive one—if you want to be *sure* that the conclusion is correct—then you do have to catch all the exceptions, and that's plausibly very hard. But you don't always need to be as sure as that:

---

<sup>20</sup>If you squint just right, mid-last-century rule-utilitarianism can provide examples: the argument for keeping your promises involves the conduciveness of the practice to overall utility, and although in some particular case it may be clear that utility is best served by breaking your promise, you are supposed to ignore that fact and do what you said you would. Thompson (2008, Part III) is a useful discussion of the logic of this move. (For that latter suggestion, I'm grateful to Michele Palmira.)

I shouldn't have taken *this* flight (too late, I realize that I was bound to miss my connection), but missing a connection is normally a recoverable, albeit inconvenient, mistake. Or: although outsiders get quickly swamped, experts can often see where the problems are at a glance, anyway when an argument is in their area of expertise.<sup>21</sup> Lastly for now, perhaps each wave of argument can be beta-tested for some period of time, by seeing what happens as its conclusions are put into practice; by the time the next round of argument is launched, the previous wave will have been somewhat debugged.

No doubt we avail ourselves of these and other fixes, but the inexhaustibility, surprisingness, and occasionally catastrophic upshots of the defeaters mean that we shouldn't think our problem is thereby solved.<sup>22</sup> The three elements of our predicament were: first, the defeasibility of practical inference forms; second, the requirement on competently deploying a defeasible inference form, that one exhibit sensitivity to its defeating conditions; and third, the need to allow that the conclusion of one practical argument always may be, and often is, a premise of another. Taken together, just about whatever practical argument we are in the middle of can be expected to be a cognitively overwhelming and realistically unmanageable task.

In many ways, we are somewhat better off than the Asimoveans. We retain a great deal in the way of written records from generation to generation, and these records tell us, not always completely or reliably, why our predecessors thought this or the other to be a good idea. Because our generations overlap a great deal more than theirs, the older people dispensing the advice to the youth stick around long enough to be asked questions. Nonetheless, what we actually do is something rather like the Asimovean procedure. Nature has given us a method for sidestepping the unmanageable cognitive load: death wipes the slate clean of all those defeasible arguments for doing this or that.<sup>23</sup>

---

<sup>21</sup>de Groot (1965) is an early but still impressive investigation of these sorts of differences between expert and novice problem solving. A more elaborate variation might be panels of experts, each looking out for defeaters in their areas (a suggestion I owe to John Hardwig).

<sup>22</sup>Taleb (2012) and Taleb (2010) provide emphatic reminders that ordering your search path by the past frequency of a defeater for a given argument type can't be expected to handle infrequent but nonrecoverable defeaters. However, it doesn't appear feasible to focus only on major defeaters; we identify these by first finding the potential defeaters and then investing further resources to determine which among them are likely to be catastrophic.

<sup>23</sup>"Rather like": On the one hand, we don't necessarily wait for people to die before deleting the supporting argumentation (Vines et al., 2013). And we do seem to preserve texts that purport to provide practical guidance for relatively long periods of time—the Bible, the *Nicomachean Ethics*, and so on—and not just preserve them, but actively refer to them.

On the other hand, such texts accrete layer upon layer of interpretation; these interpretations are passed down, very often without their supporting arguments, and are treated as exegetically authoritative; they have been able, time and again, to make the texts they cover over say just about anything, including the very opposite of what they originally meant. In this way an apparent exception to the Asimovean phenomenon turns out, most of the time, to be another instance of it.

Notice, however, that in the just-so story, there won't be room on the stone tablets to include any sort of elaborate explanation of what the key terms in the dicta (e.g., "murder," "theft") mean. Argument over the inherited doctrine will accordingly tend to take the form of argument over the senses

## 6.4

This is not the moment to consider whether the problem has a principled solution. While it is important to tackle the problem of cognitive load, my own sense is that real headway will require a deeper understanding of defeasibility than we now have. So for now I am going to leave it as a very large loose end. Instead, let me flag a handful of new questions raised by our model of that *unprincipled* solution.

First, as the simplified model has it, each generation conducts its practical argumentation on the basis of premises that have had the qualifications and hedges with which they were formerly accompanied stripped off. When the new generation deploys these premises, they will learn to accompany them with their own restrictions, hedges, and qualifications. Can we expect the subsequent generation to recover and reconstruct the defeating conditions of the preceding generation? Will the understanding of the defeasibility conditions of their practical inferences reproduce, largely or even partially, the inferential command of the former generation?

Second, while we haven't assessed the strengths and weaknesses of the unprincipled solution—and in particular, I didn't try to show that it will give us correct or high-quality conclusions—we might well expect the inferential performance of each generation to start off badly, because they will not have a good sense of when their arguments should not go through. And we can expect their practical reasoning to end up sclerotic, as the cognitive load increases over the course of the cycle. Is there a sweet spot—practical deliberation that comes neither too early nor too late—in which performance is high?

Third, again as the simplified model has it, each new generation starts out with a smallish set of practical premises and ends up with a smallish set of conclusions that they then leave to the following generation. Under what conditions can we expect overlap in the earlier and later rounds of canonized oughts? Are there oughts that will be reliably reproduced—rederived and rebequeathed—from generation to generation? Can we expect such islands of stability to grow over time? How sensitive is the appearance of such islands to the choice of premises in the initial round? And if we *can* identify this sort of robustness, should we treat it as our best substitute for that “just really is a *fact*” of the moral realists?<sup>24</sup> Or should we understand it as merely an artefact of the procedure?

And finally for right now: how robust is the model itself? If we make it more realistic—for instance, by relaxing the assumptions that the generations

---

of such key terms. (I'm grateful to David Schmidtz for this observation.) When the tablets are more permanent, as in the case of the texts we were just considering, exegetical resolutions of such questions have tended to freeze up into dogma. The deontic inheritance of the different generations is in our own case partly constituted by the leading edge of the wave of interpretation.

<sup>24</sup>Maybe not: as John Hardwig has gently reminded me, various forms of bigotry, passed down from generation to generation, have remained remarkably stable.

are discrete, or that the transmitted contents are so very minimal, or that the society is homogeneous enough to settle on a single list of dicta to bequeath—will the effects I have described persist? More generally, inferences drawn on the basis of modeling are defeasible; can we characterize some of the more important classes of case in which we should refrain from drawing conclusions from this one?

## 6.5

Now that we have our stripped-down, simplified model of those very raw oughts, let's consider more closely what it is telling us. You might think the hypothesis we've been entertaining about what oughts are amounts to something on the order of a definition of the word: an ought is an existential quantifier over arguments for doing something. This is a semantic thesis, and whatever the pragmatics of an ought turn out to be, they will fall into place around its meaning, which consequently comprises the most important part of our philosophical treatment of the topic. Never mind what they *do*: figure out what they mean, and their cognitive function will fall into place on its own.

I doubt that way of seeing things can be sustained, and to explain why, recall that I introduced the hypothesis as covering the *central* use of an ought. But the central oughts are pretty clearly not the only oughts. Informed-desire or informed-preference theories highlight an ought supported by a heuristic: telling someone he ought to do something can be a gesture at the (one hopes!) fact that, if he thought about it more carefully, he would decide to do it. A relative of this heuristic, famously invoked by Mill, involves surveying the preferences of actually existing people who have already made the choice you're considering; telling someone he ought to do something can be a gesture at such an opinion poll (maybe a poll that hasn't yet been commissioned). Then there are oughts delivered by experts, which we'll discuss further in due course; while some of these are central oughts, namely, oughts backed up by the arguments the expert might produce, frequently expert know-how isn't articulated, and for all we know in principle can't be articulated. (There's a why, alright, but the expert can't *say* why; he might announce, for instance, that he's pretty sure there's "something to it.") In this sort of case, you're relying on the expert's training and experience, but not on any supporting considerations he could adduce.<sup>25</sup> Presumably there are *sovereign* oughts: you are bound by the dictates of—for instance—a legislature, whether there are reasons backing them

---

<sup>25</sup>Experts aren't always degree-bearing licensed professionals. When someone who has spent his life in a gritty American city tells you that such and such a neighborhood feels dicey to him after dark, but can't say why, that can be expertise speaking in just this way. (I'm grateful to Dale Clark for discussion of this topic, and to George Gerpheide for the example.)

up or not.<sup>26</sup> Call such oughts *penumbral*; since heuristics don't run out, and neither do (types of) sovereigns, there are evidently indefinitely many more of them.<sup>27</sup>

In the analytic tradition, a semantic analysis functions as a substitution license: if it is satisfactory, it allows you to replace the term or notion with its definition, or, in a slightly more sophisticated variant, it allows you to replace the term or notion, together with some of its surrounding context, with a formulation specified by the analysis.<sup>28</sup>

Thus penumbral oughts show us that we had better not think we know what "ought" means. While substituting a disjunctive definition for a term is unproblematic, substituting a disjunction that trails off into an ellipsis is problematic; you have not produced a well-formed expression bearing a properly specified content if, in the middle of it, there is a clause reading "... and fill in indefinitely many further disjuncts here."<sup>29</sup> In grade school they tell you that

<sup>26</sup>That said, laws are not normally formulated with oughts; rather, they have the form, "So-and-sos *shall* such and such, and this and that are the penalties for noncompliance." When you decide that you ought to do something, because it is required by law, you normally have an argument for doing it, one in which the law figures as a premise. Consequently, it is easy to assimilate the oughts grounded in legislation to our central oughts, and the legislation itself presents no additional oughts to accommodate.

One further point: often enough, a sovereign ought is supported by argument, but an argument *for the sovereign* and not for the subject. And more generally, the reasons someone has for telling you that you ought to do something may not be *your* reasons for doing it.

<sup>27</sup>For those to whom this penultimate point isn't obvious, consider the special case of heuristics adopted where there is a single independently specifiable success criterion. (I mean, as contrasted with more complex cases in which the target is a construct tied to many different anchors, some of which may themselves be approximation techniques; for a very useful example, see Tal, 2014.) Anything that serves as a good-enough proxy for that success criterion might be usable as a heuristic; how many such proxies come to hand is limited only by the structural richness of the world and by your ingenuity. Because there are indefinitely many ways to cut corners, and also indefinitely many tradeoffs to be made when you are choosing among competing shortcuts, indefinitely many heuristics are normally going to be in the mix: just for instance, in the academic institutional case, "best practices," what our peer universities are doing, standing in some government or commercially provided ranking system and so on.

<sup>28</sup>The delimiting phrase is there because the idea that definitions function as substitution licenses is not shared by all philosophical traditions. For instance, the pragmatists had their own very different views about what content analyses did, on which it is not a consequence of the correctness of a definition that you can substitute it for the term being defined. When Bertrand Russell and William James had their back and forth about one of the pragmatist theories of truth, the underlying methodological bone of contention was Russell's assumption that if truth is what works, you can substitute "what works" for "true" in test sentences, an assumption that both James's theory of meaning and his theory of reference disallow. (See Russell 1999 and James 1977, Chapter 14.) Again for instance, Whewell 1858, p. 175, complains about what he calls "this erroneous habit of thought," that "the Rose Tribe are [he proceeds to quote Lindley's *Nat. Syst. Bot.*] 'Polypetalous dicotyledons, with lateral styles, superior simple ovaria, regular perigynous stamens, exalbuminous definite seeds, and alternate stipulate leaves'"; however, "no one would say that this was our essential conception of a rose, to be substituted for it in all cases of doubt or obscurity." (For his own conception of definition, see Whewell 1857, vol. iii, p. 90.)

<sup>29</sup>For a representative special case of the complaint that if you can't fill in the ellipsis, the sentence has no determinate content, see Reutlinger et al. (2011), Sections 4 and 5.2. I should emphasize

a sentence is a complete thought; when there is an ellipsis sitting in the middle of it, it isn't a sentence, and what it expresses is less than that complete thought. We've just observed that the account of ought we sketched came with a trailing ellipsis: there were those central oughts, which were quantifiers over arguments, those penumbral oughts, four of which were briefly described, along with indefinitely many further heuristic-based and sovereign oughts—and possibly along with those “*just* oughts,” if we do not want to disallow all of them on the basis of our Asimovean just-so story. Substituting *this* into a test sentence, say, “You ought to do it,” gives you something on the order of: “There's a good argument for doing it, or perhaps a survey of the preferences of the experienced would tell you to do it, or perhaps. . . .” The substitution does not give us a sentence with a definite content, and so it would be uncharitable to treat the hypothesis as proposing a *definition* for ought.<sup>30</sup>

---

that ordinary people routinely operate with thoughts that contain “et ceteras,” “and so ons,” and so forth. I am identifying a commitment of the semantic-analysis approach, not a requirement on ordinary speech. That commitment is tied to a background understanding of sentences as expressing propositions, which in turn have definite truth conditions.

I should also distinguish the ellipsis that extends a well-understood series—as in, to borrow an example at random from Polya (1990, vol. i, p. 9),

$$1 - \frac{x}{1} + \frac{x^2}{2} - \frac{x^3}{3} + \dots \quad (6.1)$$

—from those in which I'm interested here: where you can't go on and say what comes next, what comes after that, and so on.

There's a tricky hedge to those substitution licenses, and it has a consequence. Definitions function as substitution licenses when the context is not “opaque,” and since there's no independent or principled account of opacity, the license has really behaved this way: you can replace the term by the definition, except when you can't. That is, you can substitute, unless the context is a propositional attitude, or unless the context makes the orthography matter, or unless it makes the pronunciation matter, or. . . . The ellipsis at the end of the previous sentence means that, by the lights of the semantic analysis tradition, it does not have a definite content. But now, here's a dilemma for that tradition: If you can't do intellectual work with such incomplete thoughts, then you can't so much as state and think about one of the methodologically central commitments of the tradition. That would make your methodological views impossible to defend. But if you can work with them, then while semantic analysis might be useful sometimes, even often, it can't be nearly the whole philosophical story; we can and will have to do philosophical work with thoughts that resist what the tradition endorsed as semantic analysis.

<sup>30</sup>In addition, that first and most important element of that open-ended list is unsuitable as a component of a reductive definition, because it invokes the notion of a good argument, which notion is open-ended in about the same way that the main list is. We invent, develop, and refine new forms of argumentation, and it is hard to believe that we could have a clean criterion telling us what counts as a good argument. The naive tend to think that it's a good argument if it takes you from truths to truths, but first, that is not very plausible when we're considering arguments about what to do, and second, part of the process of developing novel arguments is finding alternatives to truth. For instance, over the past few hundred years, probabilities have emerged as an alternative (Hacking, 1975); more recently, the vagueness literature has entertained alternative designated values (such as supertruth or Edgington's real-valued verities), intended for alternative validity concepts (Keefe and Smith, 1996, especially Chapters 9, 15, 16). The uncodifiability of “good argument” is bound to be reflected in the blurriness of semantic analyses that deploy it.

## 6.6

The alternative to the semantic approach that I am putting on the table is to skip worrying about what a class of expressions means, and ask instead what the cognitive function of a philosophically problematic concept is: what it does for you, intellectually or otherwise. If our model shouldn't be telling us what an ought *means*, we would do better to think of it as a preliminary characterization of an intellectual *device*, one which may be much more open-ended than an attempt at a definition would tolerate.<sup>31</sup> To work our way around to the cognitive function of an ought, I want to point out a suggestive analogy between the workings of what I've been calling the central oughts and deflationist truth.

On what is today probably the most popular theory of truth, anyway among us analytic philosophers, truth is a device that enables semantic ascent. That is, instead of saying, "Sally hasn't eaten her potatoes," which is about Sally, you say, "The sentence, 'Sally hasn't eaten her potatoes,' is true," which, although it's apparently not about Sally, but about a sentence, conveys the same information as before.

---

When you're after a *definition*, you need to understand the terms used in the definition before it lets you understand the term being defined. In a cognitive-function treatment, however, a device and its components may have to be identified and then retooled and made more precise *together*.

<sup>31</sup>Here's further warmup for looking seriously at alternatives to semantic-analysis approaches. If we think of oughts as an intellectual tool, semantics of the traditional sort can be successful only if the design and implementation of our version of the device are extremely *clean*. In a world shaped by natural selection and human history, there's no reason whatsoever to expect that sort of ruthlessly enforced elegance in a found object.

If you have a programming background, here's a helpful analogy. Second-generation programming languages really do have recursively specifiable semantics, but that's because their designers insisted on it. By contrast, first-generation programming languages were messy kludges. Natural languages were not designed under the influence of Edsger Dijkstra, and consequently attempts to reconstruct their semantics as though they were are pretty much doomed from the outset.

Consider a quote from Fara and Williamson (2005, p. 27), complaining about it scarcely matters what, but typifying the expectation built into the semantic-analysis approach:

the objection . . . requires accepting that the semantics for natural language is radically non-compositional. This is an extremely strong commitment, one which seems insufficiently motivated. . . .

But the default is that a found object will not exhibit the regimentation we impose on some of our artifacts, and the expectation the entirely typical quote expresses is hopelessly misguided—it is, in fact, the linguistic analog of creationism.

By contrast, while our Asimovian just-so story was quite idealized, *that's* fine. (Compare the way traditional state of nature arguments, for example, exhibit aspects of the functionality of political states by constructing stripped down and idealized models of their founding events—despite real political states being generally very messy social collages indeed.) The aspiration of old-school conceptual analyses has been to provide a semantic theory that is *true*, and so an elegant such theory can only do its job if the language or system of concepts it is representing is also elegant; and it will not be. Whereas a highly idealized functional model can convincingly elicit the workings of a very messy cognitive device.



Now, if the truth-invoking sentence *does* convey the same information as the simpler sentence, why would you ever want to use this device? Isn't it a waste of time? Well, no. Sometimes you want to assert claims in bulk, and sometimes you want to make assertions whose actual content you don't know. So you perhaps say of *Spine* that it is a very reliable journal, meaning that everything, or most everything, in it is true. Perhaps you have read your way through the back issues of *Spine* and could in principle recite its contents; even in this case being able to put it this way is an invaluable time saver. But you may not have done all that reading, in which case you don't have in mind the contents of all the assertions you're endorsing; certainly when it comes to future issues of *Spine*, you do not affirm their contents because you've inspected them yourself. And you may not be—in this case, because it's a specialized medical journal, and you aren't an orthopedic surgeon—so much as able to *understand* the contents of those assertions. But you can, in this indirect way, make them anyway . . . and that's all there is to truth. If we had to sum up the deflationist view, but in the spirit of the cognitive-function approach I'm recommending, it would be that the primary function of truth is to allow you to make assertions while *screening off* the contents you are asserting.<sup>32</sup>

I happen to think that Dorothy Grover's prosentential account of truth is a much more elegant rendering of deflationism than semantic-ascent theories, and as it also happens, her account will fit the analogy I'll presently suggest very nicely.<sup>33</sup> When Grover wants to account for the assertion in bulk of possibly unknown contents, she turns to substitutional quantification, in which a bound variable ranges over, roughly, sentence-expressing strings that are to be plugged into the matrix of a larger sentence. On this way of explaining truth, that claim about the very reliable journal comes out something like this: (*p*)(if *p* is in *Spine*, then *p*). This way of explaining how truth works looks a lot like the account of a central ought which we already have on the table. On the one hand, in the bulk-assertion uses, saying what's true consists in applying a substitutional quantifier that ranges over (something like) sentences; on the other, a central ought is an operator-like appearance of a quantifier that ranges over arguments. I think the resemblance signals an analogy of function: just as truth lets you affirm claims, while screening off their content, oughts (in their

---

<sup>32</sup>Although the deflationist view of truth is popular, and although truth can certainly be used this way, my own view is that the primary function of truth lies elsewhere. Here, however, I only need the aspect of truth that is captured by deflationism. For that rather different view about the cognitive function of truth, see Millgram 2009a.

<sup>33</sup>Grover (1992); on her view, "it's true" isn't a predicate applied to a sentence, but rather a prosentence, that is, a grammatical construction analogous to a pronoun. In the simplest case, just like "he," in "he's home," is a substitute for a previously mentioned name of a person, one with the grammatical appearance of a noun, so "it's true" substitutes for a previously mentioned sentence and presents the grammatical appearance of a sentence: "Elephants never forget." "That's true, but this isn't an elephant."

central use) let you recommend choices, and mark them as supported, while screening off the arguments that support them.<sup>34</sup>

When you say that there's an argument for doing something, you haven't given the argument, and you may not have it on hand to give. Oughts allow you to tell someone to do something, to tell him that you're not just pushing him around or making him conform to your arbitrary whims, but without having to tell him just what the considerations which make it more than that sort of thing are. To be sure, this isn't the only way we deploy oughts, just as assertion in bulk of unavailable content isn't the only way we deploy truth. However, merely redundant uses of truth (saying "That's true," instead of repeating what's true) are dispensible, while the content-screening applications are not, and likewise, merely redundant uses of ought (saying "So you ought to do it," when you've just given the argument *for* doing it) are dispensible, whereas applications that screen the argument off from the ought's client are indispensable.

The analogy of function, I want to suggest, extends past the formal resemblance between prosentential truth and those central oughts. Recall those penumbral oughts: When you point out to someone that he ought to do it, and what you have to back that up is the conviction that, if he thought about it more carefully, if he knew more and so on, he would decide to do it himself, you may be sparing yourself not just that story about the upshot of his counterfactual deliberations, but your explanation of why you think that *would* be the upshot. When you point out to someone that he ought to do it, and what you have to back that up is the conviction that, if you polled people who actually *have* tried it both ways, they would recommend doing it, or opt for doing it again, you may be sparing yourself the costs of conducting and presenting the results of the poll. When an experienced driver tells you that you ought to downshift *here*, he hasn't tried to explain what may well be know-how he isn't in a position to articulate. When a legislature passes a law, they don't need to tell you what reasons lie behind it. And, returning to those "*just* oughts," when an Oxford philosopher tells you that you *just* ought not to torture babies, he is exempting himself from explaining why not.<sup>35</sup>

---

<sup>34</sup>Readers with a background in the causation literature may be familiar with its use of the phrase "screening off"; please don't be confused by the accidental similarity.

If you think that oughts are quantifiers over arguments, then you have a choice to make when deciding what is meant by someone who denies that you ought to do it: is he announcing that there's no good argument for doing it, or is he telling you that there's a good argument against doing it? Because the function of oughts is to screen off arguments, someone who makes such an announcement is all too often unlikely to have decided which he intends; and so what on the semantic approach is an urgent question turns out on the cognitive-function approach to be a matter we can leave to those occasions which raise it.

<sup>35</sup>Return briefly to the objection, introduced in footnote 9, that oughts and arguments are mismatched, because there are oughts where there are no arguments, and especially all those "*just* oughts." (Let's register again that the objection doesn't quite make contact with the view on the table, because

## 6.7

Let's return briefly to the discussion of phenomenology kicked off by the Asimovian model; it is confirmation of a sort that once we have the screening function of an ought in view, more of the phenomenology of oughts falls into place.

The quarter century lasting from about 1980 to 2005 saw a debate between internalists and externalists about reasons for action: do your reasons to do something have to bottom out in your motivations, as the internalists thought, or can you have "external" reasons, reasons to do something whether you want to or not?<sup>36</sup> I don't know that there's a point in pursuing the question further, but we can think about the contrast between reasons that *strike* you as external and those that don't. Like the just-really view, that aspect of the phenomenology is also an anticipated side effect of the cognitive function of an ought. If you're going to help yourself to this sort of arrangement, you have to *accept* the advice, recommendation, or instruction as a starting point for your own further inferences, even though the argument behind it is screened off from you. That argument might invoke your motivations, or it might not; all that lies behind the screen of the ought. Implementing the ability to act on such practical conclusions requires that we act even when the connection to our motivations, if any, is invisible to us. Deploying oughts involves our acting on recommendations that *seem* to us like external reasons.

And since we are considering phenomenology, here's one more bit of it we can now speak to: the so-called puzzle of imaginative resistance. Discussion under this heading has been largely directed at attempting to explain an alleged asymmetry between counterfactuals and counterfactuals or counterevaluatives: that it is easy to, say, imagine a baby getting tortured, but hard to imagine that—the facts being just as they are—it ought to be.<sup>37</sup> For my own

---

existential quantifiers over arguments were claimed only to be the *central* case of an ought.) If the argument is on track up to this point, there's much less in the way of such mismatches than one might have thought, since many or even all of those "*just* oughts" now look to screen off arguments, after all.

By a regress argument, *mustn't* there be "*just* oughts" for those practical arguments to bottom out in? That depends on what the correct theory of practical reasoning is: not if oughts don't have to be derived from other oughts, and not if oughts that serve as starting points for arguments are practical observations that are open to further investigation, and not if instrumentalism is correct, and oughts bottom out in desires.

That second option is practical induction, briefly canvassed in Section 6.2, above; for the instrumentalist take on oughts, see Williams (1995a, Chapter 3).

<sup>36</sup>The debate was kicked off by Williams (2001, originally published in 1980); Robertson (2001) provides an overview of the prequel to the debate; Millgram (1996) explains the internalist position as the expression of a substantive view about practical inference; finally, Chapter 4 in the present book locates Williams's internalism in the trajectory of his philosophical career.

<sup>37</sup>Gendler (2000) and Moran (1994). The puzzle is alleged to arise out of a phenomenon that covers all moral thought, or even all evaluative thought, and has been further claimed to reach as far as aesthetic judgment (Stokes, 2006). Obviously a treatment of oughts will not explain asymmetries that do not turn on them. However, it does seem to me natural to extend the account I am about to give

part, I am unconvinced that there is much of an asymmetry, but we are now in a position to explain part of the misleading appearance of one.

Suppose that I tell you that I have something in my right pocket, but refuse to show it to you, or even tell you what it is. If I then ask you to imagine something in my left pocket that is very different from *that*, you'll have a hard time doing so: because my right pocket screens off the object, you don't really know what imaginative task you've been given.

Oughts screen off support, especially arguments. To imagine that the oughts differ, even though the facts do not, is (centrally) to imagine that there is something wrong with the supporting arguments for a practical conclusion, and that there are compelling arguments for a different practical conclusion. All that's hard to do if you can't *see* the arguments, and oughts normally screen them off. That's doubly true when the subject matter is one you don't have a lot of experience with and haven't thought much about (which would be the case for most moral philosophers and their favorite examples, such as torturing babies).

For subjects with which one has experience and to which one has devoted some thought, imagining that the arguments backing up the oughts are no good and that arguments pointing in a different direction are better is just the exercise of a competence philosophers are expected to have. And sure enough, I myself have no problems imaginatively accepting a violently counterevaluative conditional such as, "If the sort of philosophy prized by the Leiter Report were philosophy one *ought* to do, then departments it ranks highly would be good places at which to apprentice oneself." That consists in supposing that the arguments against doing philosophy that way are broken, and that there are good arguments in favor of doing philosophy the Leiter way; as someone who is in the business of considering what would follow from such and such an argument if it worked, I have no problems doing *that*.<sup>38</sup>

---

to cover "good" (provided you allow some variant of the "buck-passing account" briefly recapitulated in Chapter 5, footnote 10). And the effects of a concealed quantifier seem to me continuous with the explanation offered by Stock (2005), on which failure to execute an imaginative task has to do with not knowing how to perform it—in the cases of interest to Stock, due to missing context.

No doubt there is room for other explanations of whatever asymmetries we acknowledge between how we imagine facts and evaluations. For instance, we might bring to bear the discussion of bracketing in Williams (1973b, pp. 210ff).

<sup>38</sup> A theme of this literature is an alleged distinction between merely entertaining a possibility and richly or engagedly imagining it; for example, Gendler takes her larger point—or one of them—to be that imagination is not just supposition. Thus a natural response to my explanation of an apparent asymmetry between our counterfactual and counterevaluative imaginings is that I have misidentified the subject matter: I am perhaps accounting for a feature or side effect of mere supposition, but not of (rich or engaged) *imagination*.

The alert reader will suspect my take on the distinction to be that it is *only* alleged, and further, that I suspect in turn that it is the participants in the debate over imaginative resistance who have misidentified their own subject matter. He's right, but because we aren't provided with anything like a usable criterion for successfully "richly" imagining this or that, I'm not in a position to provide a straightforward argument to that effect, and this is not the place for a more elaborate treatment. (On this point,

Oughts screen off arguments, whereas truth screens off content; if I tell you that someone is very reliable—everything she says is true—and now I ask you to imagine that one of those things she has said is false, only without telling you what it is, you’ll find it just as hard to do as when I tell you to imagine that you ought not to do something you thought you had to, while keeping you in the dark as to what the arguments for doing it had been. Briefly, there’s no asymmetry thus far: if you know not whereof you speak, you may well have a hard time imagining it different. But our intellectual devices make it more likely, in the practical cases, that we don’t know what we’re talking about when it comes to supporting argumentation; in the theoretical cases, it’s what we’re asserting, plain and simple, that we’re less likely to have in view.

## 6.8

Oughts are an intellectual device, and when circumstances change, the devices people use quite properly tend to change along with them. If I am right, our increasingly articulated division of labor has already made today’s oughts somewhat different from their predecessors, and the next agenda item is to explain how and why.

Oughts seem not to have a place, or much of one, in the lives of the other species with which we share our environment.<sup>39</sup> The sand rat is adapted to eat the leaves of the salt bush plant, which contain a natural hypoglycemic; if the sand rat’s diet changes, it develops diabetes. The physiology of the sand rat is shaped to a highly specific environment, and to strategies for survival and reproduction that presuppose this environment.<sup>40</sup> There is no room for such a creature to consider doing things differently—as animals in children’s books do, when they wonder whether to move away from the burrow in the desert to a stylish midtown condo. Oughts are representations of well-supported guidelines for action. The primary payoff of representations is flexibility—you can change them out for other representations—and we pay the costs of developing and maintaining and updating these representations because, one day, it might be the case that we ought to be doing something *else*. Because *this* is how sand rats live, and that’s that, there’s no need for a representation to govern the activities of a sand rat. Since they are not in a position to reap the rewards of migration to a different niche in a different ecosystem, they are hardwired

---

cf. Chapter 8, footnote 8, below.) However, if one after another feature of the imagination falls into place as we examine the workings of the thinner notion of imagination—of “supposition”—eventually that will amount to the requisite argument; you can think of the present discussion as a very first step in that direction.

<sup>39</sup> Working animals may be an exception (Hearne, 1986; Diamond, 1988), but here it is no accident that these animals are trained by human beings.

<sup>40</sup> For instance, like certain other rodents adapted to a halophyte diet, its incisors are adapted to removing salt tissues from *Atriplex halimus* leaves, and its kidneys remove high salt content from its blood by producing a very concentrated urine; see Mares et al. (1997).

to do what their adaptive strategy requires them to do: dig burrows, forage for salt bush leaves, and so on. Generalizing, oughts are needed only by creatures that can change what they do; human beings' oughts mediate flexible responses to a changing environment.<sup>41</sup>

Moreover, oughts are needed only by creatures that marshal the cognitive resources available to one or more of them into multistep sequences of inferentially regimented representations. One sophisticated response to a changing environment, namely, reasoning about what to do, puts a premium on the ability to mark overall upshots of arguments. Sand rats don't deploy oughts not just because they can't change what they do, but because they don't assemble considerations into practical arguments; we do both, and that's been true for at least as long as the historical record.

However, relatively recently, humans have come to exhibit a distinctive and more extreme form of plasticity. One prominent aspect of it is the proclivity to specialize: sand rats do what they do, and that's that; we (many of us) become roofers, political scientists with a game-theoretic bounded rationality slant, computer programmers, violinists, butterfly curators, and many other things, even philosophers. Human beings not only manage highly articulated specialization of this kind, but can manage it repeatedly; over the course of a lifetime, someone might be first a child violinist, then a student of comparative literature, and then a philosopher. To have some shorthand for this, I will say that humans are *serial hyperspecializers*.<sup>42</sup>

We have exhibited much more behavioral plasticity than sand rats for a very long time. But we are starting to exhibit a great deal more than *that*, and we can expect the pervasiveness and depth of division of labor to increase. Just as tires had to improve when highway speeds increased, the devices that subserve behavioral plasticity will have been improved as the extent of specialization-driven flexibility increased—or anyway, that's what we should expect to find. Those improvements may well be unobvious: tires today still

---

<sup>41</sup> Even what Williams (1985, p. 142) dubbed “hypertraditional” societies—hidebound cultures where nothing ever seems to change, and where no provision is made for the possibility—deploy shoulds and oughts. But because the mores of such societies differ from one another, evidently they too display the flexibility that the present point requires.

<sup>42</sup> For a somewhat less terse presentation, see Chapters 3 and 4. When I say “relatively recently,” I mean that although specialization has been around for quite some time, until no more than a few centuries ago it was a marginal phenomenon. Whewell (1847, vol. ii, p. 149) informs us that according to “a writer of contemplative theology in the twelfth century . . . *Practice* . . . includes seven mechanical arts; those of the clothier, the armourer, the navigator, the hunter, the physician, and the player.” If there are only a handful of different types of experts (notice that there are claimed to be seven, but apparently only six come to mind), then even if Richard of St. Victor is overlooking a few of them, the brevity of his list shows it to come to us from a time and place in which there is not all that much in the way of division of labor.

Even now, serial hyperspecialization is not found everywhere. Some societies have been too poor to afford the panoply of experts on whom we depend, or too chaotic to allow their inhabitants to specialize, or just too regimented to allow people to switch specializations.

*look* very much like the tires of the 1960s. Even if our oughts *sound* like older oughts, we shouldn't assume that their workings remained constant as the extent to which we specialize increased.

What does specialization that goes as deep as it does in our own case involve? Serial hyperspecializers normally produce specialization-specific forms of representation. Discipline-specific vocabularies are just the tip of this iceberg, and these are not always scientific or academic vocabularies: a while back, I was up on my roof, talking to a roofer, and I literally couldn't understand what he was telling me. Serial hyperspecializers develop systems of standards, priorities, and guidelines that frame decision and choice within particular disciplines. Mastering such a specialization often means learning to deploy distinctive modes of argumentation (keeping it close to home, remember that only philosophers use transcendental and state-of-nature arguments), along with the standards of success to which arguments of those types are held. Because the cost of mastering those forms of representation, those systems of standards, those techniques of argumentation, and so on is heavy, they are not normally accessible to disciplinary outsiders.

Disciplines that agents haven't mastered—which is always most of them, because it takes many years to achieve such mastery, and lives are only so long—are replete with arguments that other agents can't understand. We need to take advice from experts; that's why our society supports so many of them. We can't know, or even understand, what all those experts know.<sup>43</sup> People who field the challenges posed by a complex and mutable environment through a strategy of flexible specialization are bound, if they can, to adopt a way of marking recommendations and decisions as well-supported, for clients who are not in a position to assess that support. So we should expect oughts to have been reappropriated and adjusted to meet the novel demands imposed by specialization. To get a sense of what those might be, let's do a little compare and contrast.

As we earlier modeled oughts, they were important because they screened off arguments you could in principle understand, but which had become unwieldy to the point of being unmanageable. However, this is a good time to acknowledge that there have always been other uses of the device. Parents have to tell their children what to do, often on the basis of considerations the parents understand, and that the children won't. Moreover, sometimes what necessitates parental oughts is that even when the children in some sense can understand the argument, they often won't take the point: children have bad judgment, they can't tell what counts as a good argument, and even when they can, they exhibit what Susan Hurley once called “epistemic akrasia.”<sup>44</sup>

---

<sup>43</sup>The phenomenon is vividly portrayed by Hardwig (1985); see also Hardwig (1988) for a followup discussion.

<sup>44</sup>Hurley (1989, pp. 130–135, 159–170); I'm grateful to Jan Murphy for the observation.

Or again, back in the day when most people were either peasants, warriors, or priests, it was common to pretend that the great many oughts serving to mark the traffic laws of a society were the deliverances of a god. When those rulers or priests imposed their traffic laws, they didn't generally insist that there was any argument waiting in the wings, over and above "The god said so."

Experts' oughts typically mark arguments you couldn't understand, not without becoming an expert yourself; once more, you couldn't possibly acquire the expertise of all the different sorts of experts upon whom you will need to rely. In this way, the specialists' ought is continuous with the parental ought, in that it is a way to mark an action as something you'd better do, even when you don't understand and can't appreciate the arguments which support that assessment: as a lay adult, you are often in the position of a small child.

However, it seems to me that expertise is nonetheless making what I was calling the central oughts even more central. Because clients can't themselves assess the support for expert advice, even when they have the time and inclination to do so, they insist on heightened transparency; this typically takes the form of having one expert assess another expert's arguments, and that in turn means that arguments are more often called for, where once a mere pronouncement on the basis of experience, say, would have sufficed.<sup>45</sup> The further and older uses of an ought that I mentioned a moment ago have themselves been affected. For instance, we now expect of our lawmakers not to pretend that a god told them what laws to enact; we expect our traffic laws, literal and metaphorical, to be designed; we expect (overoptimistically, perhaps) the designers to solicit expert advice in the course of their work: maybe advice the arguments for which only traffic management specialists would understand. We even expect parents to provide the arguments behind their oughts, when their children demand them.

These adjustments notwithstanding, room for improvement no doubt remains, in part because we haven't been thinking about the design of the device. This isn't the place to try out proposals, but let me gesture at three ways oughts could be revised, just so that you have a sense of the sort of thing I have in mind.

Unlike "good," which contrasts with "better" and "best," oughts as we have them do not come in degrees.<sup>46</sup> So there is a tension between the flat assurance conveyed by an ought and the experts' own understandings of the varying ways and varying likelihoods that their arguments can go wrong. To be sure, the experts cannot explain how problems may arise in the supporting

---

<sup>45</sup>There is perhaps a further motivation: One mark of expertise is precisely the ability to explain one's reasons clearly to nonexperts. Or anyway, whether it's correct or not, that's what the nonexperts are likely to think; if experts can't explain their reasoning, how is anyone supposed to be convinced that they *are* experts? (I'm grateful to Sara Montanari for this observation.)

<sup>46</sup>I was reminded of this by Andrew Sepielli.



arguments without explaining the arguments themselves, which normally they cannot do. But—and here’s a first and near-trivial illustration of what it would be to adjust the widget that we have—perhaps oughts can be supplemented with flags that signal how worried, in the expert’s own view, one should be about defeaters, and maybe even what *kinds* of defeaters to be on the alert for.

Next, Bernard Williams used to work with an extended version of the notion of conjunction: Truth *agglomerates*, in that, if  $p$  is true, and  $q$  is true,  $p \wedge q$  is true; imperatives do not agglomerate, because each of the commands, “Open the door!” and “Leave the door closed!” can be perfectly in order, even though “Open the door and leave it closed!” is not. Oughts Williams thought to constitute a less obvious case; however, they would agglomerate if “You ought to  $\phi$ ,” and “You ought to  $\psi$ ,” entailed “You ought to  $\phi$  and to  $\psi$ ”.<sup>47</sup>

I’ve suggested elsewhere that “true” has come to serve as a generic certification that can be passed across disciplinary boundaries (and that it is frequently a substitute for the nuanced certifications deployed among the specialists themselves).<sup>48</sup> Oughts analogously have come to certify guidelines for action provided to nonspecialists, and so likewise must be made usable by people who not only don’t have the supporting arguments on hand, but who wouldn’t have the competences to control the arguments if they did. That truths agglomerate—that is, that you can conjoin truths indiscriminately, without worrying about how the arguments supporting them interact—lightens the cognitive burden on their users enormously. Leaving to one side whether oughts *have* agglomerated, we would do well to ask whether experts’ oughts can and should be *made* to agglomerate.<sup>49</sup>

There are evidently tradeoffs. On the one hand, an ought that can be agglomerated with any other ought is user-friendly: such oughts allow a client of several different experts to solicit guidance from each of them, and to assemble all that guidance directly into a plan of action. Oughts from different experts—say, a chef and a physician—that cannot be agglomerated will leave clients wondering: is the recipe safe if I’m taking my medication? On the other hand, the costs of producing oughts that agglomerate are likely to be high,

---

<sup>47</sup>Williams convinced himself that tragedies arise because oughts do not in fact agglomerate; see Williams (1973b, Chapters 10–12).

<sup>48</sup>See Chapter 4 in the present volume.

<sup>49</sup>Old-school metaethics is sometimes motivated by the thought that ethics is characterized by disagreement, and science by convergence; moral antirealism is repeatedly offered as an explanation of the asymmetry. Both oughts and truth, when used by experts to convey advice and information to outsiders, will mask in-field disagreement. We are used to getting our science from specialists, and we generally think of ethical advice as not dispensed by technically qualified experts. (That’s partly because we have the habit of classifying expert advice as not moral or ethical, and partly because having specialists dispense information historically predates having specialists dispense advice.) Thus there’s less in the way of an asymmetry than, say, Bernard Williams believed—maybe even *enough* less to warrant asking whether we really need the type of explanation the old-school metaethicists so desperately tried to supply.

and even insuperably high. Normally, it will mean having the experts preemptively cross-check among themselves. But there are many different types of possibly relevant expertise, along with communication barriers between disciplines, which means a *lot* of often difficult cross-checking. Perhaps another candidate modification to traditional oughts would be adding indicators as to whether and when they can be agglomerated.

Third and finally for now, where the older screening device was depicted as having the side effect of making the supporting arguments unavailable (in our Asimovean just-so story, when the older generation vanishes, its practical arguments are lost forever), in principle the community of experts remains available for follow-up queries. We are used to marking some oughts as prudential, others as moral, others as legal, and so on; in the past this practice served to restrict the domain of quantification to prudential arguments, moral arguments, etc.<sup>50</sup> But similar keywords now serve a new and very important function, that of telling you which sort of specialist to call when you notice something going wrong.

## 6.9

When you have a repertoire of tools and techniques, it's the easy problems—that is, the problems that are easy, given those tools and techniques—which get solved first. Semantic analysis began by plucking the low-hanging fruit, but those fast, exciting advances are long past, and the leftover problems on which we're now working are likely to have remained on the agenda precisely because we need different tools to make headway on them. Metaethics in particular is visibly deadlocked and high-entropy; if we can see another approach, it's certainly worth trying out. With any luck, there'll be low-hanging fruit.

The alternative approach I've been demoing focuses on giving design characterizations of intellectual devices and descriptions of the job they do within a larger cognitive system and intellectual environment. This newish style of treatment can give us a sort of philosophical understanding that semantic analysis doesn't provide.<sup>51</sup> For instance, remember why we decided that the

---

<sup>50</sup>See Chapter 5, footnote 19, in the present book. The older discriminations can themselves be given a new function, that of allowing an expert to convey what sort of reasons lie behind his advice. If he's a lawyer (so he's telling you what you legally ought to do), is he telling you what's legally *safe* (what it's legally prudent for you to do)? Or what legal ethics requires? And he could also use these markings to signal that another professional's advice is motivated by a particular sort of consideration that you would regard as inappropriate—say, his fee structure.

<sup>51</sup>Newish, but not unprecedented: Craig (1990) invites his readers to imagine a society, pretty much like ours, but lacking the concept of knowledge. He then talks his way through the problems this society would have, and what its members would be missing: a generic certification for information, one that tells the person you're giving it to that he can go ahead and use the information, and also pass it on to someone else, with the same certification. (That the certification is generic and transmissible is as

penumbral oughts belonged together with the central ones: they were further content-screening indicators of support for a choice, and it wasn't surprising that we'd have let them share a linguistic marker with the central oughts. When we took the semantic analysis approach to metaethics, we thought we were abstracting away from the content and substance of ethical theory, to a shared sort of meaning. But there wasn't a shared meaning; rather, there was a shared function, namely, screening off supporting content.<sup>52</sup> When it falls into place, cognitive function turns out to be a frame that can make the semantics look much less important: once you understand what an ought *does* for you, you can treat what it *means* almost as an afterthought.<sup>53</sup>

---

much as to say that whether a user can go ahead and insert the information into whatever arguments he's constructing doesn't depend on anything special about *him*.) After you think about it for a bit, you're supposed to see that any society that didn't have such a certification would invent one, because it's so inconvenient to get along without it. And when they did, it would look, walk, and quack a lot like the concept "knowledge." Now, if it looks like a duck, and it walks like a duck, and it quacks like a duck, it's a duck; if it looks, walks, and quacks like knowledge, it's knowledge. "Knows" just *is* that sort of certification, and once you understand that, you understand just about everything that a philosopher needs to know about knowledge.

Notice that while the details of such a certification are likely to vary from society to society, for this sort of functionality—and this is a philosophically interesting move—details don't necessarily matter. For most of the twentieth century, epistemologists were in the business of tweaking definitions of knowledge to get them to line up with exotic intuitions about Gettier cases, or about telepaths and psychics who don't know they're telepaths and psychics, or about landscapes dotted with papier-mâché barns. . . . But if Craig is right, the definitions don't matter, and the exotic intuitions don't matter; after all, there are presumably many ways of implementing such certificates, each of which would differ from the others in minor respects. All that really matters is the social function of knowledge, and we come to understand that by coming to see how, if you didn't have a form of certification that worked this way, you'd have to invent it.

This is a good occasion to mark very briefly how design characterizations of intellectual devices like oughts differ from the characterization of mental states provided by analytical functionalism. Analytical functionalism was the model for the so-called Canberra Plan, on which platitudes about a subject matter are used to construct a theory that identifies, for example, mental states with (whatever occupies) various slots within the theory. (For a less abbreviated explanation, along with an argument that the Canberra Plan is unworkable, see Millgram 2009a, Chapter 10.) Unlike the present approach, analytical functionalism purported to provide a semantic analysis of mental state terms. And unlike the present approach, analytical functionalism was oriented toward function only in name. That is, unlike the present approach, it took itself to be responsible to how people ordinarily speak and what they ordinarily think. But just as you would not attempt to reconstruct the design and workings of airplanes from platitudes about airplanes, you should not attempt to figure out the functionality of the nuts and bolts of the mind by collecting platitudes; these will at best tell you how we *imagine* our minds to work.

<sup>52</sup>I've been allowing that the different oughts have meanings, likely different ones, but for all we know, there are oughts that do their job while meaning nothing. The logical positivists' insight, which we seem to have lost, was that not everything means something. We should keep an eye out for those cases also, and avoid trying to give semantic analyses of bits of language that are semantically inert.

<sup>53</sup>At this point the historically minded reader may be wondering if it's really legitimate to contrast semantic analysis with cognitive function treatments. Wasn't there an older school of ordinary language philosophers which popularized the slogan, "Meaning is use"? (See, for example, Wittgenstein 1998, Section 43, or the final remark of Section 421.) By their lights, when I explain the cognitive function of a term like ought, I *am giving* a semantic analysis. And I have had it suggested to me that linguists, for example, don't contrast semantic and use-based analyses nearly as sharply in their own practice.

“Metaethics” has traditionally been the special-purpose name reserved for the metaphysics of ethics, morals and, more recently, practical reasoning and rationality. And in the past (even the very recent past), metaphysics was repeatedly allowed to degenerate into a science of invisible and impossible entities. Since metaethics was just metaphysics, only of a special subject matter, it too degenerated into a science of invisible and impossible entities. The most prominent of these was a force, conventionally called “normativity,” that was supposed to account for the difference between what one did and what one was supposed to do. Twentieth-century metaethics consisted largely in producing theories of what this force was.

But there is a better way to understand metaphysics, namely, as intellectual ergonomics. Metaphysics tells us how we have to understand the world, if it is to be thinkable—and probably this is the only charitable way to make sense of the enterprise in retrospect. Different ages have put very different spins on that problem, but the time has come to treat it as a straightforwardly practical question, one with two sides: How must we represent our world in order to be able to get an intellectual handle on it? And how can we reengineer our world so as to improve the intellectual handle we have on it? When we come to understand oughts as ways of marking support, while screening that support off, we are no longer tempted to think of them as representing a peculiar nonphysical force: oughts are no more and no less than an ergonomic choice. And so there are no puzzles about what theory to give of an impossible entity. There are just design issues, having to do in the first place with the question of what design description best fits the inherited version of the device.<sup>54</sup>

When we are doing intellectual ergonomics, coming to understand the design of a device as it has worked for us in the past is most importantly a

---

Here it matters that we keep the traditions straight; different philosophical (and professional) traditions tend to give different meanings to the term “meaning.” I pointed out earlier that, in the analytic tradition, definitions are substitution licenses; when you give a Wittgensteinian story about how the word “five” is used (ibid., Section 1), you can’t substitute the story for the word. “Meaning is use” is true only on the ordinary-language tradition’s use of “meaning” (I imagine one that is to be underwritten by describing how a word like “meaning” is ordinarily used). In the sense the analytic tradition has given to “meaning,” meaning isn’t use.

One further contrast: Hilary Putnam took Wittgenstein to have the view that meaning is a coarse grid laid over use. But as I’ve been describing it, use is a coarse grid laid over meaning. The various oughts may *mean* quite different things, but they’re all *used* the same way.

<sup>54</sup>Have we really bypassed the old-school metaphysics of normativity? That depends in large part on whether design analyses and ergonomic justifications can be understood to be “positive” (in the sense of Comte, as reviewed in Chapter 5). The conventional view is that design analyses are “teleological,” that teleology—except where a human or divine agent is actually the designer—involves forces that would have to take the same type of metaphysical treatment as normativity, and that since naturalistic treatments are unworkable, teleology involves a commitment to pre-Enlightenment forms of thought: it might as well just be superstition.

However, I expect that design analyses can be accounted for functionally as a design feature of our own cognition, in a manner that is continuous with the account I have been giving here. If I am right, that can defuse the charges of illegitimacy; I hope to develop this train of thought on a further occasion.

prelude to redesigning it. What we are most interested in is what design options are available and what considerations can be adduced to underwrite both the continued use of an inherited device and our choice of a replacement for it. There's more to metaethics than "ought", but if similar accounts fall into place for the other central notions ("good" and the like), we're on our way to a different *sort* of metaethics: an *applied science*, one that is in the business of reengineering intellectual devices like these.<sup>55</sup>

---

<sup>55</sup>I'm grateful to Carla Bagnoli, David Copp, Maneesh Modi, and Constantine Sandis for helpful conversation, and to Chrisoula Andreou, Sarah Buss, Heather Douglas, Christoph Fehige, Sam Fleischacker, John Hardwig, Michael Huemer, Kim Johnston, Buket Korkut-Raptis, David Schmitz, Yonatan Shemmer, Aubrey Spivey, Cynthia Stark, Dustin Stokes, audiences at the University of the Saarland, the University of Modena, the University of Arizona, the University of Utah's Kaffeeklatsch and, finally its Philosophy Club for comments on earlier versions of this material. Thanks also to the University of Utah for travel support, and to both the University of the Saarland and the Arizona Freedom Center for working space and support.

## Lewis's Epicycles, Possible Worlds, and the Mysteries of Modality

If there's an obvious lesson to draw from twentieth-century metaphysics, it's that reductionism doesn't work. Reductions purport to show that things of one kind (minds, for instance, or physical objects) are *really just* things of another kind (dispositions to behave in certain ways, or patterns in the flux of sensation, respectively), the proof of the pudding being that you could give a scheme for paraphrasing away anything you might say about the first kind of thing into assertions about the second kind of thing. Over the course of the century, reduction after proposed reduction failed, and failed systematically—that is, for the same small family of reasons. So it's something of a surprise that the position we find serving as perhaps the most visible reference point within turn-of-the-millennium metaphysics is yet another form of reductionism: this time, from *modal* facts to configurations of “possible worlds.”

David Lewis called his view “modal realism,” and as I just remarked, it has come to serve as something of a reference point. But it is also seen as an extreme and idiosyncratic position, and my real concern, for which I mean to use Lewis's view as a stalking horse, is the much more widespread idea embedded in it, that statements with modal content can be paraphrased into the vocabulary of possible worlds. For this reason, and partly as well because Lewis was such an elaborate system builder, I am going to divide up my discussion of his modal reductionism into two treatments. The other one takes up his realism proper, that is, the insistence that possible worlds, global ways things might have been, exist (*really, really* exist, in just the way anything else does); it addresses itself to a specifically metaphysical motivation, namely, the conviction that modal facts are, on the face of it, spooky, unnatural, and have no proper place in the physical world.<sup>1</sup> Here I will focus on a different but equally important motivation for reduction, the worry that you do not understand

---

<sup>1</sup> Millgram (2009a, Chapter 11).

the very words you are using. (When reductions are motivated in this way, they tend to get thought of as conceptual analyses.) The divide-and-conquer approach means that I need to ask for a slightly unusual concession from my readers: If you think of an objection to the argument, and you do not see it addressed here, please check the companion treatment before deciding that I have overlooked one or another of the resources provided by Lewis's metaphysics.

I'll begin by introducing the notion of modality and by describing Lewis's modal reduction. I will take time out to explain what is required of a reduction, to document that Lewis accepted the requirements, and that he has been widely read as doing so. At that point I'll lay out one of the arguments that buried some earlier twentieth-century reductionisms, and I'll construct a variant directed against Lewis's.

Lewis's reductionism has been criticized on occasion, and on the way in I will distinguish the argument I am using to kick off the discussion from those more familiar complaints. Once again, my intent is to use Lewis as a stalking horse, and to show a widely shared assumption to be false: that ordinary claims with modal content can be paraphrased in the vocabulary of possible worlds. Because the extant criticisms share this assumption, they do not do the work that we need.

Although Lewis is in certain respects an unusual representative of the take on modality I want to contest, his position is perhaps the most thoroughly articulated and inventively worked out in the neighborhood. So I will consider at some length whether his systematic metaphysics supplies the materials for a rejoinder to my argument. As it will turn out, it does not, but, more importantly, this stretch of counterargument will put in place two secondary conclusions: one regarding the provenance of the materials upon which we draw when we assess counterfactuals, and the other having to do with the fragmentary nature of those materials. If these conclusions are correct, they are important guides for investigations of modality down the road.

Recall that we will have been focusing on a reduction that is motivated as conceptual analysis, and leaving what philosophers nowadays tend to insist are the properly metaphysical considerations for elsewhere. It is nonetheless important to see how these apparently different modes of treatment fit together, and in Section 7.10 I will take up the connections between these two philosophical motivations. I will rehearse the reasons that modal subject matter does not allow a graceful retreat to the fallback position eventually adopted by so many other twentieth century reductionist programs: that supervenience will give us everything we wanted from a reduction, but couldn't get. I will explain why the missing fallback preempts what has elsewhere become a standard move, namely, to insist that what is on offer in one's theory is not meant as conceptual analysis, but is rather nonreductionist metaphysics or ontology.

I'll wrap up by asking what the big-picture lesson of the argument is. Discussions of modality have come to take the possible-worlds way of thinking about it for granted. The philosophical uses of the possible-worlds apparatus depend on the assumption I intend to refute, that we can give possible-worlds renderings of ordinary claims with modal content. I will conclude that it is time for a dramatic shift of key in philosophical thinking about modality.

## 7.1

Like many metaphysically deep phenomena, modality is hard to define in a noncircular and illuminating way, and for now I'll introduce the notion with a gesture rather than a definition. Think about claims like these:

1. I *could* have been a contender.
2. Ducks don't smoke cigars, but they *might've*.
3. If Clinton *had* had a decent haircut, he *would* never have been elected Governor of Arkansas.
4. Water *has* to be H<sub>2</sub>O.
5. If you *were* King Mithridates, you *would have* foiled your enemies' evil plans.<sup>2</sup>

What they have in common is modal subject matter (marked, in these sentences, by the highlighted *could*, *would*, *has to*, *might've*, and so on). To a first approximation, the subject matter includes the possible and the necessary, counterfactuals (represented by sentences of the form, 'If such-and-such *had* happened, so-and-so *would've* happened'), dispositions (because they normally embed counterfactuals: if the glass is fragile, then if you had dropped it, it would have broken), and perhaps other things of the same ilk (whatever ilk that is) besides.

Modality is philosophically mysterious. When you assert that ducks smoke cigars, I know just what to look out for to confirm or disconfirm your claim. The sentence lists the relevant items (ducks, cigars, and smoking), and all of these are things that I'll recognize when I see them. But when you tell me that ducks *might've* smoked cigars, there's a part of the sentence—the "*might have*"—that leaves me wondering what to look for and how I would recognize it if I found it. Consequently, it's easy to start worrying that I don't know how to say what would make such a sentence true (except trivially, by saying that it would be true if ducks indeed *might've* smoked cigars), and if you think of being able to give a sentence's truth conditions as a touchstone for whether I know what it means, then it looks like I don't actually understand sentences that contain *mights*, *would'ves*, and so on.

---

<sup>2</sup>Props to Kazan (1954), Gerber et al. (2008), Putnam (1975b), and Housman (1965, p. 90).



Lewis's modal reductionism was meant as a philosophical response to the mystery, and it was a natural move to make given the historical background. Modal logics—formal calculi intended to represent possibility and necessity—had gained in popularity and prestige after model-theoretic treatments of a handful of the axiomatic systems seemed to put them on a mathematical footing with more traditional Frege-Russell logic. In these so-called possible-worlds semantics, each possible world (intuitively, each way things might have been) was represented by the sort of model of the universe of discourse that would have been used to model a theory in standard first-order logic. “Possibility” was then rendered as a sentence's being satisfied at some world-model, and “necessity” as the sentence's being satisfied at all world-models.<sup>3</sup> Once you do that, “Ducks might've smoked cigars” gets paraphrased as “There are some possible worlds in which ducks smoke cigars.” The “might've” has been replaced by the unmysterious “there are some,” plus, of course, the possible worlds—and although these might sound mysterious themselves, we do ordinarily talk about different ways things might have been.

There's nothing to get philosophers to jump on a bandwagon like a formal representation with mathematical cachet, but there was more to it than that: model-theoretic semantics for modal logics made logicians comfortable with modality, in very much the same way that Cantor's work had made mathematicians (and philosophers) comfortable with infinity; where before the subject matter had seemed rife with contradictions and incoherences, now one had an internally consistent way of talking about it, and one that gave you an almost mechanical way of answering questions about (for instance, and don't worry if you don't know what this means) iterated modal operators.<sup>4</sup> The philosophers quickly came to accept “true in all possible worlds” and “true in some possible worlds” as explications or paraphrases of “necessary” and “possible,” respectively, and Lewis extended the way of speaking to counterfactual conditionals (again, sentences like, “If so-and-so were the case, then such-and-such would be the case”). If the possible worlds are embedded in a similarity space (that is, if we say that worlds are “nearer” to one another if they are more similar), then the counterfactual conditional is true if, roughly,

---

<sup>3</sup> More carefully, at *accessible* world-models; the semantics included a specification of which worlds are, again intuitively, “visible” from which other worlds; different specifications of the accessibility relation model different axiomatic systems. Formal work on modal logics began with an attempt by C. I. Lewis to represent the non-truth-functional conditionals he needed for his phenomenalist reductionism, about which more below. The model-theoretic approach was developed in work by Saul Kripke and others; for textbook surveys, see Hughes and Cresswell (1996), Forbes (1985, Chapters 1–2), and Priest (2001, Chapters 2–4).

<sup>4</sup> We should always remember that having a representation of the claim that *p* (even an elegant mathematical representation) does not count as a philosophical argument for *p*. (Not *at all*: merely to *say* that *p*, no matter *how* you say it, is not an argument for *p*.) However, I'm grateful to Tim Bays and Alasdair MacIntyre for pressing me not to forget the reasons that possible-worlds semantics became so popular in the first place.

in nearby (or the nearest) worlds in which its antecedent is true, its consequent is true. So to determine the truth of a sentence like “If hybrids were cheaper, more people would buy them,” notionally travel out to the “nearest” (i.e., most similar) possible worlds in which hybrids are cheaper, and if, in those worlds, they get bought by more people, the counterfactual is true.<sup>5</sup>

Then Lewis took the step that turned him into the Alexius Meinong of the twentieth century, which was to treat the possible worlds as *things out there*.<sup>6</sup> This struck many other philosophers as outlandish, since possible worlds are, Lewis acknowledged, spatiotemporally isolated from each other—meaning that no matter how far you travel, or how long you wait, you’ll never get to another way things might have been, and so you can’t ever *inspect* one of these entities. But Lewis took the move to be justified by its philosophical payoffs: once you have these entities, you can treat the paraphrase of “necessary” as “true in all possible worlds” as a *reduction*: not just useful idiom borrowed from the mathematical logicians, but a description of what’s really going on when you (correctly) say that something had to happen. You can likewise explicate “actual” as an indexical, one which picks out the world the speaker is in, in something like the way that “I” picks out the person who says it; likewise for “possible” and “true in some world.” And, very importantly, likewise for Lewis’s proprietary semantics for counterfactuals: what it *is* (*all it is*) for it to be true that if Clinton had had a decent haircut, the folks in Arkansas would never have elected him, is that, in the nearest possible worlds in which he had a decent haircut, they didn’t. The mysteriousness of modality is addressed: the “would” in “they wouldn’t have elected him” picks out an object, or, in this case, a class of objects, in just the way that “Clinton” and “haircut” do: namely, the relevant class of possible worlds.

Lewis’s view is straightforward-sounding, but it has its complications, and because one of them will be important shortly, let me describe it now. Take a counterfactual such as “I could have been a contender.” On the way of talking that Lewis made into a reduction schema, that gets paraphrased as “In some (at least one) possible world, I am a contender.” But does that mean that *I* am an inhabitant of this other world—that I am both here and in some other place, one that’s *so* very far away that it doesn’t even count as far away?<sup>7</sup> Lewis

---

<sup>5</sup>See Lewis (1973, 1986). That “roughly” is there to sidestep complexities introduced when the ordering of possibilities by “nearness” does not have maximal elements (see Lewis, 1983–1986, vol. 2, pp. 6–10).

<sup>6</sup>Lewis (1986, pp. 98f) took the trouble to distinguish his own doctrines from Meinong’s, something that would not have been necessary if the resemblance had not been hard to miss. The differences he pointed out do not defuse the characterization: to be the Meinong of a given philosophical period is not the same thing as having precisely Meinong’s views.

<sup>7</sup>Specialists will be aware that that way of putting it is a detour around a rather different-sounding argument (Lewis, 1986, pp. 198ff), and the back and forth that it generated in the literature. I am taking the expository shortcut because, in my judgment, the moves on both sides were badly motivated. For further discussion, see Lewis (1973, pp. 38f).

handled this difficulty by adopting what he called “counterpart theory”: the contender in that other possible world is not, exactly, *you*, but someone (your counterpart) who is relevantly similar to you, both intrinsically and terms of the role or position he or she occupies in that world. So a counterfactual like this one ends up taking on a further layer of paraphrase: “In some possible world, I have a counterpart who is a contender.”

So much for our brutally rapid introduction to Lewis’s metaphysics of modality.<sup>8</sup>

## 7.2

If your philosophical motivation for a reduction is the worry that you don’t understand your own vocabulary—in this case, the *coulds*, *woulds*, *musts*, and so on—surely that worry is only assuaged if the paraphrase you provide doesn’t contain the vocabulary you suspect you don’t understand: if I do not understand the “outgrabe” in Lewis Carroll’s “Jabberwocky,” it does not help to tell me that it’s what mome raths do, because I do not understand “mome raths” either. (In fact, there’s an alternative, which is, more or less, to exhibit the relations between the terms in the problematic vocabulary and how the group of interrelated terms is collectively related to unproblematic objects or notions; this approach has come to be called the Canberra Plan, and I’ll take it up in due course.) A reduction of one sort of vocabulary to another sort of vocabulary commits itself to specifying how statements apparently about the former can be paraphrased without residuum into statements about the latter, and without importing, explicitly or surreptitiously, the suspect vocabulary into the paraphrase. That does not disallow using the concepts and vocabulary one is trying to replace in the course of *identifying* the paraphrase, but those concepts and vocabulary had better not end up in the reductive paraphrase itself. A reduction also commits itself to eliminating black boxes, by which

---

<sup>8</sup>Not only am I leaving the full discussion of Lewis’s realism to another treatment, I am also going to put to one side a further and very obvious problem. Earlier on, I excused myself from giving a proper definition of modality, and Lewis himself, in the book-length presentation of his view, said almost nothing about what modality is (even though one of the advertised benefits of the view is an analysis of modality). That’s very peculiar, because if you were going to give an argument that As are really Bs, you’d think that such an argument would have to turn on an independent characterization of the As; without such a characterization, Lewis *could* not have a good argument for his own central claim: that modal facts are really just facts about the possible worlds.

For instance, one view people often have is that “metaphysical modality” is a different sort of thing from “epistemic modality,” and an account of the former needn’t include an account of the latter. (In sentences in which, for instance, “might be” and its relatives mark epistemic modality, they are paraphrasable, with some qualifications at the margin, by variants of “for all you know.” So, “He might be home by now” would get rendered as “For all I know, he’s home by now.”) Lewis seems to think otherwise, because he promises a treatment of the epistemic modals as well (Lewis, 1986, pp. 27ff); but without an explanation of what you meant by “modality” in the first place, how could you know whether such a treatment was owed?

I mean devices in the conceptual analysis whose inner workings it cannot exhibit.<sup>9</sup>

When we get to Section 7.10, I will, as promised, lay out the play of forces that committed Lewis to a reduction. But reductionism is so unfashionable nowadays that even compelling reasons may fail to dispel the worry that I am being interpretively uncharitable. For that reason, before we go any further, I want to document both that Lewis himself accepted the demand and that his interlocutors have read him as accepting it. (If this isn't something that's bothering you, you're welcome to skip ahead to the next section.)

As I mentioned earlier on, Lewis's position has already attracted criticism, and the relevant complaints turn on two related worries. One was that Lewis had no explanation of what made his other possible worlds *possible*: since they were, on his account, just like this world, weren't they merely more *actual* worlds?<sup>10</sup> The point of the objection is that an unexplained grasp of possibility must in fact be concealed in the appeal to possible worlds, that if our concern is that we do not understand modality, we should be equally concerned that we do not know what a possible world is, and that the obligations of the reductive paraphrase have not been met. That is, these objections take it for granted that Lewis is to be read as attempting a reduction. (Bear in mind that our present concern is not whether the criticisms are effective, but only how they were motivated.)<sup>11</sup>

When Lewis addressed this complaint, he nicely exhibited the reductionist shape of the project. Lewis reiterated that his indexical understanding of

---

<sup>9</sup>Here is a toy example of a black box: "A sentence containing modal vocabulary really means whatever David Lewis says it means." Even though that last phrase contains no *woulds* or *coulds*, unless I know what Lewis's gloss is, I'm not in a position to tell whether it really is a content-preserving paraphrase, and I'm not in a position to tell whether or not it reintroduces the dicey terms and concepts we are concerned we do not control. (For all I know, what Lewis says it means is full of "coulds" and "woulds.") That was, again, a toy example, but appeals to black boxes play a large and disreputable role in much recent philosophy, as when moral philosophers invoke the preferences agents would have in idealized circumstances; for discussion, see Millgram (2005a, pp. 69, 74, 85 n. 39). To foreshadow, Sections 7.5–7.8 will be devoted to a black box in Lewis's treatment of modality, namely, similarity relations among possible worlds. Another of the black boxes—the context-dependence of such similarity orderings, thus also of the counterpart relation—will come in for discussion in the notes.

<sup>10</sup>Complaints in this ballpark include Shalkowski (1994), Plantiga (1987), Sider (2003, Section 3), and Chihara (1998, pp. 280, 286; see also p. 80, which fields a similar complaint regarding Lewis's "worldmate" relation). Divers (1997) discusses the interplay between the two classes of objections; Divers (2002, Chapter 7) provides an overview of the back and forth around the former.

There have been other attempts to show attempted reductions of modality to rely on undischarged modal notions. For instance, MacBride (2001) primarily discusses Jubien and focuses on the modal content of metaphysical categories such as "property," "object," and "matter": an object *cannot* be instantiated, matter *must* be spatially and temporally located, and so on.

<sup>11</sup>Indeed, there are discussions that more or less explicitly describe Lewis's project as I do. Sider (2003) and Plantiga (1987) have also called Lewis a reductionist—though in Plantiga's case, the label is attached to a somewhat different point. Divers (1997, p. 144) is willing to use the word. Sider, Divers (2002, p. 106), and Chihara (1998, pp. 81f, 207n) agree with me on the substantive claim—though Chihara does not use the term himself.

“actual,” on which it picks out *this* world, entailed that other worlds could not be actual, and it must have seemed to his opponents that he was making a point of missing the point of their objections.<sup>12</sup> But his response exhibited his reductionist commitments. Since the object of the exercise was to reduce away the modal notions, those notions should not appear in a characterization of the items to which they were being reduced: for unreduced mere possibility to reappear as a feature of those other worlds would mean that the reduction had failed.<sup>13</sup>

The second family of objections had it that which possible worlds (or occupants of possible worlds) there are determines what comes out true when you quantify over them; but how can you say which there are other than: the *possible* ones?<sup>14</sup> As before, the force of these objections was that the possible-worlds rendering surreptitiously deploys unreduced modal concepts and primitively modal opinions, and so it does not reduce them. So it again exhibits the agreement of Lewis’s interlocutors that he was to be read as committed to a reduction of modality.

And as before, Lewis confirmed that commitment in his own counterarguments; he responded by attempting to specify the range of possibilities as ways of recombining elements of the actual world.<sup>15</sup> As before, our interest is not in whether his rejoinder was successful, but in what it shows about what he was trying to do. By attempting to address the complaint on its own terms, he was accepting those terms: because the range of possibilities operated in his account as a black box, he needed to show that it could be opened up, and its workings reconstructed using only modality-free materials.<sup>16</sup>

---

<sup>12</sup>Lewis (1986, pp. 97ff). Note that Lewis’s response predates the publications listed in footnote 10, which makes the belated and reiterated complaints an indication of just how hard it was to swallow.

<sup>13</sup>Here’s some further textual confirmation of the stance. Lewis treated it as an objection to competing views that they had to help themselves to “primitive modality” (Lewis, 1986, pp. 151–155, 167f, 179f), and he stated flat out that “[p]rimitive modality is bad news” (p. 242). Compare Lewis (1999, p. 208).

<sup>14</sup>See, for instance, Melia (2003, pp. 114f), Lycan and Shapiro (1986, p. 358), Divers and Melia (2002), where the claim focuses on alien properties, Bremer (2003), a rebuttal treating individuals the way Lewis had wanted to treat the worlds, Divers and Melia (2003), refelding the initial complaint, only now about possible individuals. Chihara (1998, pp. 282ff) considers the worry that a reductionist analysis of modality, constructed with an eye to making one’s modal views come out right, is not reductionist *enough*. Paseau (2006, p. 724) briefly entertains the reply (which he does not endorse) that the range of possible worlds should be taken to be the right one, whether we can specify it or not. Divers (2002, Chapter 7) is a recap of previous exchanges.

<sup>15</sup>Lewis (1986, pp. 86–90).

<sup>16</sup>This is as good a place as any to speculate about why Lewis called his view “modal realism,” not (as I’ve been characterizing it) “modal reductionism.” Even if realism and reductionism are opposites of a sort, you can be both realist and reductionist if you are a realist about one kind of thing, and a reductionist about another; normally, as Putnam once remarked, reductionist projects presume that whatever they are reducing their problematic items to *are* real. (Thus, a phenomenalist is normally a realist about sense data.) Reductionist projects are often epistemologically motivated, and when they are, they follow the epistemic order: that is, what you’re reducing *to* is whatever is easier to know. (If you think that material objects are to be reduced to sensations, that’s because you take sensations to

As announced, I intend to sidestep this back and forth, because it takes for granted the assumption I mean to contest, and which I am using Lewis's view to reconsider. You can believe that understanding what a possible world is requires a primitive grasp of possibility, and that quantifying over possibilia gives you real results only if you have prior opinions about what is possible, while *still* agreeing that the content of a possible-worlds paraphrase is that of the modally loaded claim it purports to render. Instead, I propose to consider counterfactuals. These comprise the hardest-working division of our modal apparatus; ordinary people do not have much in the way of strong opinions about what is possible or necessary, but they have a great many opinions, opinions on which they rely in their day-to-day lives, about what would have happened if . . . Accordingly, the objective of the coming argument will be to demonstrate that counterfactuals cannot be paraphrased into the vocabulary of possible worlds. If the argument is successful, it will show that Lewis's reduction fails, but, and more importantly, it will do so by showing that an especially important class of modally laden claims cannot be given a possible-worlds paraphrase.

If the worry you are trying to defuse is that you don't understand those *coulds*, *woulds*, and so on, you have to show that (perhaps with some adjustments at the margins) you do understand the very *coulds*, *woulds*, and so on that people actually utter. And this is (one reason) why the reduction Lewis was considering was a modality-free paraphrase of, specifically, ordinary modally laden claims about ordinary things.

### 7.3

Lewis used to complain that the most frequent response to his view was an incredulous stare, but that an incredulous stare isn't an argument.<sup>17</sup> We *will* get an argument; before we start in on it, however, here's a bit of warmup.

Early on in the back and forth of twentieth-century metaphysics and epistemology, Roderick Chisholm laid out the objection that I mean to use as a

---

be the sort of thing of which one is immediately aware, and so because you take yourself to be solving the skeptical problem of how one ever knows anything about material objects.) When reductions are not epistemologically motivated, they're often driven by a prior picture of what exists on its own, or in its own right. (If you think that biological systems are to be reduced to physical systems, that's because you take physical objects to exist in their own right, but biological objects to exist only by grace of the configuration of physical objects.) Lewis's modal reductionism follows neither the epistemic order (again, a typical complaint about the view was that you can't observe other possible worlds) nor the usual views about what exists all on its own (it's not standard to think of the might-have-beens as the substances). So Lewis may not have found it natural to describe himself as a reductionist because the starting and ending points of the reduction he was proposing were in these respects unusual.

<sup>17</sup>The characterization made it into a widely circulated collection of humorous "proofs that *p*" (compiled by Hartry Field); see Lewis (1986, p. 133) and Lewis (1973, p. 86) for the complaint.

pattern, the occasion being reductionist formulations of phenomenalism. On phenomenalist views such as those of C. I. Lewis,<sup>18</sup> ordinary material objects are really nothing over and above patterns in the flux of actual and possible sensation, and the content of a statement about such an object, for instance, “The cat is asleep on her pillow,” is given by many, many counterfactuals along the lines of “If I were to have a sensation like *so* [the one I normally describe as the feeling of turning my head], I would come to have a visual sensation like *that* [the one I would normally describe by saying that it’s of a cat sleeping on a pillow]”—that is, by counterfactuals about the sensations.

Leave to one side obvious difficulties about stating these counterfactuals without mentioning material objects (pillow, cat, . . .) in the course of picking out the sensations; waive the objection (since C. I. Lewis himself happily conceded it) that there are too many such counterfactuals to list—perhaps infinitely many. Chisholm focused instead on the hard-to-miss fact that pretty much any such counterfactual is only true other things being equal (or *ceteris paribus*, in philosophers’ Latin). For instance, while it’s true, other things being equal, that if I were to turn my head (or have that head-turning sensation), I would see (or seem to see) the sleeping cat, nonetheless, if someone were to smash me over the head with a sledgehammer, just as I was starting to turn, then I *wouldn’t* see the cat. Now we might try to accommodate such facts by adding extra clauses to the counterfactuals: “If I were to have a sensation like *so*—and no one were to smash me over the head with a sledgehammer—I would come to have a visual sensation like *that*.” But, and this is the problem, the *ceteris paribus* clauses are stated in the vocabulary of material objects (this one mentions a sledgehammer), that is, in the very vocabulary that the reductionist phenomenalist is committed to paraphrasing away.

To be sure, you could try rendering the bit about the sledgehammer into further counterfactuals about sensations, but these will require their own *ceteris paribus* clauses, once again couched in material-object vocabulary: If I were to have the picking-up-the-sledgehammer sensation, I would feel the heft . . . unless a nerve were cut . . . and so on down the regress. The reductionist promise of paraphrasing away the vocabulary of material objects turns out to be empty.

Here’s one more quick dress rehearsal, another argument on the same pattern due to Hilary Putnam. Logical behaviorism is the view that statements framed in psychological vocabulary can be translated into (or at least analytically entail) statements framed in a nonpsychological behavior-description vocabulary; for instance, if you’re in extreme pain, then you’re more likely to scream. Now behavior is produced by indefinitely many psychological states

---

<sup>18</sup>Clarence Irving, not to be confused with David Lewis; I’ll always give the earlier Lewis’s initials, and a freestanding “Lewis” will always be David Lewis. For the phenomenalism, see C. I. Lewis (1956), and C. I. Lewis (1946); for the objection, see Chisholm (1948).

jointly, and in particular, sometimes we suppress behavior prompted by one psychological state on the basis of another. Putnam memorably called the characters in his illustration “super-spartans”: extreme pain makes super-spartans want to scream, just like anybody else; however, the super-spartans want, even more strongly, not to show the pain, and so they hold it in. The relevant entailments or translations only hold if other things are equal, and there are always going to be *ceteris paribus* clauses that have to be stated in psychological vocabulary. As before, the reductionist promise of logical behaviorism turns out to be empty.<sup>19</sup>

## 7.4

With these model arguments before us, let's turn back to David Lewis's modal reductionism. Take that last item in our short list of sample modal claims: If you *were* King Mithridates, you *would have* foiled your enemies' evil plans. On Lewis's way of rendering the content of that counterfactual, it comes to something along these lines: in the nearest possible worlds in which Mithridates is your counterpart, he foils their plans. So to assess the truth of the sentence, we notionally travel out to the nearest worlds in which you are King Mithridates (in which he's your counterpart), and if, in those worlds, he foils his enemies, it comes out true.

Now scratching the surface of even that quite minimal description reveals a great many embedded modal facts. For instance, a king is someone such that, were he to issue a command to his ministers, it would be (*ceteris paribus*) obeyed; this is, after all, constitutive of being a king, or anyway was so until the advent of merely ornamental royalty. A king is someone such that, were he to die, one of his children would succeed him. A king is someone such that, were he to wear a ridiculous garment, no one would dare to comment on it. —*Are* these really necessary? Well, if too many counterfactuals like these are not true of you, we'll start to wonder whether you really are a king.<sup>20</sup>

On the modal reductionist account, paraphrasing away these further counterfactuals means picking out the counterparts of those counterfactual Mithridates in still further possible worlds—possible worlds that are near to the ones they occupy. To check the truth of the *first* counterfactual, we notionally

---

<sup>19</sup>Putnam (1975a); the argument to follow, which is modeled on the two arguments of which I have just reminded us, is a slower and easier-to-follow version of Millgram (2009a, Section 11.5).

<sup>20</sup>We're using Lewis to think through our more general concerns, but alert readers may be worrying that perhaps the argument we are embarking on has too narrowly defined a target. On the views of other theorists about modality, we do not need to identify our counterparts in other possible worlds; rather, we simply stipulate that it is *you* we are considering. So note that we can concede the point without affecting the way the argument plays out: the example would work equally well if the antecedent of the counterfactual was “If you were a king . . .”; the problem, either way, is determining whether you have what it takes to be a king. See Kripke (1980) and Kaplan (1979) for background.



travel out to the nearest possible world(s) in which you are Mithridates, and see if you foil your enemies; but to check whether he really *is* the counterpart required by the antecedent of the conditional (among other things, whether he really is a king), we have to consider the truth of *another* counterfactual (his orders would be obeyed), and so we notionally travel out from some Mithridates-containing world to the nearest possible world(s) to it in which your counterpart's counterpart gives an order, and we determine whether it is obeyed; also, we travel out to the nearest world(s) in which your counterpart's counterpart dies, and we determine who succeeds him . . . and so on. (See Figure 7.1 for help visualizing the spheres of “nearby” possible worlds, and the ways they are positioned in the natural representation of the example; evidently, this would be an all-too-appropriate occasion to revive the word “epicycles”.)

But who are *those* counterparts? I adapted the counterfactual whose content we are trying to paraphrase from a bit of Housman's *A Shropshire Lad*, so allow me that Mithridates' counterparts have to be, at any rate, human beings: surely that's how Housman would have meant it! But now, scratching the surface of a human being yields just as many modal facts as scratching the surface of a

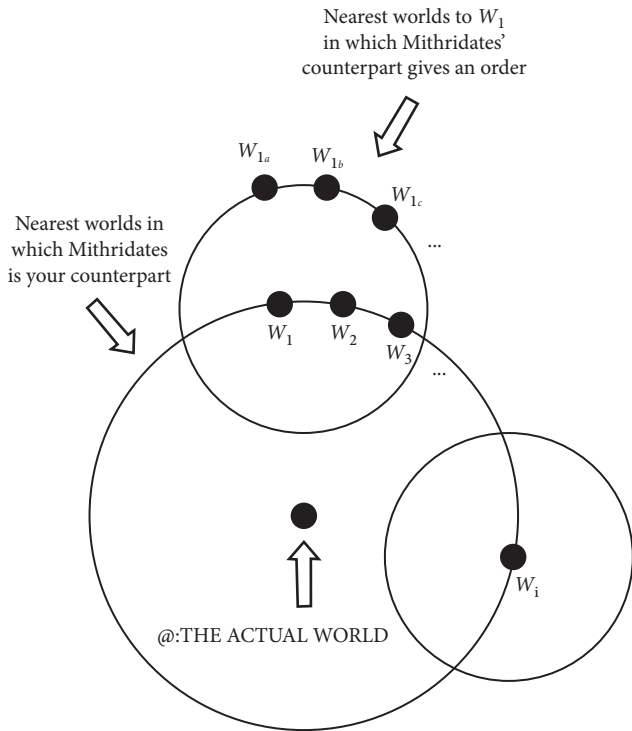


FIGURE 7.1 Lewisian Epicycles.

king. A human being is something that wouldn't vanish if you were to breathe on it. (Things that look like people but vanish when you breathe on them are *ghosts*.) If the putative human walks and talks normally, then it has to be true, if it is a human being, that, if you asked him any of a great number of unexpected questions, a suitable answer would come to his mind most of the time. (In other words, he has dispositional mental states; these will become important in due course.) So we owe a possible-worlds paraphrase for these modal facts as well.

When you try to paraphrase ordinary modal talk into possible-worlds talk, using Lewis's recipe for doing that, you find yourself marching down a regress, and one that looks very similar to the regress that refuted phenomenalism. But let's make the problem that Lewis is facing conform fully to the shape of our historical paradigms. Bracketing worries that maybe, even in principle, there's no such thing as a modally thin concept or description or object, suppose we *have* such a description—not "king," but as close as we can get without building in unwanted counterfactuals—which we use to pick out objects in nearby possible worlds occupying a suitably kingly role. Then the counterfactual we started out with will have to be rendered roughly along the following lines: objects in nearby possible worlds that serve as your counterparts and also satisfy this description (they spend time in a throne room, carry a sceptre, and so on) foil their enemies . . . *unless* (and here comes the extended *ceteris paribus* clause) they wouldn't be obeyed if they gave orders, or they would be laughed at if they were to wear a ridiculous garment, or anyway, *too many* conditions like the foregoing hold; or they would vanish if breathed on, or they have no dispositional mental states, or . . . (The standard logical behavior of *ceteris paribus* clauses is on display here: it is obvious that the exceptions are not going to run out, no matter how far you extend the list.)

The force of the *ceteris paribus* clause is *not* that if a king were, for example, disposed to vanish if breathed upon, he wouldn't manage to foil his enemies. Rather, it excludes these cases from the scope of the claim, so as to make it match the sense of the ordinary assertion: someone who says that a king would foil his enemies does not mean to commit himself one way or the other as to what happens in such oddball cases. If we are to track, even roughly, the semantic intentions of ordinary speakers, the paraphrases that are the cash value of the modal reduction must contain such *ceteris paribus* clauses. These deploy the very vocabulary—modal vocabulary, this time—that the reductionist is committed to eliminating. The uneliminated modal freight is not just modally thin possibility and necessity, but counterfactual conditionals (which is, to remind you, why the argument to this point is not a recapitulation of previous objections to Lewis's reductionism). Once again, the reductionist promise, this time of possible-world treatments of modality, turns out to be empty. Lewis's modal reductionism runs aground on the very same argument that defeated some of its prominent twentieth-century reductionist predecessors.

## 7.5

The argument is straightforward enough, but the lessons about modality that I am after will emerge only from the back and forth of objections and replies, and let's start in on those now.

Recall that the proposed reduction of counterfactuals invokes a similarity space in which we are to imagine the possible worlds as embedded: the more similar they are, the closer. The enterprise is to reduce modal statements to configurations of possible worlds, and the worlds' similarity relations surely count as an aspect of their configurations. So if the similarity relations between possible worlds can be made to do the work of *ceteris paribus* clauses, we can drop the redundant clauses, in which case the reduction goes through as is, and the argument I have been developing against it fails.

Let's spell out the objection a little more slowly, and while we're at it, let's vary the example, since I observed earlier on that there is always *more* packed into a *ceteris paribus* clause: it is not a human if, were it to hear the code word, it would transform itself back into a bug-eyed monster and return to the mother ship. Mithridates' nearest counterparts, we pointed out, had better be human beings, but "human" is a modally thick concept. If the *ceteris paribus* clauses that exclude such possibilities contain modally thick vocabulary (as does this one), the reduction has not succeeded.

However, the objection runs, Mithridates-like chunks—I'm putting it this way so as not to beg any questions about counterparthood—of possible worlds similar to this one don't transform themselves into bug-eyed monsters; that in itself would make such a world very strange, from our point of view. And so worlds close to (which again means similar to) those worlds also do not contain Mithridates-like chunks that transform themselves into bug-eyed monsters. After all, since Mithridates' counterparts don't transform themselves into BEMs, would-be counterparts of theirs who did would be quite dissimilar to *them* (leaving aside for the moment worries turning on the near-transitivity of "similar"). So, since the nearby worlds *are* the ones in which Mithridates' apparent counterparts *do* have the right modal properties, we can drop the clauses of the reduction whose point is to add that guarantee. And so the reduction does go through after all: thus the objection we are canvassing on Lewis's behalf.

To address it, we need to stop and think for a moment about Lewis's conception of a similarity ordering over possible worlds. There are two relevant constructions we might put on it: On the first, similarity would be an objective or metaphysical relation, having to do with patterns of what Lewis called *universals*: the natural properties picked out by our deepest theory of the world (which will be, if materialism is true, "something not too different from present-day physics, though presumably somewhat improved"). Alternatively, possible worlds could count as more or less similar in virtue of whether the

person uttering the counterfactual in question takes them to be; that is, the similarity relations over possible worlds could be a matter of the similarity judgments attributable to one or another ordinary speaker.<sup>21</sup>

Obviously, what you are committed to reducing depends on what philosophical work the reduction is supposed to do for you. Here we are considering a reduction motivated by the worry that you do not understand your *would've*s and *could've*s and *might've*s (recall that I am considering the specifically metaphysical motivations for such a reduction elsewhere); on the assumption that you are an ordinary speaker of an ordinary natural language, that means the reduction has to capture the content of ordinary speakers' ordinary counterfactuals about ordinary objects.<sup>22</sup> Now of course people can be wrong (indeed, very, very wrong) about what would have happened, if. . . . But treating the physics-style universals as providing the underlying measure of comparative similarity makes ordinary speakers out to be wrong in the wrong ways; the systematic errors attributed to them by putting an objective or metaphysical construction on similarity shows that their semantic intentions are not being tracked. For instance, we allowed that ordinary speakers do not mean to commit themselves to what would happen to a Mithridates who was, in virtue of the modal facts, not really a human being but a bug-eyed monster in disguise. A world in which something that looks like Mithridates turns into a bug-eyed monster and returns to the mother ship (or, for that matter, is not obeyed by ministers, or is laughed at for wearing ridiculous garments, etc.) might well be quite similar to ours at the level of basic physics.<sup>23</sup>

---

<sup>21</sup>For discussion of universals, see Lewis (1999, Chapter 1; the quote is on p. 37). Divers (2002, p. 123) is an example of a philosopher taking it for granted that appeal to ordinary judgments is the appropriate alternative.

Now, there may be convergence between the two prongs of the fork; for instance, Lewis appropriated Davidson's uses of the Principle of Charity, and in particular he held that the naturalness of an interpretation (of someone's psychology or utterances) constrained its eligibility. (That is, we interpret someone as meaning *green* by "green," rather than *grue*, because green is a more natural property than grue; see Lewis 1983–1986, vol. 1, Chapter 8, and Lewis 1999, pp. 45–55.) So we should not assume that the dilemma is a clean choice, and I am constructing the argument to follow with that in mind.

<sup>22</sup>A popular nonrealist view was that possible worlds were one form or another of representation; one of Lewis's arguments for his own view was that such representations couldn't capture the truisms of ordinary modal discourse, as in his treatment of "ersatzism" (1986, Chapter 3). But if that inability rules out ersatzism, then, by parity of argument, it's not an option, for Lewis anyway, to opt for the possible-worlds dialect and what can be represented in it, and just give up on the aspects of ordinary modal discourse that it can't accommodate. Tracking ordinary users' semantic intentions matters.

<sup>23</sup>Allowing for the sake of the argument that "basic physics" is counterfactual-free, and yes, those are scare quotes. Why? Take a look at Wilson (2006).

Rayo (2013, pp. 134f) has observed that the normal way for a Lewisian to pick out counterparts involves pointing to other counterparts; for example, part of what makes so-and-so Humphrey's counterpart is that he was born in the counterpart of South Dakota. He notes as well that we don't have a way of picking out South Dakota (or Humphrey himself, for that matter) using only that alleged vocabulary of basic physics.

Since ordinary speakers mean to exclude these cases from the range of commitments they assume when they advance our sample counterfactual (on the basis of their similarity judgments: space aliens are just too *weird*), universals are not suitable for capturing the content of their ordinary counterfactuals.

To make the similarities that drive the ordinary counterfactuals out to be simply *physical* similarities—as a way of showing the similarity metric to be a matter of what is objectively, willy-nilly *there*, as opposed to something derived from what’s resident in speakers’ psychologies—requires that a big difference, as far as some counterfactual is concerned, be a big *physical* difference. But that’s not the way it normally goes. As some of the Anscombian have recently emphasized, in physics there’s no answer to the question What comes next? when it is asked in the register of natural history. (What comes next depends on whether the flower we are examining is, say, trampled by a passerby; it is mere first-this-then-that.) Whereas in biological thought, and for that matter when we are considering intentional action, the question, asked in that distinctive register, is in place. (Next, the buds unfold into small pink blossoms.) In a similar vein, a character in a recently mainstreamed graphic novel points out that “a live body and a dead body contain the same number of particles.” A big biological difference can be a very little physical difference, and, conversely, a big physical difference can amount to, say, a very small economic difference.<sup>24</sup>

We are not quite done with the topic, and I will return to it below, when the time comes to consider supervenience as an alternative to reductions. In the meantime, let’s agree to require that when we unpack the black box of comparative similarity, its inner workings must be made out in terms of the psychologies of ordinary speakers: what do *they* take to be more or less similar, when they are considering alternative ways things might have been?

---

<sup>24</sup>Thompson (2008, p. 41), and Moore et al. (2005).

Here’s another way to say it: of course a world in which one encounters bug-eyed monsters in disguise must be, intuitively, somewhat different from our own, and maybe even very different in ways that one or another science would recognize. But the task we are considering on Lewis’s behalf is that of coming up with a counterfactual-free way of saying how to count those differences, in order to show that counting differences that way preempts the problem posed by *ceteris paribus* clauses for his proposed reduction. If the similarities and differences are made out at the level of special sciences like cultural anthropology (a science that might have a good deal to say about what makes aliens *alien*), it is very hard to believe that they prove free of counterfactuals. If they are nonetheless to be made out at the level of physics, then we need a way of rendering similarities and differences that appear in the descriptions of, for example, cultural anthropology into the vocabulary of physics. Recall that I began the chapter by reminding us that one clear lesson of twentieth-century metaphysics is that such reductions don’t work.

Now a typical response on the part of Lewisians is to appeal to the role of context—especially, conversational context—in foregrounding *some* physical dimensions of similarity. However, then the work of selecting the features in virtue of which we judge worlds to be more or less similar is taken up by ingredients which are either psychological or logically on a par with psychological facts. This is to transfer us to the second horn of the dilemma we are examining. N.B.: There are further difficulties with appeals to conversational context; I will register one of them in footnote 36; for another, see Millgram (2009a, Section 7.7).

Human psychologies are small-finite, and this has as a consequence that judgments of the comparative similarity of whole possible worlds must—with unimportant exceptions—be constructed from reactions to local features of those worlds. The reason is not just that most possible worlds are too complicated for any human mind to survey adequately. Because there's not enough cognition to go around, we generally have to be choosy about where we invest thought and deliberation. Human beings' practical perspectives on the world are quite local; consequently, our well-considered judgments are also, almost without exception, quite local.

Let's distinguish *thoughtful* from *thoughtless* judgments of comparative similarity. We are interested in counterfactuals because we rely on them so heavily in our intellectual and practical lives, and we are not unreasonable to do so. If judgments of similarity underwrite our counterfactuals, then a reliable counterfactual must be tied to well-considered similarity judgments. Counterfactuals whose contents are given by judgments of similarity that have had no thought put into them are worthless. (As the computer scientists say, *GIGO*: Garbage In, Garbage Out.) So we can disregard merely thoughtless judgments of similarity.<sup>25</sup>

Possible worlds are overall or global ways things might be (or might have been). Counterfactuals are enormously important for real life: we make decisions on the basis of our opinions about what would have happened, if. . . . And for that reason, we (often) do our very best to be thoughtful about our counterfactuals and to get them right. So whatever opinions underwrite those counterfactuals must be commensurately deliberate, that is, we must be about as thoughtful about them as about the counterfactuals themselves. With very rare exceptions, we do not bother having thoughtful opinions about the comparative similarity of global or overall possibilities.

We can confirm this assessment by considering one of those exceptions that prove the rule: physicalism, for present purposes the view that everything is *really just* configurations of physical objects and properties. Take the version of it that amounts to a supervenience claim: if the nonphysical objects and properties were different, the physical objects and properties would be different, too. One way to recast this sort of physicalism in Lewis's conceptual apparatus is this: possible worlds in which nonphysical similarities and differences are not tightly tied to physical similarities and differences are extremely different from our own world. In other words, physicalism itself amounts to a comparative similarity judgment that takes whole possible worlds as its objects, and

---

<sup>25</sup>On occasion, Lewis appealed to our willingness to produce thoughtless opinions about the similarity of items—such as cities—that are too complicated to understand; if cities, he was suggesting, why not possible worlds (Lewis, 1973, p. 92)? This was a mistake on his part, because it entails a violation of the conservation of intellectual effort; you can't get usable, reliable counterfactuals out of thoughtless, off-the-cuff answers to questions like "Which city is more similar to Seattle, San Francisco, or Portland?"

one that has had thought put into it by physicalist philosophers. But now that we have an example of an actual similarity judgment that takes entire possible worlds as its object, it is obvious how *unusual* this sort of assessment is. I don't mean that physicalism is a minority view among nonphilosophers, but rather that most people (philosophers included) hardly ever invest any thought at all in similarity judgments of this generic logical type.<sup>26</sup>

However, we do work up thoughtful opinions about more local aspects or features of alternative possibilities. So if Lewis's semantics for counterfactuals deploys the psychological resources we have available, the global comparisons must be assembled out of the local ones.

There is a second observation about those local judgments of comparative similarity to be taken into account. Similarity has the role, in Lewis's account, of what I earlier called a black box; for the reduction to succeed, we would need to show that the black box does not conceal modal notions on which we are tacitly relying. Now, because our thoughtful local judgments of counterfactual similarity are cognitively expensive—we are, again, only interested in similarity judgments that have had thought put into them—and because there is only so much attention and deliberation to go around, we generally form such judgments only when their objects are of interest to us. But we are creatures who live modally saturated lives, and almost all of the objects of our attention are counterfactually thick. Almost everything we care about (or worry about, or strive for) has a usually elaborate counterfactual aspect. We've already seen a couple of examples: many people admire royalty, and whether someone is royalty is largely a matter of what counterfactuals are true of him. Being a person is largely a matter of what counterfactuals are true of you. And it is easy to continue in this vein: People move to big cities because of the cultural opportunities—the things they *could* do, which they know they won't have time to do. People are distressed or joyful because of what *almost* happened to them. People care as much or more about whether they *can* be victims of violent crime as whether they are *actually* the victims of violent crime.<sup>27</sup> So almost all of the local similarity judgments we are considering will have as their

---

<sup>26</sup>This is a good point to field an objection that has no doubt occurred to the alert reader: that the counterfactuals true of Mithridates are true of him in virtue of his modally flat, this-worldly, *physical* properties. If it *were* true of someone that he might morph into a bug-eyed monster on receipt of the signal, surely features of his anatomy that account for it will turn up in the autopsy video.

Suppose you don't want to make physicalism, as just introduced, necessary: after all, most of us think that there might be someone, or something, who walked and talked just like we do, and morphed into other shapes for good measure, but who was hollow inside; it's merely that such a possibility is very strange. Then the coordination of at-a-world physical and modal properties is local: that's what it's like *around here* (in the big and variegated possible world universe, as ordered by similarity). But the reduction we are being offered is not meant to work only locally.

<sup>27</sup>Ruth Chang alerted me to the emotional importance of near-misses; the point about how important inviolability is to people comes from Nagel (2007), which addresses itself to the centrality of the modal good in the design of political institutions. I am told that Jerry Fodor has also noticed the point about the cultural opportunities.

objects not modally or counterfactually flat properties or states of affairs, but instead objects or properties or options characterized in just the modally laden way that the reduction under consideration is committed to eliminating.<sup>28</sup>

In the way of thinking we are working our way through, modal and counterfactual claims are to be understood by way of a picture in which the ways things might have been are *global* (they are ways *everything* might have been, all at once) and *modally* or *counterfactually flat*. The similarity ordering that is meant to defuse the argument we constructed in Section 7.4 must thus have these global, flat ways things might have been—possible worlds—as its objects. The judgments of similarity we actually have on hand are, we saw, local rather than global and have modally and counterfactually thick intentional objects. So if Lewis's reduction is to work, there must be a way of assembling judgments of the latter sort into similarity orderings of the former sort. In a moment, I will proceed to consider whether this is possible in principle, but first I want to register a bit of nuance.

I am trying throughout to stay as far as I can within the spirit of Lewis's proposal, once again, because I think that in doing so I can hit a target of more general interest; the reason I think so is that similarity mostly functions as a placeholder, a representative for whatever does the job of accounting for the truth values of counterfactuals. When he entertained the notion of similarity orderings, Lewis was ambivalent about the idea of writing down such things. He took the similarities in question to be vague, incomplete, and context-sensitive, and even suggested that any way of making them completely precise might thereby misrepresent them.<sup>29</sup> Both our own objectives and Lewis's hedges mean that there's a delicate balancing act to take note of here. Lewis took "similarity" to be a theoretical notion, one that we can adjust to match our control of counterfactual conditionals. This means that you can't always take responses based on eyeballing to pick out its contours. On the other hand, the relation is used to control our assessment of counterfactuals, and so it must be deployed. That means that we must have a pretty good idea of what is, in the relevant sense, more similar to so-and-so than what else, in most situations of ordinary concern. I am going to rely on this point about our competence in what follows, as I introduce a handful of very straightforward claims.

---

<sup>28</sup>The worry that a Lewisian similarity ordering will turn out to depend on the very sorts of modal judgments that it is trying to explain appears as early as Fine (1975, p. 455).

<sup>29</sup>See Lewis (1986, p. 21), and Lewis (1983–1986, vol. 2, pp. 181f, on "tailoring"); Lewis (1986, p. 254). "Imprecise [comparative similarity] may be; but that is all to the good. Counterfactuals are imprecise, too. Two imprecise concepts may be rigidly fastened to one another, swaying together rather than separately, and we can hope to be precise about their connection" (Lewis, 1983–1986, vol. 2, p. 6). See also a remark on "the questionable assumption that similarity of worlds admits somehow of numerical measurement" (p. 12), and related discussion on p. 163, as well as in Lewis (1973, pp. 50–52, 67).



## 7.6

In Section 7.3, I described Putnam's argument against behaviorism as another instance of the argument pattern that we deployed against the possible-worlds reduction of modality. Lewis paid close attention to that argument, and his own views in philosophy of mind were constructed as a response; the technique he developed has been adapted by his followers into a popular and systematic approach to the problems of metaphysics (the so-called *Canberra Plan*). Although he seems never to have considered it himself, the Canberra Plan is the best way Lewis had to take up the task we have just outlined, and (once again because I think we can learn a number of more general lessons about modality from it) I will first describe the Canberra Plan and how it might be adapted to the construction of counterfactual-free similarity orderings over possible worlds. I will consider two related obstacles to producing such constructions. Then I will take time out to draw morals about the logic and function of *ceteris paribus* clauses, both in counterfactual contexts and in general.

Recall that Putnam's objection to logical behaviorism exploited the idea that connections between mental states and behavior are mediated by other mental states. (What you do, when you want something, depends on what else you want.) Lewis was happy to allow that, and his variant of functionalism—"analytical functionalism," to distinguish it from Putnam's computation-oriented formulation of functionalism—explicitly accommodated that idea.<sup>30</sup> The device he used required, first, collecting the platitudes of "folk psychology," then conjoining them into a single, very long sentence, and finally replacing the psychological vocabulary with variables bound by existential quantifiers. Doing that ("ramsifying") gives you a theory in which platitudes like

6. If someone is in pain, he tends to scream  
and
7. If someone has a very strong desire not to scream, he tends not to  
scream

reappear as segments of that very long sentence and look something like this:

8.  $(\exists x)(\exists y) \dots$  if someone is in  $x$ , then he tends to scream, & if someone has  $y$ , then he tends not to scream, & if someone is in  $x$  and has  $y$  then. . . .

The idea is to take all the theoretical relations at once and treat the theoretical entities as the occupiers of slots in the matrix that the relations jointly constitute. A theoretical entity is picked out, more or less, as the occupier of slot number  $n$  in the theory, and rearranging these characterizations into

---

<sup>30</sup>Lewis (1999, Chapter 16), and Lewis (1983–1986, vol. 1, Chapters 6, 7, and 9).

definite descriptions allows you, in principle, to eliminate the problematic theoretical vocabulary, which means that the technique lives up to the formal requirements of a reduction. In fact, the reductive paraphrase is never executed, but these multiple-slot definite descriptions allow you to identify the innocuous (in this way of thinking, the purely physical) states or properties which as a matter of fact occupy the slots in the theory: folk psychology may be full of beliefs and desires, peculiar mental states on which (you are concerned) you may not have a satisfactory philosophical grip; but once you know that, in human beings, only such and such neurological states are related to each other and to perceptual inputs and behavioral outputs in the way the ramified folk theory says, you can pick these out as what the beliefs and desires in fact are.

The “woulds” and “coulds” with which we started cannot be smoothly inserted into the template of analytical functionalism, but assembling a theory of similarity in the way the model suggests is more promising—though in order to explore the application of the Canberra Plan to Lewis’s modal reduction, we are going to have to allow ourselves to depart from the letter of what is now a widely applied recipe. The approach is evidently a way—evidently, the *only* way—to make a similarity ordering over possible worlds serve Lewis’s reduction of modality, and here’s what it would take: We collect the psychologically available judgments of similarity from the “folk”; we conjoin these to obtain the folk theory of (counterfactually relevant) similarity. We elicit from this theory—perhaps via the recipe’s step of replacing theoretical terms with bound variables, or perhaps in some other way—a matrix of similarity relations between the different objects, properties, states of affairs, and so on. (Recall that these objects, states of affairs, etc., for the most part bear modally and counterfactually thick characterizations.) Finally, we identify the modally or counterfactually flat *physical* properties, states of affairs, and so on that in fact occupy those slots in the matrix. We are then to reconstruct a reduction-compatible rendering of the similarity ordering over possible worlds using only these modally or counterfactually flat elements.

It is hard to believe that this is a program we will actually get around to implementing. But is it possible in principle? If it is, then perhaps, even if we cannot exhibit the Lewisian reduction of modality to possible worlds, such a reduction can be known to be possible in principle. And if it can, perhaps almost as much philosophical work can be done with that conclusion as if we had the reduction on hand. As the reader no doubt expects, I am about to argue that it is not possible, even in principle.

## 7.7

A von Neumann-Morgenstern utility function is a way of summarizing a particular agent’s preferences, and it is possible to do so provided the agent’s preferences satisfy a handful of actually quite demanding consistency

requirements.<sup>31</sup> Over the past decades, an objection has crystallized to many of the uses made of expected utility theory: that it is almost inevitable that human beings have both *too few* and *too many* preferences to have utility functions. That is, as a matter of psychological fact, a human being's preferences are far too sparse to induce a utility function; moreover, they are not sufficiently consistent with one another to induce a utility function (or a usable approximation to one, or even a usable range of them). I am about to use this train of thought as a model for considering what can be had by way of similarity orderings over possible worlds. To anticipate, if our comparative local similarity judgments are too sparse to be assembled into comparative similarity judgments whose objects are global or overall alternative possibilities, and if they are mutually inconsistent in ways which prevent them from being so assembled, that will amount to a decisive criticism of the Canberra-Plan approach to similarity orderings over possible worlds. A reminder: we are going to be as interested in *why* the approach fails as in the brute fact that it does.<sup>32</sup>

First, we have *too few* local similarity judgments. To say what I mean by that, I first need to remind you of the contrast, which we invoked in passing in considering Mithridates' modal properties, between occurrent and dispositional mental states. Occurrent states are what's explicitly before your mind, and dispositional states are those that would come to mind when suitably prompted: you're not always thinking, "My name is \_\_\_\_" (so it's not an occurrent belief), but if someone were to ask you your name, the answer would be immediately forthcoming (so it *is* a dispositional belief). Most of anyone's mental states are dispositional, and because occurrent psychological states are so few and far between, the judgments to which the specification of a similarity relation over possible worlds must help itself will inevitably be by and large dispositional. Your *actual* similarity judgments (the ones that, in Lewis's way of thinking, you have in *this* possible world) are quite sparse indeed. A moment ago, were you actively thinking of any of them at all?

A dispositional psychological state is a state that you have counterfactually.<sup>33</sup> Consequently, unpacking the black box of similarity in Lewis's reduction means bringing to bear judgments of local similarity that ordinary

---

<sup>31</sup>See Mandler (2001) for an overview. Here and below, "consistency" is used in the technical sense in which it figures into such discussions, and without any endorsement on my part of the implicit suggestion that these are conditions our preferences *ought* to satisfy, and that "inconsistency," in this sense, is a fault and grounds for complaint. For an argument to the contrary, see Millgram (2005a, Chapter 10).

<sup>32</sup>Lewis himself used to assume that people have (or approximately have) utility functions; that was perhaps an allowable error when he was writing, but, now that Daniel Kahneman has been given his Nobel Prize, is so no longer. Over and above Kahneman's work with his late collaborator, Amos Tversky (Kahneman et al., 1982), representative contributions to this body of work include Shafir (1993), and Ainslie (1992).

<sup>33</sup>Here I am going to ignore the complexities in unpacking dispositions into counterfactuals that travel under the heading of the "conditional fallacy"; the problem is introduced in Shope (1978).

speakers *would* have in one or another sort of counterfactual circumstance. Let's allow ourselves psychological resources from other possible worlds; after all, we can only assemble a usable comparative similarity ordering if we are willing to do so. Again, those psychological resources are your counterfactual judgments about local aspects of similarity. But now observe the vicious circle we are facing. In Lewis's way of thinking, the judgments you would have are those that your *nearest* (i.e., most similar) counterparts have, in suitably specified circumstances. So in order to specify the similarity relations over possible worlds, we need to determine who your nearest counterparts are. But what counts as near and far in the possible world similarity space is a matter of what similarity ordering is chosen. So to determine who your nearest counterparts are, you must first have the similarity relations—which is what we started off asking about in the first place.<sup>34</sup>

Here's an illustration of the problem. A former girlfriend used to dye her hair a very dramatic blonde. Now, she could have dyed it green, and she could also have dyed it blue. I never considered which state of affairs would be more similar to the actual state of affairs, but I'm pretty sure that, had I been prompted, I would have been able to express an opinion. Now, *which* opinion is the one that ought to get factored into a reconstruction of my theory of counterfactual similarity? If the counterpart who thinks that green is more similar is closer to the actual world, then *this* opinion belongs among the raw materials of the Canberra-Plan theory. But if the counterpart who thinks that blue is more similar is closer to the actual world, then *that* opinion belongs among the raw materials. It is quite plausible (though without seeing the details, it's hard to be sure in any particular case) that the differing inputs will make a difference to which counterpart *is* considered closer: after all, if dying her hair blue does count as more similar to the way things are than dying her hair green, then a counterpart who thinks otherwise is *surprisingly* mistaken (and so, *ceteris paribus*, farther away)—and, of course, vice versa.

Occurrent judgments of the relative similarities of alternative possibilities are quite sparse. We need to leave to one side counterfactual judgments whose

---

<sup>34</sup>Here the argument turns on which of your counterparts is "nearest," and not on who your counterparts *are*. But in Lewis's own scheme of things, that's up for grabs, too: who your counterparts are, in a particular possible world, is a matter of who is intrinsically similar to you, and occupies a suitably similar location in that world. In Lewis's picture, it's a mistake to think of your modally extended self as stably composed of your modally thin counterparts, in the way that your temporally extended self is stably composed of momentary time slices. Lewis thought of the, as it were, modality slices of a person in different possible worlds being slices of the *same* person as on a par with the fact that certain stretches of asphalt are all parts of I-5. There's no deep metaphysical fact underlying the "unity over space" of I-5, and if the highway commission were to decide to rename part of I-5 to be the David Lewis Memorial Freeway, they wouldn't be making a metaphysical mistake. I am not pressing this problem because Kripkeans will be willing to treat your reidentification in other possible worlds as a primitive, and we are after points in our treatment of Lewis that will travel; but for a representative statement of the opposing view, see Stalnaker (2012, p. 60).

deployment involves a vicious circularity—meaning, all those would-be judgments of local similarity that are not tightly enough anchored to the occurrent judgments to prevent the sort of problem we just saw from arising. And so when we try to unpack the appeal to similarity, we find that we do not have enough in the way of psychological materials to exhibit the inner workings of the black box and to show that the reduction works: there are too few judgments of counterfactual similarity available, and the black box, it turns out, is almost entirely empty. We are, in principle, not in a position to show that similarity can do the work of those *ceteris paribus* clauses.

## 7.8

Now let's imagine that the materials that have just proved to be out of reach are nonetheless at hand. Again, those materials are similarity judgments about local features of possible worlds, rather than assessments of possible worlds in their entirety. Remember those von Neumann-Morgenstern utility functions, and remember that it is now a commonplace that individual preferences are not regimented into those patterns: human beings rarely, if ever, *have* utility functions. An easy explanation, though likely not the only one, is that imposing decision-theoretic consistency on independently generated preferences is too hard a cognitive task. When you stop and form a preference over two objects of choice, you have to have a special reason to check if it is decision-theoretically consistent with other pairwise preferences you have formed on other occasions. It would be impossible, or next to impossible, to check that the new preference is consistent with all the preferences you have already adopted, or even with most of them. So, again, unless there is a special reason to do so, there will be no reason to expect consistency from pairwise preferences formed on different occasions.

That explanation can serve as a model for the argument at hand. If similarity judgments are generated one by one, to address local concerns, and there is no systematic and concerted effort to render them consistent, it would be an unbelievable coincidence if they *were* consistent. This is certainly true of generic (rather than counterfactual-specific) judgments of similarity. Consider that, during the Cold War, Hungary was like the Soviet Union, but the Soviet Union was not like Hungary; or again, a coffee shop in Nashville formerly displayed a cinnamon bun that looked, its advocates claimed miraculously, like Mother Teresa, but Mother Teresa did not look like the “Nun Bun.”<sup>35</sup> I expect that human beings are rarely if ever in a position to perform a sufficiently

---

<sup>35</sup>I'm grateful to Dedre Gentner for the first of these examples. Medin et al. (1993) collects evidence that independent processes produce uncoordinated similarity judgments; the Hungary example is a variation on an instance reported by Amos Tversky (p. 259). Gentner and Rattermann (1991) documents some of the ways in which judgments of similarity are tied to developmental stages.

ambitious consistency check on their independently formed local similarity judgments; it is simply beyond their (our) cognitive capabilities.<sup>36</sup>

If that is correct, it is not just that we do not have enough in the way of raw materials to reconstruct the global similarity ordering that Lewis's semantics for counterfactuals requires; we also have *too much*. It is evidently overdetermined that Lewis is not in a position to show that a similarity metric can do the work of the *ceteris paribus* clauses that frustrate his reduction of counterfactuals to possible worlds. If he is not, the objection we were considering lapses and, as anticipated, the reduction fails. But there is a more interesting lesson to take home from our guided tour of Lewisian arcana, having to do with the deeper reasons that the local judgments aren't suitable raw materials for the global ones, and I now turn to that.

## 7.9

The argument we have just completed emphasized that our judgments of local, counterfactually relevant similarity are typically mutually independent; it tells us that in our modal cognition we get by with problem- or topic-specific sketches of the modally important features of the circumstances, and that we do not ordinarily assemble these sketches into a global, consistent, and counterfactual-free Big Picture of the modal facts around us. (We cannot come to have that sort of a theory; the Canberra Plan reads an analysis of a given subject matter off just that sort of theory; that was why the Plan turned out to be unusable.) But how do we manage to navigate using these partial and routinely jointly inconsistent sketches? That is too large a question to take on here, but we can consider a runup to it. If this is how we use our local similarity judgments, we should expect to find them prepared, so to speak, for the use they get. Do we?

---

<sup>36</sup>For a complaint about Lewisian similarity orderings of possible worlds that is plausibly a side effect of the cognitive limitations at which I am gesturing, see Pruss (2007).

There is now a standard way of responding to examples like those I have just given: to insist that similarity is context-sensitive and that the context changed, mid-sentence, in both examples. And some philosophers have developed the habit of gesturing at the moment-to-moment variability of the similarity space when faced with one or another problem, as though the problems were thereby solved. But taking the relevant similarity relations over possible worlds to be fluid and context-sensitive makes it *harder* to show that they do their job, not easier. Recall the problem of the previous section, that there aren't enough in the way of available materials to reconstruct enough of a similarity ordering over possible worlds to save Lewis's reduction. If there are many different sets of similarity relations in play, and we switch off between them, moment to moment, then the materials available for reconstructing any one overall similarity ordering are vastly fewer. In other words, the appeal to context, invoked as a way of addressing inconsistencies, makes the sparseness problem *worse*: if before we did not have enough for a single, stable similarity ordering, we will hardly have enough for the many ephemeral context-dependent orderings.

It is also worth reminding ourselves that not all incoherences can be conjured away by appealing to shifts in context—a point emphasized by Lewis (1973, p. 13).

Return to the example of two sections back: would dying her hair blue have been more or less similar to the way things actually were than dying her hair green? Let's suppose that I judge the latter to be more similar: if I do, that opinion is advanced as correct only *ceteris paribus*. After all, if she were dying her hair green because her FSB spymasters had instructed her to poison a wealthy Russian emigre with exotic radioactive materials—well, that would make it much *less* similar. If I am seeing the territory correctly, such implicit *ceteris paribus* clauses are there to allow for the friction between independently generated judgments of counterfactual similarity and constitute, as it were, logical preparation for working with otherwise inconsistent materials. The price of thus softening (or allowing us to paper over) these inconsistencies is that opinions that embed *ceteris paribus* clauses (at any rate, the type of *ceteris paribus* clause capable of serving this logical function) are not well-behaved under conjunction: that's precisely what it takes to make the inconsistencies go away. And this gives us another way of saying why the Canberra Plan won't work here: the first couple of steps in ramsifying a theory are to collect all of the claims we make about some subject area, and then to conjoin them into a single, long sentence. But you should only be willing to assert the theory (the single, long sentence) if you take conjunction to be truth-preserving, and because our local similarity judgments contain implicit *ceteris paribus* clauses, it isn't.

We were examining local similarity judgments because they were meant to underwrite the behavior of our counterfactuals; let's confirm that the phenomenon surfaces there as well. Consider the following counterfactual: if I had a second car, it would be a Hummer. If that's true, it's true only *ceteris paribus*; if I were an avid UFOlogist, my second car would be a black Cadillac, to mislead the government agents who UFOlogists believe are persecuting them, and who drive black Cadillacs, into thinking I was one of them.<sup>37</sup> In Lewis's way of thinking, possible worlds in which I am an avid UFOlogist are farther away (i.e., less similar to the actual world) than the nearest world in which I own a second vehicle. That is to say that the initial underlying judgment of similarity (a way things might be in which my second car is a Hummer—this is now an incomplete possibility rather than an entire possible world—is more similar to the actual way things are than a way things might be in which my second car is a Cadillac) is true only *ceteris paribus*: its conjunction with "I am an avid UFOlogist" comes out untrue.<sup>38</sup>

---

<sup>37</sup>Mitchell (1999, p. 229).

<sup>38</sup>There is an older literature focused on what can now be seen to be a misconception arising out of the assumption that local sketches of the modal territory can be assembled into a global, internally consistent map. Suppose we are evaluating a counterfactual such as, "If I had looked in the mirror, I would have seen my own reflection." On the assumption that we live in a deterministic world, the antecedent of the counterfactual requires one of the following three alternatives: a deeply and pervasively different past, different natural laws, or a miracle—a "jump" whereby I inexplicably come to look in the

Generally, then, the local similarity judgments we are contemplating must be understood as containing *ceteris paribus* clauses, and this allows us to account for a claim about the logic of *ceteris paribus* clauses that I made in passing earlier on. The function of these *ceteris paribus* clauses is to anticipate and accommodate the potential conflicts among independently formed judgments; because one does not normally stop to survey one's other views about counterfactually relevant similarity before forming a particular judgment about it, that judgment might later on have to coexist in your intellectual world with just about *anything*. What comes under the heading of *anything*? Since Nelson Goodman, it has been a truism that there are infinitely many ways that any two things can be similar, and infinitely many ways in which any two things can be dissimilar. So the *ceteris paribus* clauses implicit in counterfactual-supporting local similarity judgments will exhibit a distinctive sort of open-endedness; there can always be further similarities or dissimilarities which properly suspend the judgment. That just means that there is always, as I remarked, *more* built into a *ceteris paribus* clause.

Knowing what underwrites the logical behavior of such *ceteris paribus* clauses allows us to dispose of two objections to our original argument, which I have found often occur to Lewis's followers. First, on what we might call the statistical conception of *ceteris paribus* clauses, such a clause means: the claim holds with high frequency, or with a small number of exceptions. And this might lead someone to expect that the *ceteris paribus* clauses in the proposed reduction of counterfactual claims to possible-worlds claims can be ignored, or that they will wash out, or that they can be exhausted. (If we go on paraphrasing for long enough, eventually we will run out of exceptions, or at least the exceptions will be *few* enough to be negligible.) Second, someone might defend an application of the Canberra Plan in this way: the idea behind the Plan is to take all the theoretical relations at once; as it was once put to me, Lewis's favorite move was to stuff *all* the relations into the box and close the top over them. Since the *ceteris paribus* clauses are just more relations between the theoretical entities, there's no reason, you might think, why they too can't be integrated into a reduction.

---

mirror. And so the problem of how (putting it in Lewisian vocabulary) to assess the relative distances of such worlds from our own came to seem pressing and indeed received a great deal of attention.

The *ceteris paribus* clauses attached to local similarity judgments are—with insignificant exceptions—bound to be triggered by any of these alternatives. That is to say, having rendered such a judgment and having been informed that one of these alternatives is now part of the story, you will retract the judgment of similarity; when it comes to matters such as miracles or systematically different pasts, we become agnostics about similarity and counterfactuals: all bets are off. These puzzles arose in the first place because it was assumed that the local renderings must be glued together into a single “possible world”; the assumption that we are considering the world as a whole is built into the use made of the premise of determinism. The function of such *ceteris paribus* clauses tells us that we should have known better than to try.



However, you can only collect *all* of a *definite* number of sentences; conjunction is an operation well-defined over finite sets of sentences (and which can, with a little bit of ingenuity, be well-defined over countable sets of sentences), but not over *indefinitely* many sentences. We have just seen why it is characteristic of *ceteris paribus* or “other things equal” clauses that there are always *other* “other things” to be “equal”; we can no more count or survey the ways in which there are always further things to go wrong than we can count or survey the ways in which objects or states of affairs can resemble or fail to resemble each other. That means that we can’t collect all the relations at once; there’s no theory, containing explicit renderings of all of those “other things,” to ramsify (i.e., once again, we have explained, from a slightly different angle, why the Canberra Plan is a nonstarter here).

We have now underwritten our earlier willingness to treat *ceteris paribus* clauses as inexhaustible. However, on the Lewisian understanding of *ceteris paribus* clauses, there aren’t really indefinitely many things to go wrong; *ceteris paribus* clauses don’t essentially contain an ellipsis, or if they do, unpacking those ellipses comes to an end. That understanding is an error about the logic of *ceteris paribus* clauses generally, and we now have a subject-specific explanation for how it is that the *ceteris paribus* clauses which appear in the course of paraphrasing counterfactuals possess the distinctive logical open-endedness on which our argument turned. In fact, as we have just seen, even Lewis’s own apparatus, taken together with platitudes about the available dimensions of similarity, commits him to this logical feature of *ceteris paribus* clauses.

## 7.10

We still need a reason for enforcing the reductionist demand, one that goes deeper than: Lewis happened to have accepted it.

A claim is not contentful merely because it is phrased in familiar words, and a philosopher’s duty is to give his claims content. Reductionists owned up to their duty by promising to translate away the vocabulary they proposed to discard. If a philosopher wants to let go of the reductionist way of giving his claims content, then he must supply a substitute; acting as though his words meant something, when he has not done any of the work required to make them do so, is not an option.<sup>39</sup>

When previous reductionist programs failed, their adherents retreated to claiming that one kind of thing (that they had failed to reduce) *supervened* on the other (the kind of thing they’d failed to reduce it to). Supervenience has been the traditional substitute for reduction, and for a long time, until it was supplemented by the Canberra Plan, it was the only substitute in general use.

---

<sup>39</sup>These remarks are meant to address confusions that appear to be quite widespread. A typical instance is Divers (2002, pp. 28f).

When such a move gets made, the picture stays the same: biological objects and facts, say, are really just configurations of physical objects and facts—only without the obligation to say *which* configurations. The move is often accompanied by the announcement that the claim being made is about ontology or metaphysics (more recently, about what the “truthmakers” are) and is not intended as conceptual analysis. Put in less hifalutin language, although you are giving up on saying what your words had meant, you are still trying to say what they were—collectively—really *about*.

But one way or another, a philosopher offering a claim about what is really what else has to answer the question, “Where’s the beef?” The beef which a supervenience theory supplies is the claim that if, say, the biological supervenes on the physical, then the biological facts couldn’t be different without the physical facts being different, too.<sup>40</sup> Lewis himself was a fan of supervenience in other domains, and so let’s consider whether modal supervenience was available to him as a fallback position. In Lewis’s picture, the modal facts are really just facts about the configurations of possible worlds. Even if he couldn’t say *which* configurations, he ought to have insisted that the *coulds* and the *woulds* and the *musts* supervene on the facts about the possible worlds: if the modal facts are just a matter of how the possible worlds are configured, you can’t change the modal facts without changing the possible-world configuration.

Or was that really an option? Again, the cash value of supervenience is that, if the supervening facts were different, the supervened-on facts would have to be different. But that’s a modal claim, and Lewis acknowledged that it was: “[W]e have supervenience when there could be no difference of one sort without differences of another sort. . . . Clearly, this ‘could’ indicates modality. Without the modality we have nothing of interest.”<sup>41</sup> Unless the claim can be accounted for using the apparatus of possible worlds, the position is self-refuting. Can it?

What the set of all possible worlds is (and how it’s configured in “similarity space”) was not, on Lewis’s way of thinking, a contingent matter: it *could* not be different than it is. So Lewis himself was not in a position to so much as articulate the fallback supervenience claim. But we can ask whether, regardless of Lewis’s own view of the matter, the position is theoretically viable. Making out counterfactuals about the set of all possible worlds would involve modal realism about other possible super-worlds, each of which is a way the set of all possible worlds might have been.<sup>42</sup> Since we are imagining worlds over

<sup>40</sup> See Lewis (1999, pp. 33–39) for complications.

<sup>41</sup> Lewis (1986, pp. 14f).

<sup>42</sup> As per Skyrms (1976, p. 327 n. 10); the option is one that Lewis explicitly considered and rejected: “it makes no sense to repeat the very method you think has failed, only on a grander scale. . . . There is but one totality of worlds; it is not a world; it could not have been different” (Lewis, 1986, p. 80). “It is futile,” he subsequently wrote, “to want the entire system of worlds to satisfy a condition, because it is not contingent what conditions the entire system of worlds does or doesn’t satisfy” (Lewis, 1986, p. 125).

and above the possible ones, this extension of Lewis's approach amounts to the currently popular idea of supplementing possible worlds with "impossible worlds"—or rather, with universes of them.<sup>43</sup>

To think intelligently about the counterfactual covariance of modal facts with configurations of possible and impossible worlds, we would need to make sense of counterfactuals such as "If the configuration of possible worlds had been different, then . . ." For this, on the approach we are trying to extend, we would need a similarity ordering over the impossible worlds and over universes of impossible (together perhaps with possible) worlds. But we have just argued that we are not even in a position to work up a usable similarity ordering of *possible* worlds. *A fortiori*, we are not going to have the wherewithal to construct the far more demanding similarity ordering. How much does any sane person have in the way of thoughtful opinions about which features of *impossible* worlds make them more or less similar to each other? And how much does anyone have in the way of thoughtful opinions about what features of alternate universes of worlds make them more or less similar to the Lewisian universe of possible worlds? Once again, thoughtful opinions are the product of attention, deliberation, and, generally, cognitive work. No one in his right mind has paid any attention at all to such matters; therefore, no one in his right mind has the thoughtful opinions that would underwrite the sort of counterfactuals required to make sense of the supervenience of the modal on possible worlds.

The point is that we can't afford to be casual about the failure of a strict reduction. It is not as though Lewis could have retraced the steps taken by earlier embattled reductionists, and backed off to a weaker but still contentful modal supervenience account. When you back off from insisting on a reduction, you have to replace it with something else, if you're going to end up saying anything at all. When early analytic philosophers did their metaphysics (all the while denying that they were), they spelled out the contents of their claims linguistically, as theses about what could and couldn't be given eliminative paraphrases. More recent philosophers have spelled out the contents of their more modest claims via the Canberra Plan and via supervenience. But

---

Divers (1999) has attempted to extend and defend Lewis by offering a "redundancy interpretation" of modal claims about the machinery: statements like "It is possible that there are many worlds" are allowed, but they are flattened down to "There are many worlds." So notice that the way Divers provides of stating supervenience claims about modality disbars them from doing any of the work that we needed supervenience for: we can no longer capture the thought that, were the modal facts to vary, the configuration of the totality of worlds would have varied as well; compare, on this point, Divers (2002, pp. 55–57). However, Divers (2002, pp. 208f) explains how some types of reductionism about possible worlds—that is, views on which possible worlds are themselves to be reduced—can accommodate supervenience claims naturally.

<sup>43</sup>Yagisawa (1988), which gives a very funny two-front argument against Lewis exploring the option of "impossible worlds"; Lewis (1986, p. 7n) did not regard the suggestion as a friendly amendment to his view.

these devices are not available, which was why, when Lewis discussed modality, he wrote like an old-style reductionist. If you back off from old-style reductionism, but don't say what you're backing off *to*, then you haven't managed to make contentful claims, and if the Canberra Plan and supervenience are unavailable, it's reductions or nothing.

Let's go back to an option we were considering a while ago, that a Lewisian similarity ordering over possible worlds might be constructed out of the objectively (that is, physically) present properties of those worlds, and without appealing to our own psychologies, but rather to universals. Can that claim be softened out to: the similarities and dissimilarities we discern among possible worlds *supervene on* (even if they can't be reduced to) their objective, universal-based features? A supervenience claim has it that there's no difference between (say) nonphysically characterized states of affairs unless there's also a physical difference. The point of the move to supervenience is to prescind from telling you *what* the physical difference in such a case is. And so it follows that a supervenience theorist can't back up a claim to the effect that a big difference of some nonphysical kind is a *physically* big difference. And so it follows that a supervenience claim cannot do the work needed to save Lewis's account from our initial objection to it.<sup>44</sup>

---

<sup>44</sup> At this juncture, you might be wondering whether I'm being sufficiently charitable. After all, it is quite normal to assert counterfactual conditionals about matters that could not be otherwise, as when, working a mathematics exercise, I say to myself: suppose that the height of the formula *were* greater than the theorem allows, then this variable *would* take such and such a value . . .

Would Lewis have gone along with this suggestion? We are told that "nothing can depend counterfactually on non-contingent matters. For instance nothing can depend counterfactually on what mathematical objects there are, or on what possibilities there are. Nothing sensible can be said about how our opinions would be different if there were no number seventeen, or if there were no possibility for dragons and unicorns to coexist in a single world. All counterfactuals with impossible antecedents may indeed be vacuously true. But even so, it is seldom sensible to affirm them" (Lewis, 1986, p. 111).

Let's imagine that Lewis would in the end have to have allowed for some way of construing counterfactuals of this sort. Why can't it (whatever it is) be used to underwrite the counterfactuals about the set of all possible worlds that would allow us to make sense of modal supervenience?

Some way of understanding such counterfactuals there must be. Whatever it is, however, it is not to be made out using the apparatus of possible worlds. So it must be done some Other Way, and that Other Way will ultimately need to be given a philosophical explanation. The Other Way is going to have to be very powerful indeed, if it is going to handle counterfactuals about how the possibilities themselves might have been different. There is no surface difference between those counterfactuals that require treatment via the Other Way and those that are amenable to Lewis-style possible-worlds renditions; laymen don't distinguish between the counterfactuals of mathematical reasoning, and counterfactuals about their garage work. So why shouldn't we expect that, when we have the account of Other Way counterfactuals, it will handle the phenomena that Lewis's modal reductionism was supposed to handle? In short, our final worry that we are being uncharitable turns out to presuppose a further and distinct account of (anyway certain kinds of) modality, one which we can reasonably expect to make Lewis's own account superfluous.

Divers (2002, p. 98) offers a companions-in-guilt response to the complaint that Lewis and his followers cannot handle such counterfactuals: nobody else can explain them either. So note that this reply is irrelevant to the point I am making here. Note also that the inability to handle such counterfactuals

## 7.11

Lewis gestured at a method of converting ordinary modal language (woulds and would haves, coulds, musts, and so on) into a possible-worlds paraphrase, and in this he was typical. He was also attempting to enforce what he thought were our ontological commitments to the possible worlds, and in this he was atypical. There is only so much mileage to be gotten from disabusing us of something that no one else believed anyway, so let's focus on the common ground: pretty much everyone in the business takes it for granted that ordinary modal speech, including counterfactuals, bald claims about possibility and necessity, and so on can be converted into the possible-worlds vocabulary.<sup>45</sup> Even those who insist that possible worlds have modality built into them think that the paraphrase is available. Even those who are philosophically unhappy with the possible worlds respond by attempting to paraphrase possible-worlds talk away; the assumption implicit in this way of proceeding is that the vocabulary of possible worlds gives us the right expressive power (it captures the content of ordinary modal speech); we just want a different way of getting precisely *that* expressive power. What we have seen, however, is that possible-world renderings don't match the commitments of ordinary statements: as when, in our example, you took on *extra* commitments, to what something which looked a lot like Mithridates, but was really a bug-eyed monster, would do.<sup>46</sup>

---

is an objection to Divers's proprietary treatment of apparently modal claims about the modal machinery (that is, of what he calls "extraordinary" modal claims, mentioned in footnote 42 above); since counterfactuals are of a piece with the rest of the machinery, that treatment should extend gracefully to cover "extraordinary" counterfactuals as well, and it does not.

<sup>45</sup>Compare Chihara (1998, pp. 144ff); typically authors who allow that the expressive power of a language with modal primitives and the possible worlds vocabulary can differ think that what you can say in the former, you can say in the latter. For example, Melia (1992) argues that the possible-worlds vocabulary allows you to say *more*. And when Lewis (1986, pp. 10–13) faced up to difficulties in rendering ordinary modal expressions into a regimented modal vocabulary, he resolved the problem thusly:

If this language of boxes and diamonds proves to be a clumsy instrument for talking about matters of essence and potentiality, let it go hang. Use the resources of modal realism *directly* to say what it would mean for Humphrey to be essentially human, or to exist contingently.

In other words, the possible-worlds vocabulary is perhaps more powerful than the language of quantified modal logic, and powerful enough to render ordinary claims about essences, etc.

However, for dissent about the counterpart-inflected vocabulary, see Fara and Williamson (2005).

<sup>46</sup>One might wonder whether having to surrender the possible-worlds renderings of counterfactual discourse nonetheless allows us to keep possible-worlds renderings of thin modal vocabulary. Perhaps "Mithridates might have worn a chocolate crown" can still be construed, without change of content, as "In some possible worlds, Mithridates wore a chocolate crown." But we can now see that the ordinary objects that figure in such sentences, like Mithridates, are modally thick: to be *Mithridates* is for indefinitely many counterfactuals to be true of one. If possible worlds are modally extensionless, then not even thin modal claims about ordinary objects can be paraphrased into the possible-worlds vocabulary without change of content. Thus, when Stalnaker (2012, p. 11) offers the relatively modest-sounding stance that "it is not reduction but regimentation that the possible-worlds framework provides—a procedure for representing modal discourse, using primitive modal notions, in a way that helps reveal its structure," the apparent modesty is insufficient to sidestep the problems we

No one has a very good philosophical account of modality, and in my view, the habit of talking about possible worlds is in significant part to blame. When the possible-worlds way of paraphrasing modal vocabulary hit the diaspora of analytic philosophy departments, it was adopted as a distinctive if odd manner of speech—a kind of scattered regional dialect. In the South, they say “you all” (or, in some of the more rural areas, “you’uns”); in the philosophy departments, they came to say things like “in some possible world, you are a contender.” The accent is easy to acquire, and because it was assumed that what you could say in one way, you could say in the other, these philosophers acted as though there really was not much more to it than an accent. But the philosophers (and philosophers-in-training) who did so for the most part had the impression that they thereby *understood* modality—maybe not everything about it, but enough for their purposes. Maybe one didn’t know what possible worlds *were*, and maybe there were other issues to argue about, such as whether other possible worlds really *exist*, or whether you could reduce the modality away, but one at any rate had a clearer way of talking through one’s modal claims. And so we find philosopher after philosopher observing how hard it would be to do your thinking about modal issues without possible worlds.

There were two pernicious consequences. The first was that a great many arguments came to assume the equivalence of claims in the two vocabularies; those arguments must now be reassessed, because possible worlds *don’t* provide a clean way of keeping track of our ordinary modal claims: as we’ve just argued, possible-worlds renderings of ordinary counterfactuals don’t preserve the contents of those counterfactuals. The second was that philosophers stopped thinking hard enough about modality, because they took it that it was something they *already* (basically) understood.

It should be obvious that to acquire a funny accent is not to understand anything you didn’t understand before. But we now have an argument to add to that truism. Possible worlds are neither a transparent alternative representation for our ordinary modal assertions, nor an aid to understanding. And that’s a problem not just for Lewis, but for everyone who talks that way, which means most analytic philosophers. The peculiar dialect *doesn’t* capture, and is not a passable surrogate for, the content of ordinary modal discourse. It’s a mistake to think that manipulating the pictures associated with the dialect is going to help you understand the mysteries of modality.<sup>47</sup>

---

have identified; unless we can capture the content of ordinary modal discourse, there is no framework, no procedure, and no revelation.

<sup>47</sup>I’m grateful to Irene Appelbaum, Aisling Crean, Kate Elgin, Don Garrett, Kathrin Koslicki, Beatrice Longuenesse, Ceil Maslen, Fraser McBride, Shaun Nichols, Doug Patterson, Laurie Paul, Guy Rohrbaugh, Candace Vogler, and Matt Weiner for helpful conversation; to Chrisoula Andreou, Sarah Buss, Ben Crowe, Christoph Fehige, Alan Hajek, Clif McIntosh, Laura Schroeter, Scott Shalkowski, and Mariam Thalos for comments on earlier drafts; and to audiences at Brigham Young University, Kings College London, the University of Reading, the University of Colorado, the University of Notre Dame, and the University of the Saarland for valuable feedback. Thanks also to the University of Utah’s College of Humanities for sabbatical support.

## Progressive Necessity

For most of the twentieth century, Anglo-American philosophy predominantly approached its problems as exercises in semantic analysis, and modality was no exception. Making metaphysical sense of necessity, possibility, and the *weres* and *would'ves* of counterfactual conditionals was to be accomplished by determining what sentences containing them *mean*, and to do so the model theory of modal logic was drafted into the service of possible-worlds accounts of modality.

The program was genuinely promising at its inception, meriting and obtaining a far-reaching commitment of intellectual resources. But, about a half-century into the possible-worlds approach to modality, the semantic approach to the philosophical understanding of modality has not panned out. It's time to try something different.<sup>1</sup>

In my view, we would do better to concentrate instead on figuring out the *function* of modal cognition, and my objective here is to motivate the investigation and indicate the direction I think it should take.<sup>2</sup> This is going to be a lengthier-than-usual discussion, and so here, quickly, are the bases I want

---

<sup>1</sup>Very briefly, the semantic approach is locked into a two-stage program: first, provide a paraphrase of ordinary modal content in the dialect of possible worlds; next, address whatever philosophical questions you have about those possible worlds and their inhabitants. Such a program can only succeed if the paraphrase captures that ordinary modal content; it seems to be taken for granted, across the spectrum of philosophical opinion, that it does so. Unfortunately, an argument that has made previous appearances in the history of analytic philosophy—it was used to dispose of phenomenalism and logical behaviorism—shows that the possible-worlds paraphrase of ordinary modal discourse cannot work. For that adaptation of the older argument, see Chapter 7 in the present book; also see Millgram (2009a, Chapter 11).

<sup>2</sup>The question has also been raised by Nichols (2006) and because we see things to a certain extent the same way, here and below I will highlight differences of emphasis and opinion. A preliminary contrast: Nichols takes up the function of modal cognition as a subtopic and announces his primary concern to be “what is involved, psychologically, in making modal judgments” (Nichols, 2006, p. 237). I mean to put this latter question to one side, as an implementation issue.

to touch, the order in which we'll get to them, and what I want to be doing as we traverse them. First, I will introduce a puzzle about modality to which I think philosophical treatments of the topic must be responsible. Then I will narrow the scope of discussion somewhat; when it comes to making sense of the complex landscape of modal phenomena, a salami-tactics approach seems most likely to succeed. I will be taking up necessity now and leaving counterfactuals, possibility, dispositions, and so on for later. I will float a trial balloon meant both to illustrate and advance the new research program: that necessity is a component of an attention management technique. Although I mean to persuade you to take it seriously, my main interest in exploring it is to get a feel for the investigation of modal metaphysics, on the assumption that it consists in the examination of a family of intellectual devices.

I am going to fold in a further hypothesis, that much of the necessity we work with has been produced or adjusted to manage problems posed by the division of labor. That suggests that the preview I am after should be focused on a particular specialization. But which? On the one hand, our oldest visibly successful sciences of the necessary are mathematical, and we can reasonably hope that a discussion of mathematical necessity will serve as a guide to the forms that necessity takes in less ancient disciplines. On the other, however, mathematics has long been an autonomous enterprise; such enterprises take on lives of their own, and after enough time has passed, no brief discussion can do justice to them. So instead I am going to tell a just-so story about the emergence of a protomathematical practice and the highly simplified *ur*-form of necessity that appears along with it. Think of the move as analogous to a state of nature argument in political philosophy; these arguments illuminate the legitimacy of the political state, even though no political state could be much like the toy states they describe.<sup>3</sup> In this case, the story will be meant to show how division of labor can figure into the way necessities are assigned.

As I talk my way through the protomathematical just-so story, I will make a series of suggestions. First, necessity will normally be differently inflected from one specialization to another. Second, we should expect necessity to be

---

<sup>3</sup>Here's what I mean by that. No state that is a going concern has only the single function that some state of nature argument or other highlights. Real states not only provide police protection and a defense department, but in addition run schools and museums and memorials and post offices and parks; they put on parades and fireworks displays; they regulate financial institutions and food production and medical services; and that's just the beginning of a very long list. Any actual state is a very complicated institution with vastly many functions, and because a state of nature argument makes visible only a single aspect of such an institution, the "states" that appear in such arguments are much thinner than any actual state. Nonetheless, state of nature arguments can make some of the various points of having political states quite compelling.

Some philosophers (and political activists) take the *only* legitimate activities of the political state to be those exhibited by some state of nature argument; as the above remarks will have indicated, I do not. Thus, and analogously, I do not mean to suggest that necessity, as it figures into the mature enterprise of mathematics, is exhausted by the lessons of the just-so story I will present.



*progressive*: on average, over time, more and more facts will become necessary. Since it doesn't *look* that way—on the contrary, “becoming necessary” registers as an oxymoron—next up will come a survey of differences between how necessity works and how it presents. I will try to show why an intellectual device that performs the function that necessity does is almost bound to give rise to a series of metaphysical illusions; at this point I will be in a position to frame more familiar intuition-driven philosophical treatments of necessity from the perspective of the cognitive-function program.

After taking up the question of what to make of modal inference, and thus of work in modal logic, I will explain how, on the account under development, necessity, apriority, and analyticity should move roughly in tandem. And I will illustrate their connections by diagnosing the error in Mark Johnston's so-called Missing Explanation Argument—that is, his attempt to resuscitate and get to the bottom of the argument in Plato's *Euthyphro*. Finally, looking back over the discussion, I'll assess the prospects of cognitive function analyses in this subject area.

## 8.1

Human beings commit a great deal in the way of mental effort to modal cognition. People work hard at figuring out what *would* have happened, had such and such taken place. They consider what *might* (yet) happen if they do this or that. They make announcements about what *couldn't* happen, and about what *has* to happen. And their conclusions both control decisions about what to do and views of how things stand. A little less obviously, but as importantly when it comes to appreciating just how pervasive modal cognition is, almost any empirical concept used in ordinary life has counterfactually articulated conditions of application. Dispositions are a special but exemplary case: blown glass is fragile (and perception follows conceptual structure: you *see* it that way); its fragility is constitutive (if it's not fragile, it's not blown glass); the fragility consists almost entirely in the truth of counterfactuals such as: if it were whanged, it would shatter.<sup>4</sup> Or again, John Stuart Mill analyzed physical objects as “permanent possibilities of sensation,” and this much is right in his view: if a great many counterfactuals about sensations were not true—for example, if you pushed it, you would feel it pushing back; if you looked at it from a different angle, you'd see a suitably perspective-adjusted image of it—you would normally retract your claim that what you had before you was a physical object.<sup>5</sup>

---

<sup>4</sup>“Almost entirely,” for the reasons noticed by Shope (1978) and discussed under the headings “conditional fallacy” and “finkish dispositions.”

<sup>5</sup>Mill (1967–1989, vol. IX, Chapter 11).

The deep philosophical puzzle posed by our attention to modality is nicely captured in a business concept due to David Finkelstein.<sup>6</sup> The Counterfactual Psychic Hotline is to work just like other psychic hotlines, only instead of telling you what *is* going to happen in the future, the Counterfactual Psychic answers your questions about what *would* have happened, if, say, you had told your boss what you really thought, or accepted the *other* proposal, or chosen the road less traveled by, instead of the freeway entrance you *did* choose. Belaboring the point a bit, here's why it's funny. On the one hand, people would pay \$4.98 per minute (or whatever it is such phone services charge these days); it matters enormously to people what would have happened; it makes all the difference in the world. On the other hand, you could get away with it, because you never see what merely might have happened, only what does happen. And so it doesn't matter what the Counterfactual Psychic says; it makes no difference at all. The puzzle, then, is how both these observations can be true together. How can something that makes no difference at all make all the difference in the world?

There are reasons, reasons we don't want to ignore, for taking the different types of modal cognition—thought about possibilities, about necessities, about counterfactual conditionals, about causation, about dispositions, and the like—to be intimately related. However, we shouldn't assume that all these different aspects of modal thinking serve the same function.<sup>7</sup> For that reason, the right strategy is to take the territory piecemeal: to start with what look to be the more straightforward elements, and later on proceed to the more challenging forms of modal thought. I propose to begin with necessity; if you're a professional philosopher, I'm after the sort of thought that your colleagues are wont to express with the box operator and to attempt to paraphrase with locutions like “true in all possible worlds”; if you're a nonprofessional (in some ways, when it comes to this topic, a positive intellectual advantage!), the sort of thought that gets expressed by utterances like “It *has* to be that way.”<sup>8</sup>

---

<sup>6</sup>This is the Wittgensteinian philosopher of mind, and not the philosopher of quantum mechanics of the same name.

<sup>7</sup>Or functions: we shouldn't assume that one of these topical divisions of modal thought serves only one function. It's normal, in creatures that are “designed” the way humans and other animals are, for a trait that takes root because it does one job to be subsequently extended and adapted to do more jobs; that's true all around, and it's true of cognition in particular. As Marvin Minsky once put it, the mind is a kludge.

<sup>8</sup>Recall from footnote 1 that these paraphrases are unsuccessful; I'm just gesturing at what they're after. Now, it's a commonplace that necessity and possibility are interdefinable:  $\Box p \equiv \sim \Diamond \sim p$ . However, because we're thinking about cognitive function, we may have to treat them separately. Compare: fractions and decimal notation are interdefinable, but they're *used* differently. Or again, roman and arabic numerals are interdefinable, but they're taught at different stages of a curriculum. One more guess: the notion of possibility is actually that of *possibilities*, and presupposes something on the order of a state space or a choice set. This in turn will involve us in complexities that I think we can avoid—or anyway postpone—in making sense of straight necessity. Briefly, when I suggest starting off with necessity, I mean necessity as opposed to, among other things, possibility.

The concern which the Counterfactual Psychic Hotline makes vivid for modality in general arises for necessity in particular. Within the tradition of French positivism and its subsequent appropriation by John Stuart Mill, it was argued that necessity markers could be deleted without loss: facts are observed, but the necessity of a fact is never observed. Accordingly, the part of a scientific theory that specifies that, say, causal laws are *necessary* is not responsible to any observable evidence. Instead of announcing of a black hole that an object inside its event horizon *cannot* escape, you could just as well say that it *does* not escape. Likewise for logical necessity: instead of saying that when the premises of a properly structured argument are true, the conclusion *has* to be true, you could just as well say merely that it *is* true. You never see the necessity, and so it makes no difference at all.<sup>9</sup>

Still, it's hard to believe that necessity doesn't somehow make a great deal of difference. Necessity crops up frequently enough in ordinary discourse for command of markers like "have to," "must," and the more idiomatic "got to" to count as required for even initial mastery of a language. If necessity didn't have a real use, one that made it pretty much indispensable, you would expect much more variation in the extent of its deployment than you see; in some places, it would be invoked frequently, in other places, only on occasion; in

---

Nichols (2006) *does* start with the cognitive function of possibility, which he suggests operates as a marker for risks and opportunities; judgments of possibility, he proposes, are generated in the first place by running imaginative simulations, and *impossibility* marks "imaginative blocks," that is, the experience of one's inability to imagine something.

One reason that I am not adopting Nichols's point of entry into the investigation of the function of modal cognition is that I'm unconvinced that imagination can be invoked as an independent explainer. When you ask someone to imagine something, you are giving them a task, and what matters for a philosophical understanding of imagination is what counts as succeeding or failing at it. Most of us perform that task very badly; that's why people who *are* good at it can make a living as novelists, playwrights, and so on. But you can expect people engaged in it to help themselves to whatever expedients come to hand.

Modal judgments are likely *one* of those expedients, and *sometimes* a sufficient expedient. Nichols allows that imagination may be constrained by modal judgments: the child doesn't imagine that the shark will bite him because he's been told that the shark *can't* get up on the beach. However, that seems to understate the extent to which imagination is shaped by modal judgment; sometimes announcing the relevant modal judgment counts as having complied with the demand "imagine such and such." In such cases, exercise of the imagination is a mode of reporting modal judgments and not a basis for them.

I imagine that even a preliminary survey of the territory will rule out identifying the faculty of imagination with an on-board simulator. In footnote 10, I am going to ask you to imagine how selection pressures would work on modal cognition, were it not to serve any vital purpose; in Section 8.4, I'm going to take a shot at imagining a species of primate that happens never to have evolved; starting in Section 8.2, I'll imagine the emergence of a simple mathematical practice. And as you follow along, you can judge for yourself how likely it is that performing these imaginative exercises consists simply in turning on a simulator module (the one boxologists allege comes as standard equipment).

<sup>9</sup>For a more leisurely rendition, see Chapter 5 in this volume, and Millgram (2009b). While it is possible to retain the term "necessary" for marking consequences of a law of nature (and Mill did help himself to a variant of this usage), one could equally suffice with observing that those consequences are true.

some places, not at all. In the coming argument, I will maintain that there is a great deal of variation in the contouring and operation of necessity, both over time and from subject area to subject area; but necessity in its various forms is found pretty much across the board.<sup>10</sup>

I'm going to run with the assumption that necessity does real cognitive work. My guess is that if we can figure out the roles it plays in our cognitive economy, we'll end up knowing everything about necessity that matters for philosophical purposes. My guess is that if we do that, the semantics will take care of itself. Maybe it will become obvious what statements with this sort of modal content mean, but, more likely, it will turn out that it doesn't matter all that much, and that we'll be able to make philosophical headway on the problems that concern us without producing that kind of semantic analysis along the way. First and foremost among these, a satisfactory philosophical treatment of necessity should resolve the apparent paradox of something that makes no difference at all nonetheless making a great deal of difference.

---

<sup>10</sup>That very quick argument has a more general version, for the conclusion that the intellectual energy invested in modal thought isn't all wasted motion. Here's an evolutionary just-so story. Suppose that modal cognition really didn't serve a vital function (or several of them). Sooner or later, someone would come along who spent less time and energy on it than everyone else, and so he would have more time and energy for the other cognitive tasks that really were useful. He would do better, on average, at those tasks and would tend to out-reproduce the competition. Among his more-numerous-than-average descendants, who would, if the trait was heritable, be more like him in spending less time and energy on modal matters than the initial population, some would come along who spent still less, and they in turn would be more adaptively fit than their predecessors . . . and, in the end, the disposition and ability to engage in modal cognition would be ruthlessly selected out of the population. It hasn't been: so there must be *something*—something important—that thought with modal content does for us.

We don't want to treat this last train of thought as any more open-and-shut than it is. People spend a lot of time and energy on music, and appreciation for music is obviously a spandrel, that is, a side effect of some feature that *is* subject to adaptive pressures. (Perhaps the sensitivity to rhythm, for instance, is what comes of having an auditory pattern recognition module attuned to the regular footfalls of running predators.) Still, if you stripped away the music from creatures like us, what you would get would be: creatures just like us, only minus the music; they wouldn't be missing anything further that mattered.

Are there adaptive pressures that explain why musical dispositions haven't been selected out of the population, or should we think of music as a lucky accident? It's hard to say. Possibly the cultural capital invested in our musical abilities makes having the ability to exploit it an advantage. Al Gore, when he was a candidate for the presidency, was said to dance al-gore-rhythmically, disparagement which no doubt hurt his chances with the young and hip; even if music appreciation is, so to speak, noise in the system, if you were sufficiently unmusical, it would cost you.

However, you can turn off the music when you need to concentrate; whereas you can't turn off the modal dimension of thought and experience. And it's hard to imagine stripping away the modal cognition from human beings and being left with anything that didn't require full-time special-ed staffing. Anticipating a bit, modal cognition is inferentially central. Recall that the just-so story with which I began this note was itself almost entirely composed of counterfactual conditionals. If you don't *get* arguments like that one, you're not going to be able to understand an awful lot about what's going on around you.

## 8.2

Philosophers attempting to introduce “necessary” into a beginners’ conversation often paraphrase it with ordinary locutions such as “*got to*” or “*has to*,” and when they do so, it’s genuinely surprising how typically they attempt to convey the force of necessity with a peculiar verbal emphasis, one that more or less sounds as though they’re trying to lay an egg.<sup>11</sup> Falling back on intonation in this way is a telltale of philosophical confusion. We can do better by asking what necessity is for.

Let’s imagine a mathematical state of nature, one which adapts John Stuart Mill’s views about the likely history of arithmetic.<sup>12</sup> In the practice of *protoarithmetic*—since we’re going to have distinguish its developmental phases, let’s say *Stage I protoarithmetic*—the members of some faraway tribe that time forgot have learned to count. However, they don’t compute sums the way we do: if a shepherd finds that he has three sheep, by counting them off, and then another five sheep, also by counting them off, to arrive at the sum he counts all of them off, getting eight. It goes without saying that more advanced notions, such as multiplication, are beyond the tribespeople. By the lights of such a shepherd, addition by counting sheep is the experimental investigation of an empirical subject matter.<sup>13</sup>

Members of the tribe already know, from extensive experience, that when you sum three and five this way, very large results are unrealistic; it won’t turn out that this shepherd has *thousands* of sheep. But as far as they are concerned, it’s something of an open question exactly how many there *will* turn out to be: maybe next time the sheep are counted off, there will be nine, or six, or even eleven or twelve, and perhaps the more adventurous among them, as they wait with bated breath for the outcome, place bets. There are

---

<sup>11</sup>Compare Hume’s approach to necessity: he found himself worried about what the feeling—the one you express with the egg-laying grunt—is copied from, rather as though it were a very special color, one whose lack of location on the spectrum might somehow account for its cognitive importance (Hume, 1888/1978, pp. 77f, 88, 155–165; for a summary, see Millgram, 2005a, pp. 222–223).

<sup>12</sup>Mill (1967–1989, VII:224–261); the train of thought here is modeled on that of Craig (1990; see Chapter 6, note 51, in the present volume for an overview). State of nature arguments are closely related to “creature construction” arguments, as introduced in Grice (1975) and revived by Bratman (2007, pp. 49–50).

<sup>13</sup>How *high* can they count? Our own system of numerals presupposes control of both addition and multiplication; in normal applications the usual alternative invoked in philosophy classrooms, on which a numeral is explicitly constructed by appending a string of successor symbols to an initial symbol (for zero or one), is obviously not practicable. For the purposes of our just-so story, at Stage I, we can suppose the members of the tribe to have memorized a series of symbols that allows them to count as far out as they need. Perhaps they can agree on further symbols when there is a felt demand, and sometimes there are expedients that serve when the memorized symbols run out; see footnote 14, below. Evidently, Stage I protoarithmetic is small-finite and weakly finite in the sense considered in Dummett (1978, pp. 258f).

live alternatives, and what matters for present purposes is that the existence of alternatives imposes costs.

For one thing, the discovery procedure is itself time-consuming. It's not too bad when you're only adding three and five this way, but if a neighboring shepherd has a hundred sheep and buys three more, it will be tedious to figure out how many he now has. (And may be onerous if he falls asleep while counting sheep and has to start over.<sup>14</sup>) Moreover, until the counting procedure has been executed, he will have to hedge his bets; turning up at the feed store, our first shepherd orders enough feed for—we can imagine him pausing to consider what would be safe—up to twelve sheep, just in case the count comes out high. Extra feed costs money and storage space, and when his neighbor purchases feed for up to a hundred fifty, the specifically financial costs to him are substantial.

So imagine further that the tribe produces a homegrown Prometheus, the inventor of an incantation for closing off those alternatives: it makes them simply unthinkable.<sup>15</sup> The incantation clearly has to be applied with caution. But once it is confirmed through repeated experimentation that three and five come out eight (anyway, frequently enough so that even if there are occasional exceptions, it still pays to close off the alternatives), the members of the tribe can be told (and here's the incantation) that the sum doesn't just *happen* to be eight: it *couldn't* have been any different, it *has* to (has *got* to) be eight. The spoken incantation comes paired with a magical inscription, written "□," as in " $\square(3 + 5 = 8)$ "; and once the spell is cast, that  $3 + 5 = 9$ , that  $3 + 5 = 10$ , and so on are no longer alternatives at all. The correct sum can now be learned by rote—an enormous time saver on its own—and all those shepherds can stop ordering extra feed. (Call this *Stage II* protoarithmetic.)

Informing you that something is necessary is letting you know that, as one might put it colloquially, anything else is just not going to happen. For practical purposes, it's telling you that you can *ignore* the alternatives.<sup>16</sup> Now, attention is a scarce resource. If you are a parent, you will be very aware how very energetically children compete for it, and nobody works that hard for anything that isn't in short supply. Google won the search engine wars not merely because it allows its users to find one web page or another, but because

---

<sup>14</sup>However, sometimes there are workarounds: in Hmura (1996, pp. 217f), a rancher impugns the ability of a forest ranger to count high enough to have an accurate number for his herd. The ranger's ironic response: "I admitted that I could count only to 100 [the number of cattle permitted in that area], but if the count went beyond that, I could start another hundred."

<sup>15</sup>Of course there are no magic incantations, and in real life the effect is produced by such training methods as requiring schoolchildren repeatedly to chant the correct sums.

<sup>16</sup>By way of preempting a possible confusion: to say what something *tells* you (say, to ignore something) is not to say that that's what it *means*. When a man in a 1940s bar makes sneering comments in a scatological register, that's telling someone he's being challenged to a fistfight. That's not an analysis of what the comments *mean*, but to understand what's going on, we don't need to understand what they mean: we're interested in function, not semantic analysis. I'll return to this point on p. 211.

its ranking system is such an efficient guide to allocating user attention. When a cognitively important resource is in short supply, you would expect people to have developed ways of conserving it, or budgeting techniques, or, at the very minimum, ways of warning against unneeded waste. So it is almost inevitable that we would have a way of marking the limits within which this limited cognitive resource can be usefully allocated. And once we are aware of this functional role, and we cast around for the linguistic and conceptual device that occupies it in our culture, we recognize that this is what necessity does for us, and that it is, at any rate, one reason why we have it. Necessity is an attention management device.

Notice that the function explains the philosophically distracting emphasis on the near side of the contrast that necessity introduces.<sup>17</sup> The point of necessity is to tell you *not to bother* paying attention to the alternatives: to prevent you from wasting time on them. If you proceeded by marking the alternatives themselves, say, as impossibilities, you would have to waste your time and intellectual energy figuring out just what the alternatives being ruled out *are*.<sup>18</sup> That would defeat the point of having the device in the first place, and so we mark what we *will* want to be thinking about with a “do not go beyond” sign.

I’m not suggesting that *whenever* we want to ignore something, we mark its logical complement as necessary. (Compare: the state’s Department of Transportation has a great deal of advice for motorists that it chooses not to put on the street signs.<sup>19</sup>) I also don’t want to imply that such signage always prevents facts that many people consider to be necessary—the truths of logic and mathematics, just for instance—from coming in for a great deal of anyway certain

---

<sup>17</sup>During the heyday of ordinary language philosophy, J. L. Austin famously pointed out that when it comes to “real,” it is the contrasting states that determine the import of the utterance: when you call something “real,” you had better have in mind, explicitly or tacitly, one or another way it might fail to be, as when, announcing that it’s real money, you might on one occasion mean that it’s not Monopoly money; on another, that it’s not counterfeit; on a third, that it’s not a negligible amount of money (Austin, 1962, pp. 70f). Trying to make sense of “reality” as a kind of metaphysical solidity *so* special that it can only be conveyed by making egg-laying noises—that was a mistake. Philosophers should rather be attending to the specific shortcomings that “real” tells you have been avoided in one or another case, and (though I’m now saying a bit more than Austin himself did) the reasons we would have a marker word with this sort of flexibility. If I’m right, Austin’s treatment of “real” is a road sign pointing us toward the correct approach to necessity; here, too, it is where we’re *not* looking that determines what’s being conveyed.

<sup>18</sup>Why not just mark off the possible? There are many, many somewhat overlapping possibilities; if the point is to conserve your cognitive resources, embarking on marking them, one after another, would be a mistake. The necessary is the common denominator of the possible—that’s what the “true in all possible worlds” locution gets right—and so you mark *that*.

<sup>19</sup>Sometimes alternatives are excluded one by one, and we have a large, colorful, and not-always-modal vocabulary for tagging them individually: not just “impossible,” but “non-starter,” “dead on arrival,” and the like. In the sciences, “well-understood” is a dismissive epithet, marking a phenomenon or topic as undeserving of further research; it functions in certain ways as “necessary” does. (I’m grateful to George Gerpheide and David Humphries for advice on linguistic practice in the engineering and science worlds.)

sorts of attention; I will return to this point in Section 8.7, where I will introduce the discriminations we need. Moreover, I want to emphasize that the hypothesis we are entertaining is not advanced as a reduction of modality to anything else: what you are told to ignore, on this hypothesis, are *alternatives*: itself an irreducibly modal notion.<sup>20</sup>

I reintroduced necessity, or rather a precursor thereof, within a simplified model of what will shortly become a professional context, and while I am going to pursue the imaginary history of that progressively more specialized discipline, keep in mind the continuities of its necessities with the very low-key and everyday expedients upon which we routinely rely. A former student has proposed writing detective fiction in which, rather than establishing who committed a murder, the Poirot-like character would figure out where someone's keys were.<sup>21</sup> I would be an appreciative audience, because at my age far too much of my time is spent searching for the keys I had in my hand just a moment ago, or wondering whether I turned off the stove before leaving the house, or whether the tap was left running. In the ensuing conversations, necessity is invoked as a way of narrowing the space of possibilities, and proves valuable in just the way it did for our protoarithmeticians. ("You don't need to check the freezer. The keys *can't* be anywhere you haven't been. You haven't been back in the kitchen. The keys aren't there.") To be sure, not everyone is experiencing the gradual onset of dementia; and no doubt there are other ways of solving these problems.<sup>22</sup> But already we can see how to split the difference between nineteenth-century positivism and the pervasiveness of necessity: necessities which are not seen but observed—observed, that is, in the way religious holidays are—can allow an enormous improvement in time management and, more generally, contribute to efficiency in resource allocation.

### 8.3

Figuring out what one needs to pay attention to and what can be ignored is an intellectual achievement. Attainments of this sort are normally gradual. Faced

---

<sup>20</sup> Alternatives are possibilities, and while on the philosophers' usual way of thinking, these are absolutely modally flat, in real life we can intelligibly ask how an alternative could have been different; like claims about material objects, claims about possibilities generally have counterfactual conditions of application. The modal notions seem to be inextricably intertwined.

<sup>21</sup> Amanda Johnson's working title for the project is "Life's Little Mysteries."

<sup>22</sup> For instance, the so-called Internet of Things is the hype of the moment, and we are told that very shortly we will be able to ask our stove whether we left it on; Tile is a startup launching a product that tells you where your keys are.

Notice that these low-key necessities are sensitive to empirical facts that can easily change. If there is a child in the house, the keys *can* have gotten somewhere I haven't been, and might well be in the freezer. Philosophers tend to describe such necessities as "epistemic," as opposed to "alethic" or "metaphysical." We will soon be in a position to have an opinion as to how different these types of necessity really are.



with a novel environment or class of problems, one does not know at the outset what can safely be disregarded; the only reasonable policy is to proceed with open eyes. It is only over time—often, a very long period of time—that we can with confidence determine that one class or another of features of the environment can simply be tuned out. If we typically don’t learn all at once what we can afford not to think about, and if we use necessity to mark the line between what we have to pay attention to and that of which we can confidently take no notice, we should expect necessity to be progressive: as our understanding of a subject area increases, new necessities ought to be introduced. The more experience we have with something, the more necessity we will find.

We can discern the phenomenon not just in our state of nature story but in recorded history. William Whewell gives a remarkable example—“the echineïs, a small fish, which was said [during the times of the Roman empire] to stop a ship merely by sticking to it”—which he thinks exhibits “the absence of any steady apprehension” of Newton’s third law of motion. The notion that a fish might have endangered the shipping lanes in this way strikes us now as ludicrous, because impossible, but it does not seem to have struck anyone that way back then.<sup>23</sup> Again, once still more ambitious conservation principles were accepted, patent offices ceased working through applications for perpetual motion machines one by one; they no longer had to consider each on its own merits, wondering whether perhaps *this* one somehow worked, because perpetual motion machines had become impossible.<sup>24</sup> (Notice, however, that to this day they continue to receive patent applications for perpetual motion machines, and in largish numbers; as we will discuss presently, a necessity may stick inside a discipline, but not outside of it, or vice versa.) One further example: Biologists have learned to proceed on the assumption that any biological phenomenon will have a physical implementation. This is good practice, though only recently; a little over a century ago, vitalism was still a sensible option, because treating physical implementability as a constraint didn’t buy you all that much. Biologists express the assumption by saying that there *must* be a physical implementation, and philosophers mistake this for a metaphysical doctrine—an old-school metaphysical doctrine—which they call “physicalism,” when what it amounts to is an attention allocation policy: focus exclusively on physically implementable theories and explanations!<sup>25</sup>

---

<sup>23</sup>Whewell (1847, vol i., p. 263); possibly this is the “ship-holder” that Aristotle mentions in *History of Animals* II.14 (1984, at 505b17 or so). Whewell (1857, vol. i, pp. 188f) provides a lengthier version with amusing quotation from Pliny.

<sup>24</sup>I’m grateful to Jonathan Weinberg for the illustration. Once again, developments in science resemble developments in day-to-day life. When a small child insists that the crayon markings on the wall were made by her teddy bear, and not by her, her parents don’t need to consider how likely that explanation is: teddy bears *can’t* do things on their own. But our heritage of myths and folk tales indicates that this modal status is relatively recent.

<sup>25</sup>Physicalism, like naturalism, suffers from not amounting to a storable doctrine, and I used to suspect they both properly take an emotivist account: the one comes to something on the order of

To be sure, we need to qualify the expectation that as time goes on, more and more facts will turn necessary: the progress of necessity certainly is not monotonic. Whewell, who also thought necessity to be progressive, was not alive to the way that, since Kuhn, we have come to expect scientific revolutions to roll necessities back, and frequently enough for the so-called pessimistic meta-induction to instruct you to anticipate that *whatever* you now take to be a firmly established scientific result will ultimately be overturned.<sup>26</sup> I have been suggesting that necessity is often tied to areas of expertise; when areas of expertise vanish, as they do every so often, the no-longer-needed cognitive resources will include not only the map used to navigate a specialization, but the tags on the map that marked no-go zones; when the map vanishes, the tags on the map will vanish also. In addition, every now and again a specialization changes dramatically; its necessities can be lost and replaced this way as well.

If necessity is in this way progressive, any necessity has an inception date; there are conditions on which necessities ought to be suspended or revoked; some may even come with implicit expiration dates. For on the hypothesis we are developing, necessity is a way of telling people what not to think about, and those policies must occasionally be reconsidered. There would be, for instance, no scientific revolutions unless some scientists thought about the very things to which necessity markers tell them not to pay any attention.

But this is not how necessity presents: necessities seem to be unconditional, to be eternal, and in fact to justify our attention-management policies, rather than being determined by them. (For example, our policy of taking  $3 + 5 = 7$  to be worth no consideration whatsoever makes sense *because*  $3 + 5$  is necessarily 8: It *can't* be 7.) Consequently, if necessity is progressive, the phenomenology of necessity is evidently systematically misleading. I will return to this concern once we have extended our just-so story-in-progress somewhat further.

---

“Hurrah for the sorts of explanations you find in physics textbooks,” and the other to, “Hurrah for the sorts of explanations you find in science textbooks of any variety.”

Just in case the difficulty is unfamiliar, here's a very quick way of seeing how it goes, adapted and compressed from Stoljar, 2010. The first-cut rendering of physicalism or naturalism is usually something along the lines of: your explanatory ontology is restricted to the contents of, respectively, physics or science textbooks. Are they *today's* science textbooks, or *future* science textbooks? We can be pretty sure that what's in today's science textbooks will be superseded by tomorrow's. And who knows what's going to be in the science textbooks of the future? (Of course what's in the textbooks is supposed to be a proxy for the genuine criterion of physicalistic or naturalistic acceptability, but no one is able to produce *that*.)

But if the suggestion we are considering is on target, the force isn't just “Hurrah!” Rather, physicalism is a cognitive policy: dismiss biological hypotheses outright if they're not physically implementable! The necessity of physicalism is a way of implementing that policy. We know how to argue intelligently about policies of this kind, so, if that's right, we can finally *start* discussing the merits and demerits of physicalism and naturalism.

<sup>26</sup>Kuhn (1970) and Laudan (1981); for an overview of Whewell's ideas, see Millgram (2014a).

## 8.4

Although analytic philosophers have ways of distinguishing one type of necessity from another (e.g., logical necessity from metaphysical necessity, and both of those from nomological necessity), and although logic is the province of the logicians, physical law that of the physicists, and so on, our philosophers don't take the metaphysics of necessity to be driven even partially by division of labor. So I had better pause to motivate my willingness to accept that as a working hypothesis.

Imagine a primate species resembling *Homo sapiens* in many ways, but exhibiting much less adeptness in the division of labor. The difference we are after will be both in the extent to which this species' members can flexibly take on specialized tasks and roles and in their ability to *respecialize*, that is, to switch from one specialized task or role to another. *Piltdown Men* (just to have a name for these imaginary primates) do become farmers, and fishermen, and goat herders (so they can occupy different social or more generally ecological niches), but they fix on their specializations early in life, and there is only a small, fixed menu of options. A Piltdown Man cannot reprogram itself to adapt either to newly emerging niches in its ecology or even to an already-extant niche other than the one in which it started out: once a goat herder, always a goat herder.<sup>27</sup> These creatures are evidently a good deal more limited than we are, and, on the assumption that an organism's cognitive suite normally matches its ecological strategy, we should expect them to be cognitively more limited as well. So let's ask what sort of attention management techniques semi-humans of this sort ought to have.

Because Piltdown Men do not change what they do, and thus cannot adjust to novel circumstances, we had better imagine them occupying an environment that is stable over the long term; they wouldn't do very well—and wouldn't turn up—in any other environment. If the environment is stable enough for long enough, there will be time for both cultural and biological evolution to produce modes of representation that are efficiently tailored to the handful of niches that a Piltdown Man might occupy. For example, a farmer Piltdown Man will register seasons as fall, winter, spring, or summer; each condition disposes him to a suitable set of behaviors. In the spring, he has an urge to put down seedlings; then he takes great joy in weeding and watering, mulching seems like the natural thing to do, and so on. But there is no need for this creature to imagine a fifth season, with its own distinctive weather patterns and horticultural demands. *We* can do so; we might well, in the course of writing fanciful children's books, depict times of year in which the sky rains green glop.<sup>28</sup> But if his cognitive repertoire is fully optimized

---

<sup>27</sup>For variants of this thought experiment, see Chapters 3, 4, and 9.

<sup>28</sup>Seuss (1949).

(not that natural selection always, or even nearly always, manages full optimization), a Piltdown Man will simply be unable to represent “seasons” he doesn’t need to know about. So where, after the bedtime story, a human mother might tell her worried child that it’s *got* to be either winter, or spring, or summer, or fall (thus implying that there’s no call to toss and turn all night, fretting about the green glop season), a Piltdown Man, unequipped to represent seasons other than the ordinary ones, does not need to distinguish between the options to which attention must be paid and options that are guaranteed never to come up. Where we mark the limits of usefully allocated attention with flags like “necessary,” fully optimized Piltdown Men have no use for such a device: the alternatives on which we might otherwise inefficiently squander our attention are invisible to them. An Orwellian shorthand for this might be that evolution is able to teach Piltdown Man one or another form of Newspeak.

The demand for a cognitive device like necessity evidently emerges somewhere in the continuum between Piltdown Men and ourselves. Human beings, like Piltdown Men, specialize to occupy diverse ecological niches and, in doing so, produce niche-specific systems of representation that allow them to represent effectively the features of their environment to which their specialized strategy requires them to respond. Unlike Piltdown Men, humans move from niche to niche (as when former real estate agents become web entrepreneurs) and colonize new niches (as when an electrical engineer looks for business opportunities in the nascent electromagnetic space launch industry, or as when a former employee of a ski resort developer quits his job to write a web-based ride sharing application). When they do, they articulate the cognitive apparatus they use to navigate those niches in ways that are substantially more elaborate than what is required for the more roughly characterized, Piltdown-ish activities. When Old World peddlers in America became first grocery store owners, then supermarket managers, and then department store chain owners, they adopted ever more sophisticated conceptual tools. An abbreviated way of saying all this—adopting the “view from sideways on” of our species—might be that *human beings are serial hyperspecializers*.<sup>29</sup>

The additional functionality that we find in our design accounts for the use of the attention management device we have been investigating—and that in more ways than one. First, because serial hyperspecializers explore novel niches, they occasionally find themselves having entered a niche whose representational demands they do not yet understand. During that exploratory phase, a serial hyperspecializer will not know what it can afford to ignore, and

---

<sup>29</sup>The “from sideways on” locution is used in the Pittsburgh-influenced philosophical world to register the incoherence of an attempt to adopt an impossible point of view on yourself: as though you could look, not out of your own eyes, but at those eyes themselves, from a vantage point to one side of them. The phrase, used with this point, is suitable for a species of proto-humans incapable of making sense of or using a vanity mirror or a camera, but not for us.

so a great deal must be visible to it that it eventually (subject to qualifications I will return to below) can ignore entirely. Second, because serial hyperspecializers are transient occupants of the ecological niches they inhabit, there is often insufficient time to produce the optimization of a representational system: the one that has as an effect the invisibility of irrelevant facts, options, and so on. Third, serial hyperspecializers normally need to communicate with the occupants of other niches. The inhabitants of these other niches will often themselves represent a great deal that from the point of view of one's own niche can be ignored. Although communication across niches is usually conducted in a sort of pidgin, rather than the full vocabulary of either niche, it will often entail the ability to represent facts, state spaces, and the like that for one's own purposes are never salient.<sup>30</sup>

On the one hand, then, efficient navigation of a niche involves ignoring inessentials. On the other hand, as we have just seen, the ability to move from niche to niche means that serial hyperspecializers will typically have representational systems which make inessentials visible. The straightforward fix is to have a way of marking representable inessentials as ignorable. Serial hyperspecializers will have some device in their cognitive repertoire that works essentially the way necessity does. We are serial hyperspecializers; for us, necessity is that cognitive device.

In the life of such a creature, the demands of attention management should be expected to vary from specialized field to specialized field, and if necessity is the intellectual tool we have been describing, its shape ought to vary as well. Let's return to Stage II protoarithmetic, at which inductive generalizations about counting, such as " $3+5 = 8$ ," have been marked as necessary truths. There is, it gradually becomes clear, little point in taking them one by one; they are not independent and can even be treated as interderivable. Protoarithmetic invites systematization, and once that occurs, it becomes a poor choice to post the modal signage on a case by case basis. For suppose that " $3 + 5 = 8$ " and " $8 + 3 = 11$ " are now both treated as necessary, meaning that alternatives won't be considered, and suppose that the steps needed to derive " $3 + 5 + 3 = 11$ " from the former two sums have the same status. If " $3 + 5 + 3 = 11$ " isn't marked in a way that precludes reconsideration, and if you *do* start reconsidering it, pretty soon you're going to find yourself rethinking whether  $3 + 5 = 8$ ; since that question *is* closed, whether  $3 + 5 + 3 = 11$  had better be a closed question also. Alethic necessity is like traffic signage; the pressure we're now seeing is analogous to the realization that when traffic lights proliferate, the timings have to be carefully coordinated, often across an entire urban area.

Because there are many facts of the same general sort, and great advantages to putting them together into a single system, the workings of the system are demanding enough for competence in developing, extending and testing it to

---

<sup>30</sup>For pidgins, see Galison (1997); see also Section 4.7 in the present book.

be its own area of expertise. Moreover, protoarithmetical generalizations like “ $3 + 5 = 8$ ” have very broad application: you are able to use them not only for sheep, or for that matter rabbits, but for anything that, as we might put it, gets spoken of by using a count noun: shoes and ships and tins of sealing wax and cabbages and kings, just for starters.<sup>31</sup> But now, protoarithmetical expertise should not be a subspecialty of animal husbandry *and* of historians of royalty *and* of boatbuilding, and so on; that would lead to needless duplication in research and in training. The practical case is compelling: labor is divided, and the inhabitants of our mathematical state of nature designate a distinct social caste devoted to systematizing and extending the range of such generalizations.

We have to think of all those protoarithmetical generalizations as originally being defeasible. Three sheep and five sheep are eight sheep, and twelve sticks and thirteen sticks repeatedly come out twenty-five sticks, and two rabbits and another two rabbits come out four rabbits—but not if a border collie loses a sheep, and not if a rabbit dies or runs away, and not if they breed like, well, rabbits, and not if you drop a stick, and not if a stick breaks . . . and as is characteristic of defeasibility, such lists of other-things-equal conditions are open-ended.

Accordingly, it will be natural to have the division of labor we are contemplating be accompanied by a division of responsibility for the defeating conditions of protoarithmetical inferences. It’s not reasonable to expect someone who is devoting himself to expertise in the rapidly growing protoarithmetical system to understand its many areas of application well enough to exhibit command of their characteristic defeasibility conditions.

---

<sup>31</sup>What qualifies for being spoken of by using a count noun? Aristotle has a very interesting suggestion: something is *reproducibly dividable* just if it can be partitioned into parts that are themselves what it is. (An example of the intended contrast: if you cut Timothy in two, actually or in imagination, the parts aren’t each also (a) Timothy; but if you divide up some water, the parts are also water.) Now we can say that an individual is something that isn’t reproducibly dividable, and Aristotle goes on to argue, very elegantly, that individuals are what can’t be predicated of anything else. (For a helpful reconstruction, see Frede 1987.)

I’ll use an aside as a warmup for the next point. Notice that if Aristotle is right, work in action theory that concerns itself with how many actions are there when you, say, flip a light switch is misguided. When you cut an action into parts, those parts are also actions; therefore, if Aristotle’s criterion is on the ball, actions aren’t individuals; therefore it is a mistake to try to count them.

In footnote 8, I suggested that possibility seems to pose problems over and above those involved in finding a philosophical spin to put on necessity; now I can suggest one reason why. Alternatives or options or possibilities are, so to speak, modally thick; while metaphysicians of modality by and large assume we can work with possible worlds that are sliced *absolutely* thin, for reasons developed in Chapter 7, possibilities of that sort aren’t philosophically available. Thus, when you slice a possibility down the middle, what you get back are also possibilities, albeit somewhat thinner. (If Sandra were late, that could be due to traffic on the freeway, or due to construction on 700 South; both of those are possibilities, too.) If we accept Aristotle’s criterion, possibilities are not individuals, and we should not try to count them—even though it is often perfectly reasonable to think of yourself as considering, or choosing between, this one and that one.

The protoarithmeticians' clients—often experts in their own domains—are appropriately made responsible for controlling the defeasibility conditions of protoarithmetical facts. It's the shepherds and rabbit farmers who need to know when three sheep and five sheep don't make eight sheep, and when two rabbits and two rabbits don't make four rabbits, not the mathematicians. Other than those *I Ching* users who cast their hexagrams the old-fashioned way, there are few specialists in bundling sticks; but even when there are no relevant specialists, it's the lay clientele that are likewise made responsible for defeaters in nonspecialized applications. Thus the demarcation of professional boundaries leaves the protoarithmeticians with a subject matter in which inferences are indefeasible. What would formerly have been considered empirical counterevidence to the protoarithmetic generalizations is henceforth construed as failures in their *application*. Once clients treat these generalizations as schemata they apply to variegated empirical circumstances, the protoarithmetical caste no longer needs to think about the pitfalls the clients encounter when they do so.<sup>32</sup> (Call this *Stage III* protoarithmetic.)

This amounts to dictating very different patterns of allocation of attention inside and outside of the specialist caste. Clients outside the specialist caste do not entertain alternatives to the protoarithmetical schemas with which they are provided; as I will shortly suggest, insiders nonetheless may. The outsiders do watch out very carefully for what they understand to be application issues; the insiders do not. Within the caste, necessity has come to serve a special purpose, that of marking the indefeasibility of a class of inferences: that is, protoarithmetical specialists no longer have to pay attention to defeating conditions. (We can call the version of necessity that emerges in the transition to Stage III protoarithmetic *protological*.)

Let me highlight two features of the story to this point. First, in the just-so history we have sketched, necessity was appropriated to mark indefeasibility

---

<sup>32</sup>Return also to the question, raised in footnote 13 above, of how high our protoarithmeticians can count. If no one ever needs to count more than a thousand sheep, the system of numerals can give out around there. But the shepherds need one supply of numerals; the arborists another; the stargazers yet another—and it would be a mistaken allocation of tasks to require the specialized caste of protoarithmeticians to know about the very specific needs of indefinitely many sorts of specialized client. Accordingly, we can imagine the protoarithmeticians establishing an Academy of Numerals, responsible for assuring that symbols suitable for counting with are available no matter *how* high their clients count. Once this step is taken, protoarithmetic has transitioned from its small-finite (and weakly finite) phase to a science with a more familiar scope, that of the natural numbers—and it seems to me that it is in this transition that the distinction between numbers and numerals is first genuinely useful.

It is plausible that the infinitude of the natural numbers is our very first infinity. If our state of nature story reflects how we came to insist on an infinitude of numbers, then our first encounter with infinity arose out of division of labor. But, of course, state of nature arguments don't purport to be history.

of inference; we should treat that as a representative instance of the quirkiness that the formally very different techniques of disciplinary interface design will exhibit, as variously inflected versions of necessity are arranged to manage the needs of other sorts of division of labor.

Second, notice how we can expect these decisions to play out among the philosophers in the world of our narrative. The theoretical system and those inference schemas are the responsibility of a specialized caste, and defeasibility in application is, as a matter of policy, fobbed off on outsiders; thus two and two are *always* four. And it's not the business of the protoarithmeticians to concern themselves with the tolerances of different sorts of measurement, or with whether some of the water in the two pairs of cups is evaporating as we count; so two plus two are always *exactly* four—not 4.000008769786.<sup>33</sup> Because it's not the business of the protoarithmeticians to know how high this or that client needs to count, at this point in the story protoarithmetic never gives out, no matter how large the numbers get. Casual onlookers within the society will easily come to see their protoarithmetic as exhibiting a remarkable, almost supernatural (they are tempted to exclaim) *crystalline hardness*. It will seem to their philosophers to cry out for a metaphysical explanation. But the peculiar hypercertainty, hyperprecision, and infinite reach of their protoarithmetic is a feature of their assignment of responsibilities, one that is itself a side effect of disciplinary specialization.

Now ask yourself: suppose, in the protoarithmetic state of nature, some arithmetical truths had *really* been metaphysically necessary (whatever that is supposed to mean). Would the recognition of the real necessity of their mathematics have looked any different than the outcome of facing up to the practical demands of Millian protomathematics, in the state of nature that I have described?<sup>34</sup>

---

<sup>33</sup>Readers of a certain generation will remember the Pentium floating-point bug, which produced outputs like this one; that the chip was regarded as unacceptable shows us to have internalized this requirement ourselves. For a sense of how demanding expertise about measurement can be, see Tal (2014).

<sup>34</sup>This is a good place to distinguish the view I am developing from an illustrious predecessor's; of course, there are various differences, but here is one. In Rudolph Carnap's way of thinking about necessity, all the work would be done, in the story I am telling, by the convention that what look like counterexamples are henceforth application errors. Whereas as I am imagining it, the convention is only workable because it is layered on top of—and counts as a modest adjustment to—the empirical track record.

Aspects of the view will remind readers of the later Wittgenstein as well: for instance, harking back to the remark that while persons properly trained in multiplication mostly *do* get 625 when they square 25, the empirical fact is *further* "hardened into a rule" (Wittgenstein 1983, VI-23). While I am sympathetic to a number of moves that Wittgenstein makes, note this difference in where we hope to come out: where Wittgenstein hopes to dispel the necessity, which he thinks of as an illusion, I mean to explain its purpose.



## 8.5

Philosophers who concern themselves with necessity (they usually describe themselves as metaphysicians) don't depict necessities as I have been proposing. Necessity, on the hypothesis we are considering, has the function of marking the outer limits of attention; that is, it is signage that subserves a policy of cognitive resource management, in something like the way that stop signs and one-way-street signs subserve policies for traffic management. Where the question is one of policy, at bottom the arguments about what is and is not necessary must be practical: they must be arguments, perhaps turning on the relative merits of schemes of signage placement, about *what to do*.<sup>35</sup> However, throughout the semantic-analysis tradition's engagement with modality, the arguments have been preponderantly theoretical: that is, they have been conducted as argument about *matters of fact*. And as I have already observed, they do not for the most part take necessities to be progressive or revisable.

An instruction to ignore something permanently is tantamount to an injunction to forget about it completely. Necessity is signage that tells you to ignore all alternatives but those in the designated range; so it amounts to an injunction to forget about alternatives outside that range. Now, a policy is hard to keep in mind unless you remember why you are following it. The justification for an attention management policy of this sort will almost always advert to the alternatives being excluded: for instance, that we don't need to worry about *them* because *they* come up rarely; when *they* come up, we will classify *them* as application errors, and so on. Thus it is very likely to ensue that necessities eventually seem to their users not to be aspects of the implementation of a policy, after all. Moreover, when such an instruction results in rendering everything outside a designated range effectively invisible, it will come to seem that only the designated range is *there*; accordingly, the constraints on what possibilities are to be considered will appear to be, not the effect of a policy, but a matter of fact—presumably, of metaphysical fact.<sup>36</sup> Wittgenstein thought us to have a propensity for metaphysical illusion, across the board, but whether he was right about that or not, the impulse to modal reification is a hard-to-avoid side effect of the functionality of the device we are analyzing.

---

<sup>35</sup>Where such decisions are practical, it's much too easy to decide—as does, for example, Rayo (2013, pp. 62, 132)—that how they are made must depend on what goals or aims one has. That's to adopt the instrumentalist conception of practical rationality, on which any reason for doing something must invoke a goal; the move is often made inadvertently when philosophers who are specialized for metaphysics and epistemology, and so who don't think about practical rationality themselves, invoke practical considerations in their theorizing.

For that matter, it is a pitfall in philosophy of science as well: see Millgram (2006a) on van Fraassen's version of the mistake.

<sup>36</sup>“Effectively”: I will explain the qualification I am now marking in Section 8.7 below.

Signage generally needs to be kept simple, and that is especially true of signage used to manage the allocation of attention.<sup>37</sup> These days, it wouldn't be surprising to come across a street sign reading

### NO DISTRACTED DRIVING

Imagine the lower half of such a sign covered with fine print:

Weekdays 7:00–9:00 a.m. and 3:00–6:00 p.m. Handicapped drivers and G Permit holders exempted. Urgent calls to babysitters permitted. Online maps permitted [. . . and so on and so forth].

Any driver who tried to decipher the sign would thereby be distracted, and our attention-management signage faces tradeoffs that are a good deal like that. Again, the signage always has a *de facto* start date. Serial hyperspecializers need to keep an eye out (just a bit) for signs of instability in what they can and should for the most part ignore, especially for signals that tell them that the niche they occupy is changing, that it's no longer viable, and that it's time to jump ship and start looking for a new specialization. And those signs a niche is changing are the very ones that were marked as necessary—that is, whose alternatives cognitive thrift dictated ignoring. However, precious little of this—realistically, none of it—can be explicitly registered: can be, as it were, posted on the signage itself.<sup>38</sup> Consequently, necessity will misleadingly appear to its users to be eternal and unconditional, which goes a long way toward making it look like a very special sort of fact. And accordingly, because it is crucial that the policies it subserves be sometimes violated, adding a spoonful of doublethink helps necessities go down in the most useful way. I'll return below to the question of what strategies can keep the dose of doublethink to a minimum.

Unlike street signs, our attention management signage is largely internalized, and let's think for a moment about what appearance we can expect the cumulative signage to present over time. Suppose that we're right in taking necessity to tell you what to ignore. Once again, an argument for a policy the signage subserves will normally have to mention the very items that its conclusion dictates ignoring. That means that if the signage is effective, the

---

<sup>37</sup>Policies control choices, and choices are voluntary. Attention, however, is only quasi-voluntary; as we all know, sometimes you just can't pay attention, and to tell someone not to notice something is thereby to have them notice it. A full treatment of the deep complication that I am gesturing at here can be postponed to another occasion; in the meantime, it suffices that attention management policies are the stuff of ordinary, day-to-day life.

<sup>38</sup>Following up on footnotes 2 and 8, this is another reason why the imaginative blocking that Nichols takes to be the primary anchor of impossibility judgments is best treated as a secondary technique for demarcating necessities. It exploits the rule of thumb: If you can't imagine it, you don't need to pay attention to it. In certain very well traversed domains, this is a decent rule of thumb, but in general it's a *terrible* rule of thumb. What you understand least well is what you will have problems imagining, and ignoring what you don't understand is—as the youth of yesteryear put it—*doom*.

arguments that render it intelligible will eventually be lost, as some of their indispensable components slip out of range of everyone's attention.

When metaphysicians appeal to argument or intuition to determine, in the traditional way, what is metaphysically necessary, they are best construed as developing a map of where those who went before them placed the attention management signage. Because forgetting the reasons for designating one thing or another as necessary is a nearly inevitable side effect of the device, metaphysicians are almost always unaware that this is what they're doing: their internalized opinions about what is necessary and what is not appear brute to them, and the results are quite peculiar. Necessity-related "intuitions" amount to internalized signage. Metaphysicians take their intuitions to be revealing deep, invisible, and thus philosophically mysterious facts about the world. But they are no more uncovering mysterious metaphysical facts than someone who notices a one-way-street sign is uncovering a philosophically mysterious fact about the intrinsic, metaphysical one-wayness of the street.

When a side effect of posting signage is that you forget why you did so, it becomes very difficult, over the long term, to adhere to a consistent policy for placing signage of that generic kind. We can expect each round of attention management signage to be erected without much reference to the considerations and regularities that governed preceding rounds. And when we survey modal opinions, this is indeed what we see. Some of our inherited necessities are posted on Aristotelian substances—roughly, species and their members—and support species-genera inferences (e.g., tigers are animals, this is a tiger, so it's an animal; animals eat, so tigers eat). However, essential properties of a species may empirically go missing in individuals belonging to it. (Tigers have four legs, but you have to check whether this particular tiger does: it may have lost a leg in a boating accident.<sup>39</sup>) Some of our more recently discussed necessities mark individuals or natural kinds and *do* entail observable facts about particulars: if water is  $H_2O$ , then this glass of water here will *be*  $H_2O$ , no exceptions allowed; if Alice is essentially the child of Belinda and Charlie, then Belinda and Charlie *are* Alice's parents.<sup>40</sup> Some of our necessities

---

<sup>39</sup>Thompson (2008, Part I) is a recent reconstruction of the logic of this sort of necessity; Furth (1998) is a very thoughtful guide to Aristotle's own version of them.

Notice that Aristotelian essences support a variation on the sort of indefeasibility of inference that we encountered in Stage III protoarithmetic. If house cats have as an essential property that they live by eating meat, in particular rodent meat, which they catch by running after it and dispatching it with their teeth and claws, from knowing that an animal is a cat, a biologist of Aristotle's time would be licensed to infer that this animal has this property. My own cat does very little of that: I feed her cat food. What in protoarithmetic is handled by the ascription of mistakes in application here is fielded by Aristotle's contrast between potentiality and actuality; she has these properties potentially, in that she could hunt mice, and presumably would, if I stopped feeding her.

<sup>40</sup>Kripke (1980, pp. 112ff), and see footnote 47 below. There were belated qualms: Putnam (1990a, p. 69); Machery et al. (2004).

mark facts or statements, rather than objects or types: that  $2 + 2 = 4$ , for instance. (Necessity, as philosophers sometimes say, comes in two flavors: *de re* and *de dicto*.)<sup>41</sup> Some necessities tag rules of inference: they say, more or less, that once we grant the truth of the premises, as we draw the conclusion there's no need to worry that the inference might somehow be derailed. Some mark unavoidable presuppositions, which, however, may or may not be true.<sup>42</sup> And there's no reason to think that the list we've started in on is complete. After such a history of so many new beginnings, the internalized signage will come to be something of a palimpsest.

In Germany, the high-school class project is often to designate, clear, and mark a hiking trail through a forest; but of course, what each year's class does is create a *new* trail, and it is nobody's responsibility to maintain the trails laid down in previous years. The end result is unsurprisingly a jumble of half-discernable trails, which the hiker relies on at his peril. Our inherited attention management signage is like *that*. A theory of modality contoured to be responsible to our intuitions about what is and is not necessary—modal metaphysics understood as a *theoretical* enterprise—is antiquarianism, an enterprise on a par with producing grammars for dead languages. This sort of metaphysics of modality should be about as useful as the trail map of the Saarbrücken *Urwald*—that is, not very.

---

<sup>41</sup>For useful contrasts between the rules that govern *de dicto* and *de re* necessity markers, see Brandom (1994, pp. 499–508).

<sup>42</sup>We analytic philosophers have largely lost track of the notion of transcendental necessity, the characteristic upshot of transcendental arguments. These arguments have something like the following form: either you're thinking (or acting, or engaging in a properly scientific investigation, or whatever) or you're not. While you're deliberating about whatever topic it is, you have to presuppose that you *are* thinking (or are a unified agent, etc.); if you're not really thinking (or acting . . . if, say, you're a paranoid schizophrenic, or incapable of coming to a decision and then going through with it), you might as well give up now. You can only be thinking (be a unified agent, and so on) if such and such a precondition holds. Therefore, such and such is (transcendentally) necessary; meaning, more or less, that, anyway in the context of your deliberations, you have no choice but to presuppose such and such. (For more spelling out, see Millgram 2009a, Section 4.1.)

Now, whatever you think of an argument form that was once the engine of an important philosophical tradition, note this much about its signage: a transcendental argument can enforce commitment to a conclusion that is *false*. (In the possible-worlds way of talking, it is so little the case that a transcendentially necessary conclusion is true in all possible worlds that it may fail to be true in a world in which the argument is successfully executed.) For instance, you can step through the argument to its conclusion even if you are a paranoid schizophrenic; someone who fails to be a unified agent may step through an argument showing it to be necessary that he is unified (i.e., showing that he should entirely disregard the possibility that he isn't). Please note that I don't mean the observation as an off-the-cuff refutation of transcendental argumentation; on the contrary, the feature I have just been emphasizing fits nicely with the function of necessity as we have been limning it. Sometimes, you can mark alternatives as not needing further attention even when there is some chance that they are in fact the case.

## 8.6

The treatment we are developing goes against the grain of analytic approaches to necessity. Again, these construe necessities as a very special sort of fact, and hold theories of necessity responsible to a profile of modal intuitions. We have already indicated how this sort of theorizing is to be understood from the perspective of cognitive function analyses: the reification of attention management policies into mysterious matters of fact is a hard-to-resist side effect of deploying such a device, and the intuitions are systematically misleading. Still, there is a special case of the old-school metaphysics of modality that requires a dedicated discussion, namely, theories of modal inference, and, in particular, work in modal logic.

Such theories in the first place systematize intuitions about modal entailment. (I mean views like: if  $p$  is possible, then it is necessary that  $p$  is possible.) They are also typically constructed to accommodate various intuitions about essential properties, counterfactuals, and so on. Modal logic serves as the acknowledged anchor for discussion of modality, and one reason for that does seem to me to be intellectually respectable. The pairing of formalized modal logics with model-theoretic treatments of them had something like the effect on philosophers' attitudes toward modality that Georg Cantor's contributions to set theory had to just about everyone's comfort level with infinity; where before everything had been murkiness, confusion, and uncertainty, suddenly the concepts seemed to have been tamed.<sup>43</sup>

We should not be surprised that theories of modal inference are possible. I remarked that attention-management signage has to be carefully coordinated, in something like the way traffic signage must be. And we can sometimes inferentially exploit systematically placed signage of this kind. As you drive down a one-way street, you can be sure that you will not find a one-way street sign pointing in the other direction, that at the next intersection it will not turn out that all the other streets entering the intersection are marked one-way back into the intersection, and that lines on the asphalt will not mark off lanes for oncoming traffic. As you drive through that intersection, you will notice that the stoplights are coordinated with one another: for instance, when the light in front of you shows green, the vehicles on the cross street see a red light, and the pedestrians in the crosswalk see their own version of a red light. Consequently, if you have a green light, you may conclude that the cross traffic has a red.

---

<sup>43</sup>I'm grateful to Alasdair MacIntyre for reminding me how modal logic looked back in the day.

While it would be remiss of me not to flag a sociologically important but less respectable explanation of the centrality of modal logic in metaphysics, that is, its appropriation for purposes of Zahavian or handicap-principle signaling (Zahavi and Zahavi, 1999), we can set those motivations aside here.

Modal logics analogously specify rigorous rules for interlocking attention management signage (and signage that serves other purposes also, since recall that we're postponing treatments of possibility, and so on, to a later occasion). Signage coordination rules of these kinds support inferences; just as when you see the green light, you can infer that the other drivers are seeing red lights, from  $\Box(p \& q)$ , you can infer that  $\Box p$  and that  $\Box q$ —that is, that each of  $p$  and  $q$  is also marked with an ignore-alternatives-to-this sign. Are these inferences from premises or to conclusions with a special sort of content? If the point of posting necessity signage on  $p$  is to restrict attention to  $p$ , one frequent preliminary will be to argue for the truth of  $p$  itself. If what you are really after when you do so is showing that alternatives to  $p$  need no further attention, you will want not just an argument for  $p$ , but an especially compelling argument.<sup>44</sup> Arguments that  $p$  is (should be treated as) necessary will seem to require starting points that can themselves be marked as necessary: premises that have no alternatives which themselves need attention. This easily gives rise to the impression that one necessary fact is being derived from another. But the appearance of necessity flowing over the implications—in something like the way that truths are derived from truths in valid inferences, or as Hume once suggested, that “oughts” in the conclusions of inferences arise out of “oughts” in the premises—is misleading. We should not assume that the necessity on which one does often rely in an argument for the necessity of  $p$  is part of the *content* of the premises of one's argument for  $p$ . So far, it's just *signage*, part of the implementation of a decision—a policy—not to waste any more time and intellectual effort on alternatives to those premises.

Modal signage might undergo a transformation into a system of modal symbols with semantic values, of the sort that would warrant developing a semantic theory. Perhaps some of it already has; it's hard to tell. (Sometimes, over time, function is converted into content; although zero was originally simply a placeholder—a way of marking an empty column, and so preventing “101” from being confused with “11”—it eventually became a numeral in its own right. Exactly when did that happen? It's hard to tell.) But even when some fragment of our landscape of modal intuitions seems to lend itself to semantic theory of the traditional kind, because we've been unselfaware about what our modal vocabulary is *for*, and because recursive definitions quickly give you arbitrarily large constructions, it has been very easy for the formal machinery of such a theory to outstrip its subject matter. Once you remember that this is about guiding attention, it's hard to believe that anything is represented by all but the most trivial of the structures that modal logic invites philosophers to imagine. A systematic treatment of the conventions

---

<sup>44</sup>Thus, a typical derivation rule will have it that if  $p$  is provable in the system,  $\Box p$  is a theorem as well.

governing the placement of one-way-street signs might have consequences for how a million-street intersection could be signed, but these would lack application; there are no million-street intersections. Likewise, if necessity is guidance for attention management policies, and no creature like us could have a policy for managing attention to a policy for managing attention to . . . [iterate 100 times], then there will be no occasion for a theory that tells us what  $\underbrace{\square\square\dots\square}_{100 \text{ times}}p$  entails.<sup>45</sup>

If that is right, we should be having second thoughts about the clarity that modal logic seems to have brought to our modal discourse. The tendency to reify attention management policies has encouraged philosophers to imagine they had a subject matter for a modal logic to represent. But what these formal theories seem to be about is vastly larger than any policy we are likely ever to have on hand. Briefly, I am skeptical as to whether our modal signage should now be thought of as having or needing a semantics.

## 8.7

The hypothesis we're entertaining about the cognitive function of necessity is an occasion to reconsider its place within a network of closely connected concepts. Necessity, roughly, there being no other way things could be, is contrasted with contingency. A priori knowledge is contrasted with a posteriori or empirical knowledge; if you know something a priori, you can know it without looking, whereas you have to come to have a posteriori or empirical knowledge by observation. Finally, analytic truths hold roughly in virtue of the meanings of one's words—let's say, more broadly, that they're side effects of using the conceptual toolkit one has. They're contrasted with synthetic truths, which aren't merely more or less glorified tautologies.<sup>46</sup>

The first members of these pairs tend to travel together, as do the latter members, and until recently the members of each group weren't all that carefully distinguished. But over the last couple hundred years, that has been changing, and so much effort has gone toward distinguishing notions that

---

<sup>45</sup>If the division of labor in our protoarithmetical state of nature story was legitimate, why can't the effects we're considering be accounted for in the same way? One caste of specialists, the modal logicians, is made responsible for developing the formal logic; another caste, the philosophers, is made responsible for its applications. And it's not the responsibility of the modal logicians to know how *large* any of the clients' applications are going to be. But that allocation of labor and responsibilities was warranted in the case of protoarithmetical because there were a great many different client specializations, all of which could make real use of the various formulae for addition; it was unreasonable to have the mathematicians worrying about the problems that arise in one after another of the indefinitely many client specializations. Modal logic, however, realistically has only *one* client field: philosophy.

<sup>46</sup>In addition, the contrast between what's conceivable or imaginable and what's inconceivable and unimaginable is on the face of it a matter of mental or cognitive abilities. For reasons gestured at in footnote 8, it seems premature to me to take up this contrast now, but see footnote 50 below.

normally travel together that we have begun to lose sight of the fact I want to reemphasize: that they *do* normally travel together.<sup>47</sup> Can we explain why philosophers find notions in this neighborhood traveling in these two rough affinity groups—once again, allowing for the cross cutting at which I have just gestured?

Let's return to our ongoing narrative of the emergence of protoarithmetic from a mathematical state of nature. We left the process at its Stage III, where the severally necessary sums had been systematized into a theory whose development and expansion were assigned to a specialized caste. Because the individual inductive generalizations incorporated into the system of protoarithmetic are mutually reinforcing, the empirical evidence for the system as a whole is now visibly overwhelming.<sup>48</sup> Moreover, we remarked on a dramatic innovation introduced into Stage III protoarithmetic, namely, conventions for handling what were formerly regarded as counterexamples to its empirical generalizations; they are now dismissed as application errors on the part of extradisiplinary clients. There are thus two obvious further moves to make, having to do respectively with data collection and data retention, which will together move protoarithmetic into its Stage IV.

Both mathematical sophisticates and laymen have grown to think that alternatives to the systematized theory of counting and sums are no longer deserving of consideration, that is, their protoarithmetic has become a domain of necessary truth. But when you are not considering alternatives, there is no point in collecting further empirical evidence—or even in holding onto the empirical evidence that has already been collected. In order to save space and forestall unneeded rehashing, it's efficient first of all to determine that no further data will be collected or processed, and also to decide to dispose of those warehouses of evidence that supported accepting the protoarithmetical theory in the first place.

It's a priori when you can know it without looking—that is, at empirical evidence—and we can now distinguish two different senses that “You know without looking” might bear. In one, it's effectively an instruction, directing

---

<sup>47</sup>Kant kicked off our current awareness of the differences by noticing that you could know some things without looking, even though they didn't amount to expressions of conceptual containment relations—that is, that you could have synthetic a priori knowledge; for a reconstruction of the analytic-synthetic distinction in Kant, and his understanding of why arithmetic is not analytic, see Anderson (2004). In the mid-twentieth century, in discussions of natural kinds, and what was then dubbed the new theory of reference, it was argued that although water is necessarily  $H_2O$ —it couldn't have been anything else—that fact had to be discovered empirically: there are necessary but a posteriori truths (Putnam, 1975b). And finally, because, say, a unit of length can be defined by designating a given object as a standard, there can be a priori but contingent truths: you know without looking that the standard meter is one meter long, but it could have been a different length, for instance, if the bar of metal had been cut differently. (See Kripke, 1980, pp. 56, 76; see also pp. 36–38.)

<sup>48</sup>Gottlob Frege complained against Mill that many alleged arithmetical generalizations are never observed (Frege, 1978, pp. 9f). But of course this is unfair: it is the systematized theory as a whole that gains strength from the observations.



you that there is no need to look any more, and that you can dispense with further empirical investigation. In the other, it announces that you *never* had to go look: that empirical investigation was *never* necessary. But now, taking the former on board will end up giving you something that looks a lot like the latter. Suppose that once an empirical (a posteriori, contingent) fact becomes very well entrenched, it is cognitively efficient, first, to mark it as necessary, and then, as computer scientists say, to garbage-collect the empirical evidence for it. Once all the evidence is deleted, such a fact would *strike its users* as a priori, in the never-had-to-look sense, even though it belonged to the roster of a priori facts solely due to its being empirically well-grounded. Come to think of it, since never-had-to-look truths look the same, from the inside, as looked-hard-and-then-erased-the-data truths, and since the latter sort of truth is philosophically unproblematic, whereas the former is one of the perpetual philosophical mysteries, why suppose that any important truths really do belong in the first category?

*A priori* is to *necessary* somewhat as *knowledge* is to *truth*: if “knows” is a certificate of confidence, one that tells us that further checking isn’t needed, and you can just go ahead and use the truth that is being certified, “a priori” tells us that we were *so* confident that we could go on to erase our memory of ever having checked.<sup>49</sup> We should expect necessity and apriority to travel together, though with apriority lagging somewhat behind.

So, returning to protoarithmetic, division of labor will produce a further differentiation of the allocation of attention. Largely because it comes to be so thoroughly empirically confirmed, it also comes to be thought that no further empirical confirmation of the system of empirical generalizations is needed. When you really don’t have to pay *any* attention to alternatives, you can preemptively decide against any future data collection, and then go ahead and empty out the archives. It becomes as though no one had ever needed empirical observation to arrive at this family of truths at all. However, because the job of the specialized caste of protoarithmeticians is to systematize the compilation of established sums, and to extend it further, protoarithmeticians retain those derivations within the system: they review them, they teach them to novices (as demonstrations or proofs of new sums), and they work at producing smoother, cleaner derivations—and systematizations (axiomatizations) of the protoarithmetic theory as a whole.

That is, where extradisciplinary clients simply accept the sums, and perhaps very basic rote procedures for computing them, as not requiring

---

<sup>49</sup>Other philosophers have advanced related but contrasting notions of the revisable a priori: Friedman (2001, pp. 71–73) takes a priori principles to “formulate the necessary conditions or rules for establishing empirical knowledge”; consequently, they “cannot themselves be similarly established, and it is in precisely in this sense that they are prior to or independent of experience.” Of course, just what these rules are will change from Kuhnian paradigm to Kuhnian paradigm. White (1963, pp. 287f) proposes using the term “a priori” to mark the decision henceforth not to revise a statement.

reexamination of any kind, protoarithmetic specialists, while no longer allocating attention to *empirical* evidence for the theory, focus ever more attention on arguments conducted within the theory for its theorems, old and new. And perhaps (qualifying now that penultimate observation) a handful of them—the proto-Lobachevskys and proto-Riemanns of the tribe—attempt to develop alternative systematizations that would require revoking the aprioricity of protoarithmetic and consulting empirical evidence once again.

A few steps back I suggested that deploying necessity involves a certain amount of doublethink: you conserve cognitive resources by obeying the signage, but it is dangerous not to pay attention, every now and again, to whether the signage is correct. Specialization can mute the tension, even if not do away with it entirely. In Stage IV protoarithmetic, the specialists allocate some of their efforts to determining whether the systematized theory and its attention management signage need a rethink; they are the ones who have a use for the notion of a priori knowledge supported by nonempirical argument. Whereas for the outsiders, protoarithmetic necessity really *does* just tell one, pay no mind to alternatives. If I am correct, these asymmetries are not atypical; when you are one specialist speaking to another, you can normally direct your colleague's attention without deploying devices like necessity. As a philosopher, I can explain to philosophers what they should pay attention to, and why, and also what they should ignore, and under what circumstances. But when outsiders will not follow or be able to make use of this sort of direction, practically speaking you have no alternative to giving them the baldest possible instructions as to where not to look. Accordingly, necessity will frequently be for *extradisciplinary* consumption.

In the move from necessity to aprioricity, there is a further step that is sometimes taken, and I will briefly describe it without saying how successors to our protoarithmeticians might attempt it. The notation itself can sometimes be reconfigured so as to render excluded alternatives not coherently expressible. That has not quite happened in this case: “ $3 + 5 = 7$ ” still counts as a well-formed and intelligible expression, even if one that is known to be false.<sup>50</sup> But although analyticity (and at the limit, what we can call *Tractarian* analyticity) might seem to be a capstone exercise that cements necessity in place, and while we do for the most part *say* that analytic truths are necessary, in fact

---

<sup>50</sup>Ismael (2013) discusses the strategy and its pursuit in the physical sciences. This sort of paring away of expressive options may be foregone, however, even when we redesign our representational systems for improved efficiency: we haven't adopted a notation that makes it impossible to write out a forbidden fraction like  $\frac{1}{0}$ —we just teach our pupils that they can't divide by zero. As we know, in the example we are developing, there are Gödel-related obstacles to carrying this process to its completion. (We could perhaps ignore these in our account of our imaginary primitive society, by keeping their protoarithmetic too weak to support the incompleteness theorems.) However, some of the *effects* of rendering alternatives not merely effectively invisible seem to have taken in our own case: for a vivid illustration, see Gasking (1953).

analyticity removes the occasion for necessity. If the exclusion is written into the very words, and if you can no longer express it, you don't need to instruct anyone to ignore it. Optimizing the allocation of attention can proceed either by explicitly marking the line between what we need to pay attention to and what we don't, or by constructing systems of representation that render cognitively invisible would-be objects of attention that amount to no more than distractions. Sometimes we do both; when we do, the latter usually comes after the former; doing both is likely to be redundant.

If necessity, apriority, and analyticity have the functional characterizations I have attributed to them, they are not simply the contrasts to and opposites of the contingent, the empirical, and the synthetic. It is foolish to decide that we do not need to pay attention to alternatives, that we do not need to keep track of whatever empirical evidence we have collected, that we do not need to collect any more of it, and that we can build substantive presuppositions into the intellectual toolkit that shapes our thoughts, unless we have already looked very hard, very closely, and for a very long time. The former categories emerge *out of* the latter, as their epitomes rather than their contraries.<sup>51</sup> That means that we will often encounter a characteristic pattern of two-way inferential connections—a pattern I will discuss by taking up Mark Johnston's Missing Explanation Argument.

## 8.8

A “secondary quality” is a feature or property that you can observe and that you understand not in the first place by looking at *it*, but by attending to patterns in the reactions of its observers. At *Euthyphro* 9e–11a, Plato's Socrates argues against what we might now call a divinity-oriented secondary-quality account of the holy or pious.<sup>52</sup> On Euthyphro's proposed definition, the holy is whatever it is the gods like or love: sacrifices, perhaps. But generally, whenever someone (and even a god) likes or loves something, there's some feature of what he loves or likes that explains why he's affected that way. In particular, the gods are high-minded enough not to really care about anything but holiness—to like something because it was, say, popular wouldn't be

---

<sup>51</sup>James (1977, pp. 62f) anticipates part of the process I just described. Of the Kantian transcendental aesthetic and the categories of the understanding, he states:

[S]urely all these were once definite conquests made at historic dates by our ancestors in their attempts to get the chaos of their crude individual experiences into a more shareable and manageable shape. They proved of such sovereign use as *denkmittel* that they are now a part of the very structure of our mind.

<sup>52</sup>Plato (1997).

very *divine*. So the gods like those sacrifices because they're holy. But those holy sacrifices are also holy because the gods like them, and you can't have claims of the form "A because B" and "B because A" *both* be true: that would amount to a vicious circularity. So Euthyphro's understanding of holiness or piety is, as some enormous percentage of the students who take introductory philosophy courses are required to recite, incoherent.

Around the 1980s, it became a standard move in the philosophy world to observe that not all circularity is vicious, and this may be what made Johnston unhappy with the argument I have just reproduced: what's to say that Euthyphro's circle isn't one of the virtuous ones?<sup>53</sup> Plato is onto something, but what is really going wrong with secondary quality accounts of properties like—and let's update the example somewhat—"funny" must be rather different.

"Funny" had better name a secondary quality; nothing is funny from the point of view of the universe. If "funny" names a secondary quality, then its definition must be *something* like this: to be funny is to be the kind of thing that would make people with a sense of humor laugh (or chuckle wryly, or whatever). But surely the fact that something is funny can explain—*causally* explain—why someone (with a sense of humor) laughed. Causal explanations are empirical, a posteriori and contingent. So the link between funniness and laughter is empirical. Definitions are not empirical: they're a priori, necessary, and even analytic. So if being funny is a secondary quality, the link between funniness and laughter is nonempirical (definitional, necessary, a priori). Now, the link between funniness and laughter can't be *both* empirical and nonempirical: if A explains B empirically, B can't explain A definitionally. So "funny" can't name a secondary quality.<sup>54</sup>

The very plausible thought moving Johnston is that the connection between A and B can't be necessary in one direction and contingent in the other. Likewise, it can't be a priori in one direction and a posteriori in the other; and it can't be analytic in one direction and synthetic in the other. After all, it's

---

<sup>53</sup>For allegedly virtuous circularity in this rough topic area, see McDowell (1998, Chapter 7) and Wiggins (1991).

<sup>54</sup>Johnston (1991, 1998). For obvious reasons, Johnston peels apart what we and the tradition call "secondary qualities" into "response-dependent properties" (those conforming to the conceptual requirement that you understand them by way of observers' reactions) and "manifest properties" (those you can actually see).

Hacking (1995, p. 82) provides a real-life example of this form of argument:

[C]hildhood onset and the presence of trauma are not parts of an empirical generalization of a statistically checkable fairly-necessary-condition. They are part of the authors' understanding of multiple personality disorder. . . . There is nothing methodologically or scientifically wrong with this. I warn only against having it both ways. There is a tendency (a) to define the concept "MPD" . . . in terms of early childhood trauma, and (b) to state, as if it were a discovery, that multiple personality is caused . . . by childhood trauma. We should not delude ourselves into thinking that we first defined the disorder and then discovered its cause.

the *very same claim*, looked at from one end and from the other.<sup>55</sup> But however plausible Johnston's thought, it is, as we are now in a position to see, false. Once you realize what the explanation for our *having* necessity is (and, although this is not the burden of the present essay, apriority and analyticity), it is obvious that the apparently paradoxical pattern we are considering should be the normal case. And so it is: It is an empirical fact that (organisms that resemble) human beings do not give birth to (organisms that resemble), for instance, mice, and vice versa; indeed, that your parents were human beings causally explains your being also a human being. Very many facts of this general sort are stable enough for membership in a lineage to have become the central component of species or taxon membership: at least since Aristotle, it has been necessary that humans are descended from other humans.<sup>56</sup> It is an empirical fact, although a very deep one, that we do not split in the manner imagined by Derek Parfit—that is, like amoebae—and so “I tomorrow” necessarily picks out at most one person.<sup>57</sup> That you can't change the past is in the first place a fact of life: no one knows how to go about it (other than, as Nietzsche and Wilde insist, by reinterpreting it); consequently the medievals came to think of what is past as necessary. (And likewise, when Kierkegaard tells us that life is remembered backward but lived forward, that is an analytic truth.) That your mother is female explains—*causally* explains—her giving birth to you; that fact is stable enough for mothers being female parents to

---

<sup>55</sup>I need to anticipate a worry that careful readers might be having, namely, that the contents of the incompatible links don't really match up (and so that perhaps they're not incompatible after all). “Funny” is, by definition, what makes people with a sense of humor laugh: that connection is definitionally tight. But that a funny joke causes laughter is not nearly as tight a connection: maybe a particular audience member didn't laugh at *this* because he swallowed an olive and started hiccupping.

Analytic philosophers are used to the idea that definitions have to be exceptionless and that they can't be *kinda sorta* true; that's why some of us are always trying to give necessary and sufficient conditions for this or that. (For reasons to exempt definitions from the bivalence requirement, see Millgram 2009a, Section 6.3.) Whatever the merits of the requirement, when it is imposed on secondary qualities, the effect is to introduce stipulations into the definition, such as “normal observers,” “in standard conditions,” or, in the low-key example I gave, “people with a sense of humor.” The function of these clauses is to serve as a cancellation coupon for the *ceteris paribus* clause lurking in the background: funny things are what people laugh at, except that sometimes they don't; adding “people with a sense of humor” excludes the people who don't laugh at funny things.

Although my own illustration doesn't emphasize this fact, these *ceteris paribus* clauses can't be converted into lists; they have a logical function, and not a determinate content. That means that a cancellation coupon can't be specified as having a determinate content, either. “Standard observers and normal conditions” only does what it's supposed to when it's understood to mean: override the *ceteris paribus* clause. No one can say what “standard observers” and “normal conditions” are—although many inadvertently amusing journal articles have tried. And once we adjust for the fudge factor in the definitions, we have matching contents.

<sup>56</sup>Thompson (2008, Part I) elicits the version of this necessity that is built into our Aristotelian way of thinking about species; Rowe (1987) discusses a reform of biological terminology that would make it definitional of taxa generally.

<sup>57</sup>Rovane (1990) observes very convincingly that if the biological and social facts were as Parfit imagines, not just our concepts but our first-person indexicals would shift appropriately.

have become analytically and conceptually true. It is an empirical regularity that the earth rotates on its axis 365 times (and a little bit) as it completes one solar orbit; and that is why it is *also* true by definition that a year is made up of 365 days.<sup>58</sup>

It is intellectually efficient to adjust the signage we post around our most robust empirical knowledge so as to encode its being known to be robust; as Whewell put it, “the language which [we have] learnt to use unconsciously, has been adapted to, and moulded on, ascertained truths.”<sup>59</sup> It is intellectually efficient, once we have done so, to treat those very robustly confirmed facts as requiring no observation (no further observation, and we typically delete the supporting observational evidence we formerly possessed); our empirically best-supported conclusions will come out seeming *a priori* to us. And finally, because an attention-conserving strategy is also intellectually efficient, we sometimes adjust our concepts so that they encode especially those connections whose causal tightness has been overwhelmingly empirically confirmed. Johnston’s Missing Explanation Argument runs aground on a missing explanation, namely, of what we have necessity *for*.<sup>60</sup>

## 8.9

I’ve made a series of rapid-fire suggestions about the cognitive function of necessity, and also about its place in a system of related statuses. It may not sound that way, but these suggestions are meant to be tentative; I don’t want to insist that I’ve gotten everything right, and in any case the main illustration was

---

<sup>58</sup>In these examples, the causal or empirical regularity comes first in the order of explanation. But there is a variant of this pattern in which the analytic or conceptual truth serves to explain the matching causal connection. Although it’s true by stipulation that graduating entails having satisfied a long list of requirements, that you satisfied the requirements causally explains your graduation; people ask the rhetorical question, “Is the Pope Catholic?” because it’s analytic that he is, but his being Catholic is part of the empirical explanation of his having become the Pope; as Jerry Fodor reminds us, it’s not a genuine United States dime unless it comes from the US Mint (a necessary, *a priori* and even analytic truth), but the empirical existence of this here dime is due to its having been manufactured by the Mint (and that’s a causal, empirical and contingent fact).

<sup>59</sup>Whewell (1847, vol. i, p. 270).

<sup>60</sup>In his earlier publication, Johnston tried to cash out his motivating thought with an explicit argument (dubbed the “Missing Explanation Argument” at the time), which had roughly the following drift: An *a priori* equivalence should license a substitution of equivalents within the causally explanatory link; the substitution should preserve how causally explanatory the link potentially is. But if you substitute “it’s the kind of thing that makes people laugh” for “funny” in “he laughed because it was funny,” you get the not very explanatory “he laughed because it was the kind of thing that makes people laugh.” That argument generated a small literature, for instance, Menzies and Pettit (1993), Miller (2001), and Haukioja (2006), whose upshot was a gradual retreat from the argument to the motivating thought; the takeaway from that back and forth is that the argument I’ve just recapped was not a satisfactory expression of the thought, which best stands on its own. I’ve been following what looks like an emerging convention by continuing to use “Missing Explanation Argument,” now as a label for that thought.

a drastically simplified and stripped-down *ur*-version of a very rich practice. Again, the real world will be more complicated.

The intent was to provide, not quite a dress rehearsal, but the run-through before the dress rehearsal of how to go about investigating the necessities that form around a disciplinary specialization. Even while we're still just practicing, puzzling phenomena seem to fall into place, one after another, when they're looked at from the cognitive-function perspective. That's a promising contrast with the semantic-analysis approach, which hasn't generated this sort of forward motion for quite some time now.

One of the phenomena that was discussed deserves highlighting. When philosophers start producing credibility-straining ontologies—just for instance, when they produce theories of a mysteriously coercive force, “necessity,” that somehow constrains what can and can't happen, or describe vast structures of abstract objects that seem to be fabricated of an ethereal and impossibly hard crystal—sometimes the secret ingredient is memory loss. Old-school metaphysics is, anyhow often enough, a side effect of forgetting what you were in the middle of doing. Here, what is forgotten is a policy we had adopted, or a series of decisions our ancestors had made, and in one of those category mistakes for the record books, you mistake policy requirements and directions for a fact—a *supernatural* fact.

There may be practical benefits to the approach I am commending, over and above getting exotic philosophers' confusions out of the way. Self-awareness is a great help to doing certain tasks well. Once we recognize that we are engaged in formulating attention management policies, we may do it more successfully. And once we pay more attention to the depth of division of labor in human life and to its flexibility, we may do better at formulating policies that are suited to us, rather than to the nonexistent creatures in the Piltdown Man thought experiment.

The interesting question, it is now clear, is where the signage *should* go. Old-school metaphysics of necessity is in the business of reconstructing the internalized attention-management signage of previous generations. But *who cares* what our forefathers thought didn't need any more attention? What we want to know is what *doesn't* (now, given what we know now, and given the practical demands now being made on us) need further attention. Certainly necessity is a blunt tool, one with costs and unwanted side effects of its own. Nonetheless, from here on out the real question is: What should now be *made* necessary?<sup>61</sup>

---

<sup>61</sup>I'm grateful to Jenann Ismael, Brendan Balcerak Jackson, Lex Newman, Shaun Nichols, Valeria Ottonelli, Emily Rosenzweig, David Schmitz, Aubrey Spivey, Mariam Thalos, and Jonathan Weinberg for conversation, to Chrisoula Andreou, Ben Crowe, Christoph Fehige, Svantje Guinebert, Kim Johnston, Kathrin Koslicki, and John MacFarlane for comments on an earlier draft, and to audiences at the University of the Saarland, Cologne University, the 6th Annual Intermountain Philosophy Conference and the Cogito Research Center of the University of Bologna for valuable feedback.

## Applied Ethics, Moral Skepticism, and Reasons with Expiration Dates

The question, “Why be moral?” has been on philosophers’ agendas at least since Plato, and the antiquity of the Question (which is what I’ll call it from here on out) reasonably prompts the suspicion that moral skepticism may not in the end be rebuttable. Haven’t we waited long enough? Isn’t it time to discard the pretense that we have reason to do what morality demands? And shouldn’t we finally stop wasting our time on it?

Before buying into that conclusion, I want to point out a largely untapped source of reasons for being moral. It has been overlooked, I will suggest, in part because attention has been exclusively directed to providing a *definitive* refutation of moral skepticism. However, that fixation is attributable to a very basic error about what sort of creatures we are. Most of our reasons come with expiration dates; and, indeed, the Question is often answerable, but *temporarily* rather than definitively.

I will suggest that one category of less-than-eternal answers to the Question should receive more attention than it has. Instead of addressing ourselves to the cynical huckster who wants to know what’s in it for *him*, we should be responding to, for example, the businessman who understands loyalizing potential whistleblowers to be his *responsibility*. By way of defusing one source of resistance to my proposal, I will advance—but not properly argue for—two further hypotheses: that a central function of morality has the consequence that we should expect justifications of morality to be less-than-eternal generally, and that the cognitive repertoire available to support some such answers may make them quite short-lived.

I hope to indicate how periodically replenishing our supply of answers to the Question is an activity in which moral philosophers might engage. I will conclude with a suggestion as to which sort of moral philosopher is best equipped for the job. The most concrete, satisfying, and socially important answers are likely to come from the applied ethicists.



## 9.1

If a question is not to be hollow, it must be posed in a way that determines what would count as an answer.<sup>1</sup> When the question is of what someone has reason to do, it will thus invoke standards, priorities, and guidelines. I mean these notions as placeholders, and, so far, all I want to register about them is that, although they are used to assess choice and action, I haven't yet introduced assumptions about their structure and content. In particular, we should not assume that they must be, formally, goals or ends or desires.

*The Question*, accordingly, must be asked so as to pick out, tacitly or explicitly, a system of standards, priorities, and guidelines that fixes what would count as an appropriate answer to it. Call that the Question's *platform*; I am going to assume (disregarding Prichard<sup>2</sup>) that platforms from which we ask the Question are at some suitable distance from morality. If the standards, priorities, and guidelines we invoke *are* just those of morality, then the Question, voiced in a suitably skeptical tone, amounts to wondering whether morality is self-refuting.<sup>3</sup> And while the reflexive platform is a perfectly reasonable special case to take under consideration, here I am going to put that version of the Question to one side.

From what other platforms can the Question be launched? Moral philosophers of our tradition have considered, among others, one's self-interest or one's well-being, the desires or preferences one happens antecedently to have, and forms of practical consistency demanded by a highly purified rationality. But the traditional inventory of philosophers' standards is not an exhaustive list of available platforms. It scarcely needs mention that instead of asking the egoist's question, of why *he* should be moral, one could ask why one's clan, or country, or religion should sacrifice its interests in favor of the demands of morality. It seems to me that a further and important class of platforms has been overlooked. To identify them, let's cut to a panoramic shot of humanity, one that will allow us to see ourselves as one species among many, but nonetheless a species implementing a distinctive adaptive strategy.

Think of species as roughly placed along a spectrum, with specialized species on the one end and weedy species on the other. Specialized species are often celebrated by film in the nature documentary genre: to take an example

---

<sup>1</sup>The observation is meant to recall Hamblin's Dictum (Hamblin, 1958, p. 162); Diamond (1991, Chapter 10) suggests that when this demand is not satisfied, what is being posed may still be a *riddle*—something rather different than a question.

<sup>2</sup>Prichard (1912).

<sup>3</sup>Nietzsche, for one, thought that it was; in the *Genealogy* (1989), he develops the idea that morality "overcomes itself," that is, that once the demands of morality are uniformly enforced, the effect will be to undercut the commitment to morality itself. In particular, when the moral demand for honesty is extended and deepened to include thoroughgoing intellectual honesty, moral agents will have to surrender their justifications for morality. His view is a healthy corrective to the presumption that because morality will of course be self-endorsing, the problem that arises here is vicious circularity.

that screened in the movie theaters a while back, the emperor penguin is built to go for extended periods without eating, to prevent its egg from freezing by balancing it on its toes for months at a time, to live through blizzards in eighty-below temperatures, and to regurgitate food that it has carried, on foot, some eighty miles in its belly . . . in short, it is spectacularly engineered to exploit the resources scantily available in one of the most daunting ecological niches on earth.<sup>4</sup> It is, however, adapted to *only* this niche; the emperor penguin does not survive in the wild anywhere but Antarctica. If the environment changes and the niche disappears, the species will disappear also. Weedy species, like rats and cockroaches and arabadopsis, are less well suited to star in nature documentaries; the engineering is usually not nearly as lapidary. But they are much more durable, which is perhaps why the prediction that these species alone will survive a nuclear war has become entrenched in popular culture.

Human beings are sometimes described as occupying the weedy end of the spectrum, but this is not quite right. In fact, humans are a unique way of handling the tradeoffs represented by this spectrum. Humans occupy both of its ends simultaneously; they are weedy *by* specializing. The trick is to produce the adaptation in software rather than in hardware, and humans manage some of the most extremely specialized adaptations—call them hyperspecialized adaptations—that one will ever observe: arctic nature cinematographer, condensed matter physicist, historian of medieval English property rights, sculling coach . . . just to begin an indefinitely long list. What is more, humans are capable of more than one such hyperspecialized adaptation over the course of a lifetime. Although there are only so many times one can manage this kind of transition, someone may start out as a chemistry major, become a rabbi, later on work as a superintendent of schools, and, still later in life, reinvent himself as an author of books about the Bible. To have a short way of saying this, I'm going to adopt a slogan: *Humans are serial hyperspecializers*.<sup>5</sup>

We can think of the weedy-by-specializing creatures we just described (ourselves, but it's helpful to keep the external perspective) as occupying niches that are roughly analogous to the niches occupied by specialized species in an ecology. These niches are usually social. The shape of the niche of a nonhuman organism may be determined by its predators, its prey, its competition for that prey, and so on; all of these are likely to be other species. The shape of a serial hyperspecializer's niche may be similarly determined by the occupiers of adjacent niches, but in a dense enough population these will be, for the most

---

<sup>4</sup>Jacquet (2005).

<sup>5</sup>See Chapter 4 in the present book for further discussion. Although plasticity has come in for much attention in recent evolutionary biology, the slogan I have just introduced is not meant as a claim about our evolutionary history. While I mean my claims to be compatible with current science, the point is not that hyperspecializing strategies are the products of past selection pressures, but rather that, however we got this way, it is how we are suited to function *now*. Relatedly, neither do I mean to suggest that all human societies allow for full or even partial exercise of this capacity.

part, other, differently specialized serial hyperspecializers. We can already anticipate that serial hyperspecializers will collectively have a use for a morality: if their interrelated specializations mimic an ecology and if predator–prey relations are normal structural elements of an ecology, a host of issues are likely to arise that will provide occasions for management and amelioration.

Because adaptation to such a niche is handled in the software, when serial hyperspecializers enter one, they must identify and adopt a system of standards that will allow them to navigate it successfully. (If the niche is being created from scratch, that means generating a novel system of standards.) Niches can differ arbitrarily from one another, and consequently the appropriate systems of standards can also differ arbitrarily from one another. However, because niches are relatively constrained environments (an environment that wasn't wouldn't count as a niche), an effective system of standards for a niche will exploit the regularities that the niche provides. For this reason, such standards can often be formulated as fairly precise claims about, for instance, what performances or outcomes should or ought not to occur. Let's use that observation to tighten up just slightly the terminology I introduced earlier: from here on out I'll use the term "standards" for the niche-specific assessment and guidance devices and will leave "guidelines" and "priorities" for devices that may perhaps work within a niche, but are also, or instead, suited to broader and more open-ended problem spaces.

Academia is itself a microcosm of the phenomenon to which I am pointing, and so let me take the academic example that is perhaps closest to home. Philosophers internalize elaborate and often quite subtle standards for assessing philosophical work. For instance, they are taught to be extremely sensitive to vicious circularity; here the standard is that arguments shouldn't beg the question. That is not by any means a universally accepted requirement; outside of philosophy, it is so little imposed that the very phrase has been reappropriated. (In contemporary idiom it has come to mean, "to invite a further question.") Or again for instance, philosophers learn that the successful interpretation of a text must demonstrate competing interpretations to be in error; this is a demand with a great deal less grip in some of the other fields that engage in textual interpretation, such as literary studies. Or again, a philosopher's arguments are expected to be mutually consistent; contrast this with adversarial legal systems in which (I acknowledge that this is going to be a bit of a caricature) it's fine to argue that your client never so much as saw the vase he is alleged to have stolen, that the vase was given to him, and anyway that it was already cracked when he took it.<sup>6</sup> Or again, philosophical prose, in the

---

<sup>6</sup>The justification for the practice, as I understand it, is that when the judge does not so much supervise an investigation as determine the merits of arguments presented to him by opposing sides, clients are ill served by an attorney who fails to present arguments that the judge might find compelling. (I'm grateful to Alice Clapman for discussion on this point and for the illustration.)

analytic tradition at any rate, is expected to work very hard to leave no room for alternative interpretations of its meaning. Indeed, it is typical to broach possible misinterpretations of one's view explicitly and to warn one's readers off them. Contrast this with theater, where the standards are rather different. A good script leaves room for interpretation on the part of the actors; it doesn't specify what the characters are thinking or feeling; if it doesn't provide enough in the way of blank space for the actors, then, one says, it has too much author in it.<sup>7</sup>

There is of course much more to a disciplinary system of standards than the standards that are used to grade the product. Continuing with the example on hand, professional philosophers for the most part buy into an internal disciplinary hierarchy (metaphysics is more respected than ethics—*that* hierarchy); the pecking order of subspecialties strikes outsiders as deeply counterintuitive. Recently, philosophers have come to internalize a ranking of American philosophy departments posted on a suddenly influential web site. They internalize a schedule of injuries—for instance, those having to do with intellectual property.<sup>8</sup> And this is just a sampling of the mostly tacit code.

Serial hyperspecializers exhibit what seem on the face of it to be aptitudes pulling in opposite directions. On the one hand, a serial *hyperspecializer* will be cognitively equipped to internalize the system of standards that governs activity in a niche it has come to occupy. But a *serial* hyperspecializer will also be cognitively equipped to detach itself from one niche and go looking for a different and better one. The prompts for doing so don't look much like niche-specific standards. Paradigmatically, negative spikes in hedonic tone, as well as reactions of interest and boredom, can trigger a search for a more rewarding specialization. These are *feelings*: thus, formally quite different from claims about what ought to be done. The affective cognitive repertoire is suitable for navigating less structured environments than those provided by social niches.<sup>9</sup>

## 9.2

Niche-specific systems of standards frequently enough amount to platforms for the Question, Why be moral? This means in the first place that the Question may invoke a great many rather different standards. If questions are given their content by the standards that fix what would count as a satisfactory answer to them, that is to say that the Question is in one important respect systematically ambiguous. Here's another close-to-home illustration:

---

<sup>7</sup>I'm grateful to Carla Bagnoli for serving as a native informant.

<sup>8</sup>I'm grateful to Candace Vogler for alerting me to the central role in these standards of definitions of injuries.

<sup>9</sup>For discussion of the cognitive equipment, see Millgram (2005a, Chapter 1 and Section 5.7).

Philosophers prize tightness and, often, a sort of exquisite delicacy in argumentation (and they prize argument almost above all else); typically they like arguments on one philosophical question fine-tuned to accommodate developments in other philosophical debates. So when *they* ask the Question, they want a tightly argued answer, and preferably an exquisitely and delicately argued answer, one that is properly fine-tuned. But politicians prize robustness, which is a different matter entirely; a politically satisfactory platform is one around which a stable and effective coalition can be built.<sup>10</sup> If an answer to the Question pleases philosophers, by virtue of being exquisite, delicate, tight, and sensitive to the outcomes of debates in various philosophical subspecialties, it is unsuitable by the lights of the politicians.

It evidently follows from the redescription of our species which we have just given that there is not just one answer to the Question, Why be moral? Instead, there are indefinitely many of them. To a certain kind of old-fashioned military man, addressing the Question might mean showing that morality is honorable rather than quixotic. To a contemporary business practitioner, it might mean showing that morality makes sense as good business practice and that doing the morally right thing does not amount to fiduciary irresponsibility. To a politician, it might mean showing (by way of, as I suggested a moment ago, a suitably robust set of considerations) that refusing to cross the moral red lines would not ineptly and naively betray the trust of his constituents.<sup>11</sup> Note that none of these versions of the Question amount to asking, “What’s in it for *me*?” Their platforms are systems of professional standards, and the worries they raise can seem pressing even once crass egoism has been replaced by the thought that it is simply irresponsible or inappropriate to make the *moral* choice.<sup>12</sup>

---

<sup>10</sup>The force of the requirement is exhibited in Bernard Williams’s venture into applied ethics. The *Williams Report* (Williams, 1981b) constructs its argument around the Harm Principle, not because there is some philosophically satisfying argument for it, but because it was the only criterion on which the witnesses interviewed by the commission seemed to agree. For further discussion of the problems posed by this requirement, see Millgram (2005a, Chapter 8).

<sup>11</sup>Here’s the sort of thing I mean. Historically inclined Americans will recall the anti-Communist posturing in which Richard Nixon had to engage in order eventually to be in a position to recognize what was then called “Red China.” Politicians are in some circumstances *expected* to lie in just this way about what they are going to do. The sort of concern to be alleviated is that a politician who refrained from lying would be effective neither in advancing the interests of his constituents, nor in reshaping those interests; yet these are among the most important parts of a politician’s job. To be sure, there is an air of paradox in the thought that not lying to one’s constituents is to betray their trust; nonetheless, it would hasty to dismiss even this instance of the so-called problem of dirty hands on that basis.

<sup>12</sup>In putting it this way, I mean to be neutral as to whether such an agent takes himself to have reasons, not derived from his professional niche, for being moral (so that he is worrying about whether they are perhaps trumped by niche-based reasons to disregard moral requirements), or whether he is considering reasons for morality that are directly niche-derived. For an interesting example of a mixed stance, see Munger (2007, p. 80), describing the investment firm that he and Warren Buffett manage: “this place . . . tries harder than most places to be ethical . . . [D]espite the presence of some human failing, we’ve had an amazingly low amount of litigation or scandal or anything of that sort over . . .

Notice, now, that this type of reason for morality cannot be expected to last forever. Serial hyperspecializers implement a strategy that sometimes takes them from one niche to another. Such a creature may have had the Question satisfactorily answered, in the terms of a niche-bound system of standards which it accepts. After moving to a different niche, it may find itself lacking reasons to be moral anchored in the system of standards adopted in the course of the move. When the officer trades his uniform in for civilian politics, his military code becomes irrelevant to his new circumstances; such systems of standards can be discarded surprisingly rapidly, and with them, justifications for being moral which they had served to anchor. It may not be at all apparent to the former officer what the newly internalized standards of a career politician or senior civil servant have to say about allegiance to the moral law. In such a case, he is now a potential client for a new justification of morality.

The phenomenon of temporary rather than definitive niche-derived justifications for morality does not arise solely from serial hyperspecializers traversing extant niches. The strategy of serial hyperspecialization not infrequently leads to the emergence of novel niches within the larger human ecology, and within these, of novel systems of standards.<sup>13</sup> Even in the unlikely event that a niche-bound justification of morality is one day produced for the occupants of every existing specialization, that state of affairs would prove only temporary. As new niches appear, new platforms from which to raise the Question, and so new forms of moral skepticism, come into play. And as old platforms are abandoned, old answers, no matter how formerly compelling, turn moot.

### 9.3

At this point, readers are likely to be worried that the subject has been changed on them. When philosophers worried about moral skepticism, they were not asking whether professionals of one or another sort would find professional reasons to be moral, but why *you* should be moral, *whoever* you are. It is implausible that all niches will themselves provide the materials for an answer to that Question: what could there be in the craft standards of a burgler or hit man to anchor an answer to it?<sup>14</sup> Shouldn't the account that philosophers are

---

decades. And people notice that. . . . I don't think we deserve a lot of credit for that because we early understood that we'd make more money that way. . . . I'd like to believe that we'd all behave well even if it didn't work so well financially. And every once in a while, we get an opportunity to behave that way. But more often we've made extra money out of morality. Ben Franklin was right for us. He didn't say honesty was the best morals, he said it was the best policy."

<sup>13</sup>For a historical example, see Duden (1991), which contrasts the standards of the barber-surgeon or the town *physicus* of eighteenth-century Europe with those shared by today's medical specializations.

<sup>14</sup>There have, however, been philosophers, most recently of a Kantian bent, who thought that individual systems of standards, whatever they are, must carry commitments enabling a uniform answer to the Question. Korsgaard (1996b) is a well-known example.

seeking rather give you a compelling reason not to become a burglar or hit man in the first place? And shouldn't a *compelling* reason not come stamped with an expiration date?

My objective here is not to argue that old-school assaults on moral skepticism did not have a point, and it is not to argue that the niche-bound answers will do everything the old-school moral theorists wanted. My interest is rather in redirecting some of our energy and attention to another class of answers which we might attempt, one that has been hitherto almost entirely ignored. I am, however, going to take some time to discuss one of the potentially objectionable features of niche-based answers to the Question. There are reasons for expecting answers to the Question to be temporary rather than definitive, even when they are not niche-bound.<sup>15</sup> So one of the resistance-provoking features of niche-bound reasons is likely to be shared by much of their competition.

I started discussing the justification of morality without explaining what was meant by "morality" in the first place. Now, morality has proved hard to define in a way that does not beg important questions. Accordingly, instead of offering a definition, I advance the following remarks about its function, which will I hope will prove uncontroversial.

Morality regulates our interactions with one another; that is perhaps not the only task it has, but is certainly central. It does not, however, regulate every aspect of each of our mutual interactions; we look to it in the first place for guidance in circumstances that we might face as, to put it a bit baldly, generic human beings. The Decalogue, for instance, prohibits murder and adultery, demands that we honor our parents, and so on: these guidelines address choices that just about anyone might be in a position to make. They contrast with matters of disciplinary practice and expertise that are much less often—except where they encounter considerations that might be invoked in generic contexts—thought of as moral matters. Without being very precise about it, I will say that morality regulates generic social interaction (and to

---

<sup>15</sup>Once we have the sketch of humanity as serial hyperspecializers on hand, and are alert to the likely effects of a strategy of serial hyperspecialization, a survey of the more traditional sorts of platform for answering the Question ought to make one think twice about just how much in the way of truly niche-independent platforms we have. Traditional platforms may not be *alternatives* to niche-based platforms, just because serial hyperspecializers are cognitively equipped to internalize the standards of the niches they occupy. Take the sample platforms to which I alluded earlier: Professionals typically desire success in their profession, and there are many other ways goals and ends are shaped by a serial hyperspecializer's niche; they are thus not normally independent of the niche one occupies. One's well-being is likewise shaped by the niche one occupies: doing well at what one does is a large part of doing well. (This indicates how keeping an eye out for consequences of the serial hyperspecialization strategy can give content to an otherwise amorphous concept; on the other hand, that cannot be all there is to well-being, because a serial hyperspecializer can do well by switching niches.) And I gestured at reasons for thinking that not all of rationality is, *pace* Kant, niche-independent: recall how standards for successful argument varied by field.

reiterate, this is meant to pick out one of perhaps several functions of morality, and not as a definition of it).<sup>16</sup>

If that *is* what morality is in the business of doing (and here I am coming to the reasons my hypotheses are tentative, and supported by less than a tight argument), we should not assume that we know what the content of a satisfactory morality is, even to a first approximation. For if human beings are, as I have been suggesting, serial hyperspecializers, the regulations that govern our mutual but generic interactions have been all-too-evidently composed without an eye on some very important features of the lives that are being regulated.

We can identify the oversight in as early a text as Plato's *Republic*. In setting up the problem he will address, division of labor comes on the scene very early. But he implicitly assumes that the roles are few, fairly simple, and, incidentally, relatively fixed. Plato portrays his characters as able to list the professional specializations that a city will require (so he is not imagining that novel crafts will emerge routinely in the normal course of events in a well-run society). He seems to think that, for the most part, once a carpenter, always a carpenter (and it's better that way); thus there is no need for an extensive investigation of the aspects of the human soul that enable carpenters to become computer programmers, or, for that matter, members of SWAT teams. Still more significantly, he underestimates the transparency of the systems of standards that govern activity within niches. This is exhibited not in his direct pronouncements (rather, a symptom of how seriously he underestimates it is that the topic does not even come up), but in his solution to the problems of social regulation: an expert, the infamous philosopher-king, qualified by his studies in a quasi-mathematical discipline that is perhaps our earliest version of metaphysics, will write and administer the rules.

---

<sup>16</sup>This may explain why we are so often resistant to the notion of moral expertise: moral guidance is picked out in part *as* guidance of which we can avail ourselves where expertise is beside the point. However, traditions in moral philosophy differ on this point. And codes of professional ethics attempt to extend the moral mode of regulation into areas of disciplinary practice; where they do so, moral expertise is much more in place. A hospital's hiring a bioethicist does not prompt nearly the puzzlement that appeals to moral expertise in nonprofessional contexts would.

Urmson (1968, pp. 83–85) presents a consideration that can be taken as both an argument that we don't treat others as authorities in matters of morals and a quick and dirty test for whether we are willing to treat people as experts and authorities in some region of applied ethics. He observes that questions prefaced with "Is it true that . . . ?" are normally addressed to authorities and that prefacing questions on moral matters with this formula sounds distinctly odd. ("Is it true that killing is wrong?") However, notice how much less odd it sounds when the subject is applied ethics: "Is it true that medical providers should obtain informed consent whenever possible?"

(A caveat: old-school applied ethics lives up to its name by taking its task to be that of applying one or another niche-independent moral theory to a problem posed by, for example, medical practice. The commitment to informed consent, in the example I just gave, is derived from views about the importance of autonomy, not in medicine but in general. However, I certainly don't mean to endorse that way of doing it; better work in applied ethics rarely looks like this.)



Plato, like most moral philosophers since, seems not to notice that the systems of standards regulating specialist activities cannot normally be understood by outsiders. If you are a philosopher, you have a grasp of the disciplinary standards of philosophers, and with a great deal of effort, you might have a decent grasp of one or two more such systems of standards. But nobody—not Plato’s philosopher-king and not anybody else—is in a position to understand all those drastically different niche-specific concerns and activities well enough to write a decent set of rules for them. Academics experience their small-scale versions of the problem often enough. How many times have you been on some interdisciplinary committee charged with allocating resources and realized that there is no one in the room capable of assessing the comparative merits of the competing proposals? How many times have you complained of a rule written by some university administrator, that he did not understand its impact on *your* department? Because no one could have the disciplinary insider’s knowledge of so many disciplines, no one could competently perform the job of Plato’s philosopher-king.

The reason for tentativeness, then, is this. One glaring problem that we face in regulating our generic interactions is that, because so many of us are so specialized, and because specializations induce standards that are not comprehensible from outside the niche they serve, we are scarcely ever in a position to understand the impacts of our decisions on others. A central job of morality is regulating our generic interactions. So we should expect that morality will do its job only if it has been designed with a great deal of attention to that problem. That attention has been conspicuously absent. And that is why we should not assume that we know what the content of an adequate morality would be.

Moral philosophers too easily adopt the stance that, when it comes to morality, there is rough agreement on matters of content, and that may be true enough, anyway as far as the philosophers themselves go. But we should not assume that this sort of agreement is telling us more than: they are, all of them, making pretty much the same mistake.

If that is correct, making claims about structural features of plausible justifications of morality is an especially tricky business. What would count as a decent justification of morality will presumably depend on the substance and content of what is being justified, in something like the way that what would count as the justification of a scientific theory is going to depend on the specific content of that theory. We should not be confident that we know what the content of morality is. Therefore, we should not be confident that we know what would count as a decent justification of morality.

With this difficulty fully acknowledged, however, we now entertain two sorts of consideration that make plausible the hypothesis that even our niche-independent reasons for being moral will be temporary rather than definitive.

Consider an analogous problem, that of designing standards and controls to regulate financial interactions. New roles periodically appear in the financial

world: recently, hedge fund manager, structurer of collateralized mortgage obligations, ratings agency analyst. We have been all-too-painfully reminded that, as this happens, the standards and controls that govern financial institutions and transactions have to be rethought and reimplemented.

I claimed that a central function of morality is regulating the generic interactions of (as it turns out) differently specialized persons, occupying different social niches. Niches come and go, and moreover, niches undergo internal changes, *inter alia*, to the systems of standards that govern activities within them. That means that the regulatory problem to which morality is a solution changes over time, in something like the way that the regulatory problem which prompted the Glass–Steagall, Sarbanes–Oxley, and Dodd–Frank Acts changes over time. Whatever the content of an adequate morality turns out to be, we can expect it to be so only temporarily; as the human ecology reconfigures itself, it will need to be rethought and renegotiated.

Nonetheless, notice that because morality solves a regulation problem which spans niches, its content at any one time and place will seem relatively independent of the reasons for conforming available from within a niche.<sup>17</sup> This means that, when it is raised in this way, the Question will appear to the agent to be about adhering to a given code, rather than about the content of a satisfactory morality, even though that content is likely, in the larger scheme of things, to be up for renegotiation. Compare the deliberations of a trucker, wondering whether he should obey the traffic laws: he is not asking himself what the traffic laws should be, or how they might be improved (although of course they could and may well be), but rather, whether he will conform to the laws as he is presently faced with them.

To recap: What justifications we provide for morality will depend on the content of morality. The content of morality, we have just suggested, will change from time to time. We should not expect the justifications to outlast those changes; as serial hyperspecializers invent new niches and abandon old ones, we will, for this additional reason, need new reasons for them to be moral. If all that is correct, then the Question is a moving target.

Let's now take up a second tentative reason to think that even niche-independent platforms for answering the "Why be moral?" question provide temporary rather than definitive answers to it. Once again, I have not argued that niche-based reasons for morality are the only ones there are, and as the career of Eichmann reminds us, the inability to think outside the box is a deep defect in a human being.<sup>18</sup>

---

<sup>17</sup>That said, it does need to be qualified. MacIntyre (2006, pp. 183f; see also 198f) describes a case in which there is less than meets the eye to the apparently shared sign-on to a set of generically coordinating priorities. Death is agreed to be bad and to be prevented when possible and within reason, but *how* bad (and how it is bad) varies startlingly from one set of niche-specific guidelines to another.

<sup>18</sup>For some discussion of what the defect amounts to, see MacIntyre (2006, Chapter 11).

When we are thinking outside the box rather than inside a niche, we deploy a different and distinctive cognitive suite. In particular, I suggested earlier that we rely much more heavily on affective indicators, and I adduced as examples two sorts of paired responses, pleasure/displeasure and boredom/interest. Put aside the worry that novelists and filmmakers are better than professional philosophers at articulating the reasons that turn on such affective responses. Instead, notice that these responses extinguish rapidly: like jokes that are funny only the first few times, the formerly interesting idea quickly becomes boring; the delight in some recent improvement fades just as quickly.<sup>19</sup> And this means that arguments for morality built around interest, delight, and similar responses will serve only for a time.

Perhaps committing a new sort of sin makes you feel awful, and perhaps we can understand your feelings as telling you not to do it. Even if we could count those feelings as an argument, or build an argument around them, feelings fade; as recent history repeatedly shows, people get used to even the most ghastly practices almost before they know it. If I am right about how serial hyperspecializers think when they are not niche bound, niche-independent justifications for morality will draw on our affective cognitive repertoire. Extinction is characteristic of many of the responses in that repertoire. And so niche-independent justifications for morality will often have a short shelf life: even if they are compelling answers to the Question while they last, we ought to take very seriously the possibility that they will be evanescent rather than once-and-for-all answers.

## 9.4

If a language has the job of facilitating communication between everyone in a society, everybody has to be able to speak it. And analogously, if morality has the job of regulating our generic interactions, it is natural to think of it as common property: how could we get it to do that job, if everybody didn't know what it was they were being told to do?<sup>20</sup> The thought that morality is common property seems to have suggested to moral philosophers the idea that the justification for morality must be common property as well: how could we get morality to do its job, if everybody didn't understand why it was they had reason to comply?

---

<sup>19</sup>For discussion, see Millgram (2004a). In Millgram (1999) I contrasted evanescent affective responses with the stable reasons typical of morality; I now worry that I was too hasty in assuming without argument that moral reasons must be all *that* stable. There is also a tricky worry that I do not want to resolve now: when affective responses shift, do we want to say that the reasons have shifted, or rather that our sensitivity to the reasons has changed?

<sup>20</sup>This sort of approach to the topic may be the source of two related and not uncommon presumptions that I earlier mentioned in passing: that morality has in the first place to do with managing interactions of generic agents rather than specialists, and that there is no such thing as moral expertise.

The linguistic analogy suggests that this move is too fast. It is not just that we can learn to speak languages without having available justifications for learning one language rather than another, or for learning any language at all. Not everyone need have the same justification for speaking some language or other: classicists have reasons to learn Greek and Latin that are shared by no one else; my reasons for using English are quite different from my reasons for using Hebrew. Justifications for deploying a language need not be common property, and when they are niche-based, they are not. Why should justifications for morality be common property? When their platforms are professional specializations, they will not be.

We have seen a couple of reasons to suppose that niche-based reasons for morality need not do worse, as far as their impermanence goes, than out-of-niche reasons. If we are indeed serial hyperspecializers, we should expect niche-based platforms to supply a good deal in the way of raw materials for constructing interesting answers to the Question. And perhaps we will do much better with those materials than with what we have available if we confine ourselves to premises that make up humanity's common denominator. Because the design solution we implement turns on extreme differentiation, and on switching from hyperspecialization to hyperspecialization, it is likely that there will not be all that much in the common denominator, over and above the cognitive equipment supporting the strategy. (A qualification: perhaps hardwiring having to do with the basic functioning of the organism—sleeping, eating, and reproduction—is too critical to leave to the self-modifying software and will turn out to be shared and stable.)

Niche-based answers to the Question will be almost as varied as the niches themselves and can be expected to require frequent upgrades. Who among us is best suited to take on this unfamiliar version of an overly familiar enterprise? That really is to ask which philosophers have the most control of arguments conducted on the basis of the standards internal to one or another highly specialized niche. Once the query is put that way, the answer to it is straightforward enough, and we arrive at a conclusion with which I myself appropriately take leave of the subject we are discussing. The justification of morality is a discouraging topic for metaethicists, and mainstream moral philosophers have not done all that well with it. It is time to see what the bioethicists and the business ethicists, philosophers working in military ethics, legal ethics, and so on—in a phrase, applied ethics and professional ethics—can do with the question of why we should be moral.<sup>21</sup>

---

<sup>21</sup> I'd like to thank Nate Rockwood for prompting me to take up one topic of this paper, and Pepe Chang for getting me to think about the other. I'm grateful to Jonathan Dancy, Zohar Geva, and Alison Hills for helpful conversation, to Leslie Francis, Sam Black, and the latter's students for comments, and to All Souls College for fellowship support.

## Segmented Agency

A great many arguments in contemporary analytic moral philosophy turn on agency, and appeals to one view or another of what an agent is are used to defend positions regarding, among other topics, practical reasoning and substantive moral theory. One indicator of how heavy the traffic has been is the inevitable backlash, with a debate emerging as to whether anything can really be accomplished by invoking agency at all.<sup>1</sup> Here I want to take a somewhat different and perhaps more constructive tack. I am going to argue that the more-or-less shared conception of agency in play is mistaken, but the point of doing so will be to ultimately arrive at a corrected conception of agency, one which can be put to better philosophical use.

The mainstream view in moral psychology is that agents are unified, both in fact (for the most part) and ideally (disunity of agency is regarded as a defect). Accounts of agential unity vary, and include as components such theses as: agents produce actions which can be in a very robust sense attributed to them, actions they *own* because the agents are identified with, rather than alienated from, their choices; agents have “practical identities” or “ground projects” which they may lose, but cannot disown; agents do not pursue projects at cross-purposes with one another; having made a decision, they follow through on it (and do not instead act on some contrary impulse); they possess a unified point of view from which they render judgments about what is worth doing and what they will do; they reflect on their actions, and endorse their choices when they do; when they act, they act so as to understand what they are doing and why; their choices are governed by policies which dictate how competing reasons will be taken into account.<sup>2</sup> Across these variants, we find a rough but

---

<sup>1</sup> Millgram (2005b), Enoch (2006), Ferrero (2009), and Tubert (2010).

<sup>2</sup> See, for instance, Williams (1981a, ch. 2), Velleman (2000), Korsgaard (2009), Korsgaard (2008, Chapters 1 and 3), Korsgaard (1996a, Chapter 13), Korsgaard (1996b), and Frankfurt (1988).

shared picture, of a creature that has integrated its goals, evaluative judgments and other guidelines into a single and internally consistent pattern, and whose control structures generate actions that are consistent with the pattern.

In order to avoid getting lost in the crosstalk, I am going to organize this discussion around just one of these views, a systematic account of agency developed in recent work by Michael Bratman.<sup>3</sup> There are a number of reasons for the choice. First of all, the materials that Bratman uses in his theoretical constructions (plans and policies, and I will explain just what these are shortly) are as clear and straightforward as one gets in this business, but also much more pliable, much less brittle material than competitors' devices.<sup>4</sup> Second, Bratman's treatment incorporates, reconstructs and criticizes other people's ideas—there's no Not Invented Here Syndrome, a rare virtue in philosophy—and engages other philosophers with opposing views (rather than, as so often happens, pretending they don't exist). Third, Bratman's style is consistently low-key and sane-sounding; this is particularly evident in his adaptations of other philosophers' ideas.<sup>5</sup> Consequently, Bratman's account of agency ends up being a best presentation of the central ideas in the current debate, and that means in turn that if we can see how they don't work out in his rendering of them, we will be well-placed to see how (as I am going to suggest) they *really* don't work out.

Unified agency has standardly been contrasted with fragmented agency, and that contrast has made unity seem nonoptional; if you're a highly fragmented agent, you have earned, as a comedian once put it, "the demeaning epithets that are said about people who are peeling an empty banana": the lights are on but nobody's home; you're not all there; you're not playing with a full deck; you're leading (and here is one from Bratman) a "seriously fractured life."<sup>6</sup> I will be introducing another and much more viable option: human agency is *segmented*.

I will begin by laying out Bratman's account of agency and highlighting a number of its more important features. Then I will sketch a strategy, *serial*

---

<sup>3</sup>Bratman (2007).

<sup>4</sup>For Bratman's materials, see Bratman (2001), Bratman (1987), and Bratman (1999). To give you a sense of the contrasts I have in mind, Korsgaardian constitutions are an example of a less straightforward theoretical construct (Korsgaard, 2009; Korsgaard, 2008, Chapter 3), and Frankfurtian hierarchies of higher-order desires are an example of a clear but brittle construction material (Frankfurt, 1988, Chapter 2).

<sup>5</sup>For instance, when Velleman (2000) goes looking for a mental state that can't be disavowed—and in a moment we'll gloss that phrase as well—he settles on a state and a matching theory that strikes everybody who encounters it as (sorry, David) *wacko*. When Bratman adopts the idea that we need to identify a psychological state, or complex of them, that isn't disavowed, he comes up with the sort of policies we will be taking up in a moment; the move seems thoughtful, plausible, and anything but wacko.

Notice the adjustment from Velleman's "can't be" to Bratman's "isn't," marked at p. 188 in Bratman (2007, especially n. 5); however, as we will see in due course, there is a sense in which one cannot in fact disown the policies which Bratman is introducing.

<sup>6</sup>Bratman (2007, p. 298); Martin (1998).

*hyperspecialization*, that human beings, as a species, adopt early and often, and indicate how it fits badly with Bratman's account. I will argue that, when it comes to what matters and what your reasons are, if you're ever *really* wrong, wrong all the way down, that's not—on the mainstream way of thinking about these things—something that you can properly face up to: that is, face up to by deciding on a new and different way of handling situations you are in. Finally, I will return to the idea that a serial hyperspecializer normally lives its life in segments: this means, I will suggest, that we are not only missing half of our theory of agency, but that we have been looking in the wrong place for a theory of personal identity.

## 10.1

Bratman's ambition is to use an uncontroversial and unproblematic philosophical toolkit to reconstruct and thus to explicate a handful of central notions and distinctions in recent moral philosophy. I'll first briefly rehearse his agenda; then I'll lay out the toolkit; then I'll describe the construction. Last, I'll redescribe the view of agency which Bratman exemplifies at the higher resolution which his reconstruction permits.

First, a good deal of effort has been devoted, over the last four decades, to making sense of, as we might call it, the superlative attribution of both actions and attitudes to an agent. You blurt out an offensive remark, but insist, later on, that you didn't really mean it; the attitude it expressed was *there*—as Bratman sometimes puts it, it was a wiggle in the psychic stew—and in some minimal sense it was your opinion: who *else's* would it have been? But it wasn't a view you endorse, and you can legitimately disown it. Some very strongly felt desires are for things that you are quite clear you don't really want: an apparent oxymoron, accommodated by the distinction between *mere* wanting, and *superlative* ("real") wanting. Some would-be beliefs and desires *are* full-fledgedly yours, and that you on reflection endorse them is evidently either strong evidence for the superlative attribution or (perhaps in part) constitutes the attitude's "agential authority." Some actions are things you really did—as opposed to others which, although in some sense you performed them, just kind of happened, as you sheepishly tell others afterwards.<sup>7</sup> Bratman's first agenda item is to reconstruct this distinction, between "really yours" and "merely yours"—between attitudes and actions with which you identify and those you can reasonably disavow.

Second, moral and political philosophers have been, again over about the last half-century, very much interested in autonomy.<sup>8</sup> So many philosophical

---

<sup>7</sup> For an overview of some of the attempts to distinguish between full-fledged actions and less-than-actions that count only as "mere activity," see Millgram (2005b).

<sup>8</sup> See, for example, Hill (1991, Chapter 4), Dworkin (1988), and Christman (1989); Fleischacker (1999) describes its topic as "liberty," but can plausibly be taken as an account of autonomy as well.

demands have been made of this concept that it is perhaps unreasonable to hope to provide an account that covers everything that anyone has insisted must be part and parcel of autonomy, but self-government is, by anyone's lights, a central aspect of it and, Bratman thinks, the aspect with which we should be most concerned. So he hopes to provide a model of a self-governing agent.

Third, Bratman wants to be able to say what it is to value something. Partly the point of doing so is to make room for an observation about superlative attribution that we now think of as Watsonian: that identification is evaluative.<sup>9</sup> And partly (I am guessing) Bratman is tempted by the prospect of headway on longstanding open questions in metaethics.

In developing his account, Bratman adopts one widely shared constraint and drops another; I need to pause to explain what these are. First, philosophers in this field experience pressures toward imagining people as ghosts who float above or stand behind themselves, and who intervene from time to time in the workings of their own minds. For instance, in the literature triggered by Harry Frankfurt's early paper on personhood, what makes my desire superlatively mine is that it is endorsed by a second-order desire.<sup>10</sup> But what makes that second-order desire superlatively mine, rather than merely another wiggle in the psychic stew? A further desire? Won't *I* at some point have to decide that one of the desires in this regress *is* really mine? And when I admit to doing that, am I not representing myself as outside the natural order (taken here to include psychological states): something over and above, or concealed behind, the complex of psychological facts? In the face of this sort of pressure, Bratman is taking on board the requirement (which he sees to be correctly posed by Velleman) that his account not invoke, over and above the psychological structures that it adduces, the further presence of *the agent himself*.<sup>11</sup> The agent will turn out to be a psychological structure of a specified sort, one which will leave no such leftovers or loose ends.

Second, metaphysicians of agency typically insist that their theory of agency tells you what agency *is*, the implication being that there could be no properly so-called agents who do not conform to the theory. Bratman is attempting, more modestly, to provide *a* model of agency that exhibits features of interest, such as self-government. Although we are not shown alternatives, for all we know, there might be other psychic structures that amounted to, say, forms of autonomous agency supporting a distinction between really and merely wanting something.<sup>12</sup>

---

<sup>9</sup>Watson (2004).

<sup>10</sup>That is, a desire whose object is another desire: the idea is that I don't just want it, I *want* to want it—or, more carefully, I want to *act* on my wanting it. See Frankfurt (1988, Chapter 2).

<sup>11</sup>Bratman (2007, pp. 24, 196); as he also puts it on occasion, the account must be “nonhomuncular” (p. 187; see also p. 177).

<sup>12</sup>See, e.g., *ibid.*, pp. 163 (and especially n. 5), 183, 197, and 199.



In earlier work, Bratman introduced plans as his preferred philosophical medium or rendering tool. *Plans* (equivalently, intentions) are stable controllers of conduct; that is, once you have adopted a plan, you need a special reason to reconsider and discard it, and if you don't reconsider, you can be expected to go ahead and do what it says. A *policy* is an open-ended plan, one which specifies that in cases answering to such and such a description, you will do so and so. *Self-governing* policies are higher-order policies; that is, they coordinate and manage lower-order, vanilla plans and policies, along with motivating attitudes such as desires and emotions, giving green lights to some of them and red lights to others. *Effective* policies bring it about that choices and deliberation comply with them: when a policy is effective, the fit between policy and action is not merely fortuitous. A policy is *weight-bestowing* when it gives a prospective reason a weight in deliberation, and, more generally, a higher-order policy is *reason-determining* when it determines that a first-order attitude (normally, a desire) play a specific role (defaultly, an end-setting role) in deliberation.<sup>13</sup> (An example, to get the general idea across<sup>14</sup>: a recruitment committee might decide that, in the course of deliberating about whom to hire, the department will give an applicant's area a great deal of weight, will treat affirmative action considerations as a tie-breaker, and will not treat applicants' rank as a consideration at all.) You are *transparent* when the cognitive functionality of your policies is captured in their content.<sup>15</sup> Your self-governing policies are *reflexive* when they endorse their own effectiveness.<sup>16</sup> (Again to give the general idea, if a department adopts the policy that

---

<sup>13</sup>Weight-bestowness is adapted from Nozick (1981, Chapter 4); Bratman tends to use longer locutions to mark such policies (see, e.g., p. 142), and so the second term, adopted in the interest of terseness, is my own. Sometimes in Bratman's writing (as on p. 295), but not always, "self-governing" covers policies that are weight-bestowing and reason-determining.

Two observations: First, although Bratman does not emphasize it, these two concepts travel together; setting the weight of a prospective reason to zero is determining that it is not to play a motivationally effective role in deliberation and, conversely, giving it a positive weight is determining that it will. Second, although the discussion is conducted largely in terms of weights, restricting the policies in question to setting weights would be unnecessary and unprincipled. Just for instance, one could adopt the policy that a particular kind of reason is to be lexically ranked over another, even though lexical rankings are not representable by weights. The point is acknowledged at Bratman (2007, p. 300), but not systematically reflected in Bratman's terminology or illustrations.

<sup>14</sup>A variant of one introduced on *ibid.*, p. 301.

<sup>15</sup>For transparency, see *ibid.*, pp. 181, 191f. This condition excludes cases like this: you might have a policy of not starting to write until you've sharpened your pencils, made yourself a cup of coffee, read all the relevant literature, and reorganized your filing system. The real cognitive function of the policy is procrastination: it's a way of not starting to write anytime soon, and maybe never starting to write at all. But the policy doesn't say this anywhere, so to speak, and when you neurotically begin to sort through the filing cabinets, you may be unaware of what you're really doing.

Philosophers don't generally share an understanding of what it is to be conscious of something. So bear in mind that unless you have a matching theory of consciousness—the treatment of Korsgaard (1996a, Chapter 13) would perhaps suit—action governed by a Bratmanian policy need not be *conscious* action.

<sup>16</sup>See *ibid.*, pp. 183f, 189f, 194, 211, 242.

procedural decisions require a consensus, reflexivity would require that that policy be supported by a consensus.) You are *satisfied* with a self-governing policy when it does not conflict with another of your self-governing policies. Lastly, an attitude, such as a desire or an intention or a policy, is *noninstrumental* if you have it not simply because it's expedient to have the attitude itself (as when someone is paying you to have it).<sup>17</sup> Just to have some shorthand for all this, let's call an effective, transparent, reflexive, noninstrumental,

---

<sup>17</sup>See *ibid.*, pp. 84, 190 (at the latter, Bratman is discussing the negative case, of inconvenient desires that give you reasons, not to fulfill them, but to get rid of them). The casual reader may be unaware how the requirement is an attempt to steer around a very large philosophical iceberg; here I will just take time out to explain what the iceberg is.

Let's allow for the sake of the argument that there are evaluations whose correctness is agent-independent, and where the direction of explanation runs from the correctness of the assessment to the agent's attitudes, rather than the other way around; call these (following Bratman) 'value judgments' (p. 172). Now, even if there are correct value judgments galore, it's implausible that, taken together with all the facts you have at your disposal, they settle many decisions you have to make: there's *slack*. When you drive up to San Francisco, you could take either 101 or 280, and the value judgments don't settle which it will be; you just have to form an intention, and follow through on it (e.g., Bratman, 2007, pp. 159, 166, 205, 212, 233). But what goes for simple, garden-variety plans to drive up the Peninsula goes double for Bratmanian policies: it is very hard to believe that the value judgments on which you rely in choosing a Bratmanian policy uniquely determine which Bratmanian policy you choose. Bratman takes it for granted that our surplus-value commitments can do a great deal of work for us. The visible tip of the iceberg is that it seems obvious to many people that you can't adopt intentions arbitrarily—for instance, because someone is paying you merely to intend something (e.g., Kavka, 1983; Millgram, 1997, Chapter 2).

The formal problem lurking below the water is this. On the one hand, when Bratmanian policies are stipulated to be noninstrumental, that condition is meant to rule out (roughly) cases in which the policy is not held for *proper* reasons, but rather because the policy is expedient to adopt. But on the other hand, policies take up slack, which is to say that when we act on their basis, we are acting not because we have *proper* reasons so to act, but because it was expedient to adopt some such policy, and the policy we adopted dictated acting in this way. How are we to square the need for the condition with the slack-assuming role of Bratmanian policies?

The substantive worry is that a Bratmanian policy won't really be *yours*—you won't be able to take it seriously—unless you think it really is a good idea, and a good idea largely *because* the things it says are reasons, really *are*. Accepting the policy in any other way puts one in a posture that philosophy professionals have no doubt at one time or another taken toward policies adopted in department meetings. ("Resolved that we will not consider the candidates' strengths in history while making this hiring decision.") One does it, but one doesn't really think these are one's reasons: because one is play-acting, the policy doesn't speak for one (even when it speaks in one's name). Or—a slightly more elaborate example—think of polite fictions, such as the pretense required of faculty by the honor codes of some universities, to the effect that students are not cheating on their assignments. You have to pretend, in every way, that the student did his own work (while the Honor Board holds a "trial"); you know that this procedure is the negotiated settlement of a war between the faculty and the students back in the 1930s, and you know it's settled; it's your job, and you pretend. But you're pretending: no matter how thoroughly this policy controls your inferences (or "inferences"), no matter how unwilling you now are to change it, the "belief" that the student isn't cheating isn't your belief; you're sure he *is* cheating! Bratman briefly considers what it's like to have this attitude toward a group decision in which one is participating (2007, pp. 307f), but the important question here is what it's like when there are no other participants to distance yourself from: when it's *just you*. Surely a life lived on the basis of Bratmanian policies adopted in this sort of way is life according to *Dilbert*. But then, since the function of plans and policies, in Bratman's account, is to take up slack, aren't your Bratmanian policies normally to some extent or other adopted in this sort of way?

self-governing, weight-bestowing, and reason-determining policy with which you are satisfied a *Bratmanian policy*.

With this notion in place, we can expeditiously review Bratman's reconstructions of the items on his agenda. You *identify* with an attitude (a desire in particular, but the account ought to generalize) if you have a Bratmanian policy to the effect that the attitude should play a designated role in your practical reasoning: in the case of a desire, that it sets ends for you, with such and such weight or force. You *value* something when you want it, and you have a Bratmanian policy in favor of treating your desire as a justifying consideration in your practical reasoning. Acting on the basis of deliberation governed by a Bratmanian policy is *self-governed* action, which is, once again, autonomous action in the sense most of interest to us.

The pivotal role of Bratmanian policies in these reconstructions of identification and so on is motivated by the thought that what makes you the same person over time is psychological continuity. (That is, Bratman is accepting what these days gets called a neo-Lockean approach to personal identity.) In Locke, it is in virtue of the memories you inherit from past selves that you are the person you once were; recent revivals of the view allow other states to count as well, and intentions in particular: that in the future you will act on decisions you now make contributes to that future self being *you*. Now, policies, as a variety of plan or intention, are stable, and so tend to persist across time. Self-governing policies, which regulate and control one's other, garden-variety plans and intentions, function as a sort of cross-temporal spine of one's agency: it is not just that you do something tomorrow because yesterday you decided to do it, but that you decided to do it because, long ago, you adopted a policy about what reasons were going to figure into your decisions. Because such policies largely constitute you as a person who persists over time, and because they are in the business of endorsing and disavowing, when they speak, they speak for you. So what it is for you to identify with a desire, say, or an action, is for a policy of the type we have just specified to endorse it.<sup>18</sup>

Here, then, is the picture of agency we are being offered (and, once again, we are interested in it because it is a best representative of a family of models of agency that amounts to today's philosophical common sense)<sup>19</sup>: The agent is organized by and around a set of mutually compatible long-term policies.

---

<sup>18</sup>Bratman qualifies the claim by allowing that other psychic structures—he calls them “quasi-policies” and mentions ideals as one possible type of quasi-policy—might serve an organizing role very similar to that of Bratmanian policies. We are not told much about the workings of these alternatives, but in Section 10.5 I will pause to verify that quasi-policies do not, so far as we understand them, short-circuit the argument.

<sup>19</sup>Though it is also part of philosophical common sense to have qualms about how psychologically realistic such elaborate constructions can be, as when Watson (2004, p. 25) remarks: “That most people have articulate ‘conceptions of the good,’ coherent life-plans, *systems* of ends, and so on, is of course something of a fiction.”

These policies specify what counts as a reason, and how much of one, when you're making up your mind what to do; they lead you to act on (or make you balk at acting on) other garden variety desires and intentions; they are reflexively self-endorsing. Because these policies are such important contributors to your personal identity over time, and because they are policies of endorsing or disavowing reasons for action, what it is for you to identify with an attitude—for it to be *really* rather than *merely* yours—is for it to be endorsed by such a policy. What it is for you to value something is for you to identify with your desire for it. And an agent who acts on the basis of attitudes that are superlatively his, in pursuit of things he values, is autonomous.

## 10.2

At this point, I want to pause to flag a handful of objections the reader is likely to be entertaining, both in order to forestall what would otherwise be ongoing sources of distraction and to foreshadow moves and arguments that we will be taking up in due course.

First, recall the move from, A Bratmanian policy holds me together, and constitutes me as the same person over time, to, When one of my Bratmanian policies speaks, it speaks for me. (Call that Bratman's *Master Move*.) Bratman does not give an argument proper for his Master Move; it is usually marked with some such phrase as: "This suggests the conjecture. . .," "This makes it natural to suppose. . .," "Or so it seems to me reasonable to say." But evidently an argument is needed, because there are a great many things that are held together by components that do not thereby get to speak in their names. Books are held together by their covers, but what the cover says is not necessarily what the book says. States are held together by their police forces, and churches, sometimes by their censors, but what the police or the censors endorse is not necessarily what the state or the church endorses. Even when the police department announces the very same thing as the legislature's designated spokesman, it is the spokesman who is speaking for the state, and not the police department.

Cases like these can serve as models for what looks like a direct counterexample to Bratman's proposal. Fantasies don't normally speak for one (although they can of course be very revealing), and one can have fantasy policies; for instance, I might while away the lazy summer afternoons working up self-governing policies that state what ends will have what weights in my deliberation, in circumstances that I'm quite convinced will never come to pass: perhaps policies that cover how I would deliberate if I suddenly had large amounts of money. ("Whenever I'm approached by a development officer, I'll take it to be a weighty reason to offer his institution a large gift that the naming opportunity involves an embarrassing title"—i.e., I'm daydreaming

about endowing the Foolish Professorship of Philosophy.) Because I continue elaborating the fantasy policy, year after year, it is a major contributor to my neo-Lockean personal identity; in the sense at hand, this fantasy (and others like it) hold me together and constitute me as the same person over time. I have no inclination to change the fantasy, and so I'm satisfied with the policy; moreover, in the almost unimaginable event that I did get the money, I would probably act on the policy, out of sheer inertia; after all, it's not like I have other, competing habits and plans in place for handling large amounts of money. Nonetheless, the psychic function of this policy is that of a *fantasy*, and I can be entirely aware of that; when it speaks, it doesn't speak for me.<sup>20</sup>

The more minimal version of the point is that the argument for Bratman's Master Move is missing, and we shouldn't buy into his position without one. But there is another way to take the problem. In order to determine whether a distinction is correctly tied to an account of personal identity, we would need first to be clear about what the distinction is *for*. Now, there is a shared assumption at work behind the various explications of superlative agency, namely, that the difference between mere attribution and superlative attribution is *important*: that it makes a great deal of difference whether an attitude—belief, desire or policy—speaks *for me*. Explaining the alleged link between superlative attribution and personal identity requires us to step back and ask why we *care* about superlative attribution.<sup>21</sup> This of course can be a rhetorical question,<sup>22</sup> but although I am sensitive to the ways that philosophizing on

---

<sup>20</sup>Philosophers these days seem insufficiently sensitive to the distinction between fantasy and other “pro-attitudes,” especially desire. The most important difference between them is that daydreams are themselves consumed, whereas desires direct one to an object that is to be consumed. For instance, I'm told that a recent Bond movie has free running (or parkour) scenes in it. (A free runner will run up, say, a crane, jump off the end onto the ledge of a building, run along the ledge. . . .) Presumably the scenes are there because people enjoy the fantasy; Bond movies are canned fantasies that people pay \$10 or so to see. But I'm pretty sure that most of the audience members, if offered an opportunity to engage in free running, would decline an activity they are quite aware will be scary rather than enjoyable. They don't want to free-run; they just want to consume the fantasy, and more generally, people typically consume fantasies of things they don't desire. (I'm grateful to Elizabeth Calihan for bringing this to my attention.)

<sup>21</sup>Although Bratman himself does not explain why the distinction matters, other theorists do, and the point is usually forensic: we can only be held *responsible* for what we think, decide and do in the full-fledged sense. (We can name enterprises with this sort of orientation *Perp Theory*.) However, the fit between the sorts of hierarchy-oriented theories we find in this debate and the forensic social function is bad. The distinction, forensically construed, is important in the first place because it *has* a social function: if you do something rude, sometimes you get to say, “I'm sorry, it wasn't me speaking—it was the booze.” When you disown your attitude or action, it can have real social consequences, and therefore this social function couldn't be served by a subtle and complicated distinction made out at the upper reaches of a hierarchy of attitudes in the privacy of one's mind. If those who invoke the distinction aren't going to be *getting away with stuff*, the legitimacy of the excuse has to be something that others can check on.

<sup>22</sup>As it is in Foucault (1984), which reminds us that the concern for claiming authorship is a recent cultural phenomenon. (Recall that one prominent agency theorist, Christine Korsgaard, adopts the vocabulary of “authorship” to mark superlative attribution: you *author* your actions.) It's also a

this topic sometimes looks like the product of an obsession, rather than well-motivated problem solving, I do not intend it that way. When the time comes, I will propose a function (not, I expect, the only one; I mean only to be making a start on the problem I'm posing) for the distinction between what you really want, believe, and do and what you in some lesser sense want, believe, and do.

Turning now to a second and third problem: Recall that being "satisfied" with your deliberation-governing policy is a requirement imposed on its speaking for you, on its giving rise to autonomous action, and so on. The idea is that whether some candidate self-governing, temporally cross-referencing policy is really your own is only a live question if you've got a bad attitude about it—in this case, literally an attitude, in the philosopher's technical sense. After all, if you don't have the bad attitude, you can't plausibly claim to be alienated from your policy. In Bratman's appropriation of Frankfurt's idea, satisfaction comes out as a consistency requirement on the distinctive sort of policies that meet the other requirements on Bratman's list.<sup>23</sup>

However, it's unrealistic to expect a great deal in the way of otherwise-Bratmanian policies meeting the satisfaction condition. ("Otherwise-Bratmanian": policies that meet all the conditions on the list except that one.) Suppose that, as a child, you copy your policies from your environment. In

---

very oddly contoured concern: for instance, we only count *some* things as "authored" (papers, but not signed checks or letters of recommendation, will turn up in someone's *Gesammelte Werke*). Foucault is suggesting that the contours of and the concern with authorship more generally are the products of a random historical walk; once we understand the rough shape of this history, we are going to find ourselves wondering, Why care about this? Turning back to attitudes and actions, the Foucauldian thought would be: why should we build an equally contingent analog of this historically recent fixation into our theory of personhood? Why should we *care* if our attitudes or actions are "owned" or "authored"?

<sup>23</sup>The notion is taken over from Frankfurt (1999, Chapter 8), where it means that one is not disposed to alter one's psychic state. For Bratman's adaptation, note the "first" and "second approximations" on p. 35 and the "final modification" on p. 44. Why do it this way? Because it's easy to require that a superficial bad attitude won't do (say, a mere desire not to have the policy, one that amounts to no more than a whim). Once we require that the bad attitude be at a level of psychological depth similar to the one whose superlative attribution it undercuts, then it has to be another Bratmanian policy functioning to bind together the various temporal components of an agent—or a quasi-policy playing a closely analogous role.

Notice this tension. On the one hand, when defining "satisfaction", intuitively, there is an adjustment we would want to make to Bratman's rendering. Most of the time, having one policy amounts to having a bad attitude about another because they conflict in the obvious and direct sense: the policies tell me to do different things. But policies can also *undercut* one another, without conflicting, as when one policy tells me to pay no mind to gossip, and also to do such and such, while a second policy, which I picked up on the gossip circuit, agrees with the first that I should do such and such. You would think that Bratman-style satisfaction was best construed as requiring not just that one have no further policies that conflict with the policy in question, but that one have no policies that undercut the policy in question.

However, on the other, that emendation would make it *stipulatively* true that one is unable to revise one's policies in a way that counts as autonomous and fully attributable to oneself. For such a revision would have to be policy driven, and to do its job, that policy would have to undercut some further Bratmanian policy.

that case—but pretty much any plausible childhood alternative to mere copying will share this feature—you shouldn’t expect consistency. Why should your parents, the TV set, and your childhood peers all have the same, or even compatible, policies? If you are to render the randomly copied policies consistent by sorting them out, *you* sort them out, in which case the account helps itself to the very type of attribution it’s trying to explain, before the conditions for such an attribution could have been put in place, and so is viciously circular. But if you can’t sort the inconsistent policies out, fully Bratmanian policies will be rare: so rare, that if they are what account for superlative ownership of attitudes, valuing, and autonomy, these latter will themselves be too exotic to call for much in the way of philosophical attention. (Call this second problem the *genetic objection* to satisfaction.)

Third, the desirability of satisfaction should not be taken for granted. Remember the nineteenth century’s *Sturm und Drang* movement; Romantics worried that if you didn’t have different attitudes struggling to correct one another (that is, if you weren’t dissatisfied with yourself, in both Frankfurt’s and Bratman’s variants of the notion), your personality would be static, and you would be incapable of maturation and growth. (Call this the *Romantic objection* to satisfaction.)<sup>24</sup>

I do not want to pursue these objections as they are. Our observations—that we are unlikely to have many deliberation-governing policies with which we are satisfied, that there is a history of thoughtful nervousness about the settled and static personality, and that we do not have an argument for Bratman’s Master Move—will turn out to be dry runs for more pressing complaints anchored in an account of humanity’s shared adaptive strategy. A correctly functioning personality will not only often be unsatisfied with its otherwise-Bratmanian policies. It will be so especially when we understand the agent to be thinking *for himself*.

### 10.3

Psychological states and psychological structure characteristic of an organism are ordinarily to be understood as part of the organism’s strategy for responding to challenges it faces. To be sure, this claim has to be qualified in many ways: natural selection does not always produce cleanly engineered, optimal solutions to adaptive problems; in culturally plastic creatures (i.e., us), not all psychological structure is a product of natural selection. Nonetheless, an organism’s psychology should make sense as part of the organism’s life.

---

<sup>24</sup>Frankfurt (1999, p. 102) is clearly aware of the objection, and he goes to the trouble of denying the Romantics’ claims—but not of providing a convincing argument against them. (Margaret Bowman helpfully pressed this concern.)

A trivial for-instance: plans and policies control behavior over extended periods of time, and so a creature that lived only for moments—while it *might* have a mental life—would have no occasion for either plans or for policies.<sup>25</sup> In assessing Bratman’s proposal, the first question to ask is: What kinds of creatures are *we*?

We survive, and often thrive, in a world that is, albeit at intervals, deeply surprising, and by “deeply” I mean that our collective experience teaches us that there’s pretty much nothing we can’t be surprised about. We get along in our surprise-laden environment by availing ourselves of a distinctive capacity for ecological specialization. Where other species become specialists in hardware, so to speak, reshaping themselves over evolutionary time to have, say, the long neck needed to reach those upper branches, we do our specialization in software, and the software allows much more extreme specialization than one tends to see otherwise: call it *hyperspecialization*. What is more, the software is reprogrammable, which means that, over and above the collective human capability of occupying multiple niches in an ecology, an individual human can occupy them one after another.

Now, parts of our environment are, for sometimes quite long stretches of time—in suitable respects, and after a while—*unsurprising*. In making them serve us as niches, we identify their relatively stable features and develop representational schemas, systems of standards and methods of deliberation that allow us to occupy and to exploit them. But, and this happens for many reasons, humans often have to move on: to exit one niche and either go looking for another, or learn to manage in parts of the world that do not have the stability or the structure to amount to a niche.

Over time, we have accumulated a large inventory of attempts to complete the sentence, “Man is . . .” (You know, “Man is the animal that laughs”—slogans like that.) My label for this strategy is being volunteered as another

---

<sup>25</sup>Bratman’s work can be understood as a sophisticated descendent of Nagel (1978), one which takes up its discussion of the metaphysical interpretation of the self, and so clarifies the deep connection between temporal extendedness and practical rationality. Where Nagel tells you merely that what it is to be a temporally extended creature is for future reasons to give you reasons now, Bratman provides a great deal more in the way of such a metaphysical interpretation: *these* temporally extended creatures have reasons with weights set by policies that . . . etc.

In recent and related work, Ferrero (2010) has offered Bratman an account of the stability of intentions: your intentions are stable because you’ve delegated the decision to your past self, and you’re pretty sure he made the decision the way you would’ve. However, Nagel persuasively argued that internalist or present-aim-theory approaches to prudence are misguided: it’s a mistake to look for a reason for me-in-the-moment to care about my reasons at other times; if I-in-the-moment need one, I’m not really a temporally extended agent at all, but rather merely a community of time slices. Ferrero’s view is internalist in spirit, in that it attempts to give you a reason, one that makes sense to your momentary self, to heed your past self’s decisions. Evidently, the fit between Bratman’s project and Ferrero’s analysis is awkward; Ferrero’s analysis amounts to giving you further reasons to act on your previous decisions, where the propensity to *just* so act, without such further reasons, is a structural feature of creatures whose lives are extended through time.



contribution to it: *Human beings are serial hyperspecializers.*<sup>26</sup> Our present question, evidently, is what sort of agents serial hyperspecializers are going to turn out to be.

This intentionally biologized description of human life requires an example if it is not going to be misleading; because humans occupy many of the niches in the ecologies that they—after enough time—more or less constitute, for humans, ecological niches are often indistinguishable from social roles. Since the audience for this paper is likely to consist for the most part of philosophers, here is the close-to-home illustration: philosophy, seen as a profession or social role. (Here I want to leave open the question of whether that is a good way to think of philosophy; suffice it for the present that the professionalization of the activity is widespread and familiar, whatever its merits and demerits may be.) Philosophers today take it for granted that their potential employers are, almost all of them, institutions of higher education; that publication is required; that venues for publication include professional journals, books bearing the imprint of an academic press, edited collections of essays, book reviews, and—very recently—blogs. (These are examples of relatively stable features of the environment that define a niche.) In the United States, there is a widely accepted institutional pecking order, defined by someone's top-fifty list; there is a slow-to-change map of areas of specialization, that is, subfields such as philosophy of mind, history of modern philosophy, or philosophy of biology. (The first is an example of a system of standards adopted to make the niche navigable; the second is an example of a representational schema used to navigate the niche.) Against this sort of background, the deployment of plans and policies makes sense. Plain, ordinary plans might include the intention to specialize in such and such an area, to publish in such and such journals, and to write an academic book. An example of a Bratmanian policy might be the intention to count Leiter Report rankings as overriding considerations wherever they are relevant.

To reiterate, human life is an uneven mixture of the routine and the novel, and individuals may need to leave a niche—sometimes abruptly, and for reasons they could not have anticipated. In Germany, in the 1930s, philosophy professors did not read the Leiter Report, so let's imagine one of them whose Bratmanian policies include accepting the advice of his Doktorvater, not making life choices on the basis of outlandish rumors, and not letting politics impinge on career decisions. Suddenly, this philosopher discovers that he can no longer hold an academic job, that he can't use the general post office or employ his housecleaner, that his *name* has been changed, and that he has to wear a gold star. His Doktorvater tells him not to take it so seriously, that times are tough but that they're bound to get better, and that while the ruling party is overdoing it in his case, nonetheless, one has to admit that something had to

---

<sup>26</sup>For further discussion, see Chapters 3, 4, and 9.

be done about Jewish control of German cultural and economic life. The philosopher is hearing outlandish rumors about what is happening to his arrested and ‘resettled’ coreligionists. Overall, politics has come to pervade his life in ways he would have found unbelievable only ten years before.

As the illustration reminds us, working up plans and policies for genuinely unexpected change is at best an exercise in futility and at worst incoherent: if you don’t see it coming, you can’t have a plan for it. (Even if you are in the unusual position of being able to enumerate all the possibilities, when there are too *many* distant possibilities of this kind, you cannot formulate well-considered contingency plans for all of them.)<sup>27</sup> Serial hyperspecializers—us!—sometimes (not always: there are obviously other reasons as well) abandon the niches they are in because they are faced with genuinely unexpected change. So they often are not acting on plans or policies when they exit niches, and bear in mind that in dramatically new circumstances, it takes a while to come up with a reasonably adopted plan or policy; in such circumstances, we should expect that, for a while at least, their actions are not policy governed. Psychological structure, we observed, should fit into the shape of the life of an organism. So we should expect to find the psyches of serial hyperspecializers to be a mixture of plans or policies (suitable for use in relatively stable niches), on the one hand, and psychic equipment for coping with the impossible-to-anticipate, on the other. An account of agency that is made out solely in terms of plans and policies can be no more than half of the theory we are after.

If the structures of (and strictures for) agency match the shape of an organism’s life, if humans are serial hyperspecializers, and if the life of a human being is normally segmented—a philosopher for so many years, then a party journalist for a period, then a media consultant, etc.—we should expect to find that human beings are, always potentially, and frequently in fact, *segmented agents*.

## 10.4

At this point, the skeptical reader may be wondering whether our critical assessment of mainstream accounts of agency presupposes too much in the way of a controversial and very-big-picture recharacterization of humanity.

---

<sup>27</sup>The point that you can’t prepare for what you can’t anticipate applies not only to policies but to quasi-policies, and here is a place to mark a complaint that readers of Bratman often come up with, but which we can put aside. McDowell (1998, pp. 57f) has observed that not all of our intentions are “codifiable,” whereas Bratman’s policies seem to amount to codification. I want to allow that Bratman-style policies need not be fully articulated, or even articulable by an agent who has them. But we should not mistake action-guiding attitudes which you are unable to spell out for attitudes capable of providing thoughtful guidance in situations the like of which you have never encountered, and in which you have never invested any thought. That you cannot say what your policy is does not allow your policy to violate the law of conservation of cognitive work.

So let me emphasize that the problems to which the recharacterization directs our attention are problems by the lights of ordinary common sense. The perhaps overly dramatic (but nonetheless real-life!) illustration was meant to make vivid the point that we can be faced with the genuinely impossible-to-anticipate, and therefore with circumstances for which we cannot have prepared suitable plans or policies. A fortiori, we cannot have prepared suitable Bratmanian policies. Sticking with the example, it is clear in retrospect that the right deliberative move for the 1930s German-Jewish professor is to drop his Bratmanian policies like a hot potato: those who didn't (of course, along with most of those who did) ended up being packed into cattle cars and shipped off to slaughter. Now, Bratman insists that he "does not mean [Bratmanian] policies [to be] immune to rational revision . . . [his] project is not to describe some irrevocable foundation at the bottom of all further practical reasoning."<sup>28</sup> But suppose the professor *does* drop his Bratmanian policies: can this be, by Bratman's lights, autonomous action, something that *he* really *did*?

Bratmanian policies are reflexive: they apply to themselves. So one can "reflectively reassess and revise where one stands" with respect to one's Bratmanian policies—if one has a Bratmanian policy, triggered by the conditions in which one finds oneself, that tells one to revise one's policies in those conditions.<sup>29</sup> Although there are delicate issues to navigate, about how to individuate policies, and when one can square revising one policy on the basis of another with the satisfaction requirement on Bratmanian policies, we ought to allow that policies can be surrendered on the basis of an agent's further policies. (The ship which Neurath described as being rebuilt at sea, plank by plank, is familiar philosophical shorthand for this point.) Perhaps I cannot drop all my Bratmanian policies at once, but can drop any one of them, as long as there are at that time other and suitable Bratmanian policies that I am retaining. And eventually, revising my Bratmanian policies one by one, I can revise them all. What is more, policies can be rough-hewn and broadly framed, with an eye to problematic circumstances: "In an emergency, stop, look, and listen." All that granted, if a policy is not so broadly framed as to be entirely content-free, there will nevertheless be circumstances it does not

---

<sup>28</sup>Bratman (2007, p. 36).

<sup>29</sup>*Ibid.*; in his early work, Bratman introduced plans as stable in this sense: it takes special circumstances to make one reconsider them. (See *ibid.*, pp. 289f for an attempt to derive a commitment to taking the means to your ends from the requirement of means-end consistency together with stability: stability is thus being treated as a very deep feature of practical rationality.) At the time, he treated the disposition to reconsider a plan as not itself further analyzed, as determining, roughly, *how much* it would take to make you reconsider, and as subject to roughly consequentialist assessment: the disposition, along with the setting on its dial, and thus failures to reconsider plans on particular occasions, could be justified by showing that on average the results were good. Bratmanian policies amount to an alternative to both the "volume setting" model—they allow finer-grained triggers for reconsideration—and to the consequentialist mode of assessment; they are thus an advance on the earlier position.

cover. And that is not a bad thing: because one cannot have given thought to unanticipated circumstances, a policy that did cover all circumstances would be a thoughtless—a *foolish*—policy to have. Running with that last proposed policy, not everything that requires rethinking what you are doing is an emergency; or alternatively, you cannot be equipped to recognize everything that counts as *that* sort of emergency; in any case, whatever we are told in grade school, stopping, looking, and listening is an appropriate response only to *some* emergencies.

I have found that philosophers, exhibiting what seems to be a professional reflex, respond to illustrations of unanticipated circumstances by trying to think of policies that would handle them. The response is common enough to justify taking time out to explain why it is misguided.

First of all, the order of the quantifiers matters. Even if it were true that, for every circumstance you might face, there is a policy you could have that would cover it, it would not follow that there is a policy that would cover every circumstance you might face. I have already suggested that, because you can only have put so much deliberation into any policy, adopting a policy that covers all bases is almost inevitably a bad idea.<sup>30</sup>

Second, such an all-purpose policy is supposed to manage an intelligent response to whatever circumstances you encounter. Policies amount to collections of rules launched by specific trigger conditions (if you have a programming background, a policy is a lot like a Lisp `cond`). And that allows us to recognize the policy these objections are after as the holy grail of 1980s AI, the program of machine intelligence through rule-based systems. Some very bright people were unable to make this approach work; if the MIT Artificial Intelligence Laboratory failed to find the policy you are imagining, why be so confident it's there for the having? There was a reason that the field abandoned the approach and moved on.

Third, and last for now, even if a policy that covers all the bases exists in the Platonic Heaven of the Forms, it's clear enough that most agents don't know about this policy and haven't adopted it. (If they had, surprise would itself be a surprising thing.) A theory of agency should not restrict itself to agents who have adopted some policy that already solves all the challenges of agency,

---

<sup>30</sup>See Millgram (1997, Chapter 4) for an argument that an agent will not be able to live by a contentful plan that is chosen without taking into account experience of the sorts of circumstances the plan covers.

And here is a quick argument that you cannot in fact have a policy that covers all bases. For serial hyperspecializers, a policy that covered all bases would have to cover activities within specialized niches. It is characteristic of these specialized niches that their occupants develop descriptive vocabularies needed to navigate the niche, and that they are not intelligible to outsiders. The trigger conditions of policies that govern intelligent behavior within such a niche will have to be largely formulated in the specialized vocabulary. Since anyone is an outsider with respect to most specialized niches, no one can so much as understand a policy that covers all bases. You cannot have a policy you do not so much as understand. Consequently, no one can have a policy that covers all bases.

in something like the way that a methodological proposal in philosophy of science should not restrict itself to the scientists who have already arrived at the final Theory of Everything.<sup>31</sup>

Returning to the illustration at hand, I constructed the example so that the German-Jewish professor's circumstances do not trigger policy-driven reassessment of his Bratmanian policies, and that is a legitimate stipulation because one's circumstances can be impossible to anticipate, and so one's Bratmanian policies may well not be tuned to the conditions in which, at one juncture or another, one finds oneself. If we are considering such an occasion now, and if the professor in the illustration does drop his Bratmanian policies, it is neither by virtue of their reflexive application nor by virtue of their Neurathian application one to another, and therefore it is not by virtue of applying his Bratmanian policies at all.

Recall that Bratman, unlike most of the competition, does not insist that his account of attribution and autonomy lays out the only way an action can be self-governed and fully yours. So what we can say is that Bratman's account does not show us *how* the decision to drop a Bratmanian policy in the face of genuinely unanticipated circumstances can be, in the full-fledged sense, your decision, and how actions pursuant to that decision, but taken before replacement Bratmanian policies are formulated, can be autonomous action.<sup>32</sup> Even without bringing to bear the characterization of human beings as segmented agents, we can see that something has gone badly wrong: being sent like cattle to the slaughter is the very opposite of self-governing action; and if and when the German-Jewish professor ditches his Bratmanian policies and flees the country, he is, as we say, taking matters into his own hands and thinking for himself.<sup>33</sup>

---

<sup>31</sup>Christine Chwaszcza has suggested to me that perhaps the bias toward full-information accounts of preference and the like is best explained as an inheritance from an older philosophical tradition that attempted to adopt the point of view of an all-knowing God. If that is right, once we have left behind the theological interest in God's plans for us, we should also drop the methodological orientation toward (what we can call) hyperintentionality derived from that interest.

<sup>32</sup>Thus when the argument, suitably reformulated, is directed to the competition, the upshot will be that when the German-Jewish professor swerves away from his former "constitution," higher-order desires, and the like, that choice *cannot* be autonomous, or fully attributable to him.

<sup>33</sup>Chesnow (2008, p. 16) describes his parents' postmortem of their insufficiently proactive response to their similarly changed circumstances. His father ran a business in one of the Baltic states that was doing very well on military contracts in the runup to the Second World War, and failed to take advantage of opportunities to emigrate. Chesnow recalls his father recounting Kipling's instructions for trapping a monkey: you put a banana in a cage whose bars are wide enough to let the monkey's spread hand in, but not wide enough to let a fist clenched around a banana out. "My father," Chesnow continues, "finished this story with the words, 'I'm the monkey.'" When you can't give up your policy, except for reasons that the policy itself antecedently specifies, it's all too easy to end up being the monkey: Chesnow's parents did not survive the war.

As the example reminds us, we do not always respond to changed circumstances in the way they seem to demand. A lower-key example (due to Elias Moser): we've learned the hard way that relocating

## 10.5

Return to the ecological characterization I have adopted of our species. Segmented agents are a psychological adaptation to a life consisting of stints (both longer and shorter) in ecological niches that typically are also social roles. These niches are characterized by relative stability, because without stability, it would be impossible to have that part of the ecosystem serve as a niche. But there is no reason to suppose that there is similar stability everywhere else, and plans and policies only make sense against the background of a relatively stable environment. So we have to adopt a more nuanced view of the role of plans and policies—and thus, of Bratmanian policies—in human life. Intelligently formulated Bratmanian policies can be suitable guides for action *within a stable niche*. (With, however, an exception: an agent currently occupying a niche may need to think outside the box when deciding whether to push the eject button.) Such policies are not, evidently, suitable guides for the big, wide world, but, I want to insist, in *Structures of Agency* Bratman has done us an important service, that of diagramming psychological structures, and distinctions that accompany them, that serve us well in well-structured and well-understood environments.

Sometimes we're invited to think of ourselves as space-time worms, but worms don't have spines, psychic or otherwise; snakes have spines, so maybe the Bratmanian picture is an invitation to think of ourselves as space-time snakes. But we are not space-time snakes. Segmented agents have (roughly, more or less) one Bratmanian spine per segment: psychic structure that manages one's agency during the time one occupies a particular local environment stable enough to serve as a background for plans and policies. Let me emphasize that that's a first approximation: people can have a foot in more than one niche at a time, and they can spend extended periods of time outside of the constrained environments that serve as niches (in which case, they take on the ecological appearance of weedy species like rats or cockroaches and may fail to have Bratmanian spines at all). To say that each segment will normally have its *own* psychic spine is to say that these backbone segments are not normally connected one to the other in the way that their components are connected to one another: that is, one segment's spine is not joined to another's by plans or by policies. As a placeholder, we can say that they are held together by being all embedded in one temporally extended person.<sup>34</sup>

---

at-risk youth into a different environment, in the hopes of prompting new Bratmanian policies, is often unsuccessful.

<sup>34</sup>Vogler (2002, pp. 106f) complains about Bratman's earlier work that planning theory is suitable for the managerial classes (and only the contemporary version of them, because, centuries ago, managers didn't own dayrunners or scheduling applications). But it is not suitable for anyone else, and so it cannot be a satisfactory account of practical rationality, which has to be rationality for everybody. We are now in a position to amend Vogler's complaint: plans and policies are indeed usable, and not just

With this image in mind, we can say a bit more about why we should not be treating a model of a time-bounded segment of agency as a model of human agency over the course of a life. One challenge that human beings face is that of managing the transition from one segment to the next. Coping with this challenge can't simply be a matter of deploying psychological machinery that governs the segments; that machinery is effective because it can exploit the stability and constrained environment of a temporarily occupied niche; but stability and constraints on which one can rely are just what is missing in the transition from niche to niche and in the world at large. So reliance on Bratmanian spines must alternate with turns to a very different form of practical rationality.<sup>35</sup>

The problem I am posing has an analog in the philosophy of science. It is now a familiar doctrine that, during periods of normal science, scientists proceed on the basis of policies that determine what counts as successful argumentation, legitimate results, and so on; these policies express a conception of scientific rationality during that period. But during so-called scientific revolutions, those policies have to be abandoned and rethought from scratch.<sup>36</sup> Because the rethinking cannot be conducted in accord with the policies that

---

by the managerial classes, but in most niches: plans require, in the first place, stability, and if it's not stable, it's not a niche. There is something her objection is getting right, however: plans and policies work well neither in the large part of human life that is lived outside one or another stable niche, nor during transitions from niche to niche. That doesn't show that planning theory isn't part of the theory of rationality; it does show that it can only be part of the theory, because we need to understand what rationality amounts to when you are thinking outside the box.

Does my description of the human *ergon* imply that a well-lived life is spent dealing with emergencies? As I rendered the human species form, it might seem that the happy life must be spent leaping from one specialized niche to another, and that if you're not constantly throwing everything you cared about overboard, you're a failure. But surely a life without such upsets is the *better* life and the one to which we should aspire.

That your *ergon*—your design description—involves specifically *serial* hyperspecialization doesn't imply that if you don't use the capability, your life is thereby unhappy. The design solution that human beings implement includes many features that might never be used in a well-lived life: just for instance, the ability to metabolize your own muscle tissue when you're starving (or also, when you're doing exercise you haven't trained for properly), or the ability to learn more than one language natively. (Thanks to Michael Millgram for help with the first example; for second thoughts about applying the concept of *eudaemonia* to serial hyperspecializers, see Section 3.6 in the present book.)

<sup>35</sup>There is an important class of exceptions that I want to register: sometimes the niches are sufficiently entrenched, and the transitions between them sufficiently standardized, for there to be metapolicies governing transitions between niches and the Bratmanian policies appropriate to them. For instance, an employee may be governed by a Bratmanian policy at work which we can abbreviate as: be professional. At home, he is governed by a very different, much more domestic Bratmanian policy. His metapolicy is to switch from one to the other during his commute. Note that this metapolicy is underwritten by the stability and standardization of the arrangement: indeed, at this point, the segregation of the policies is supported by very-hard-to-change physical infrastructure, the differently located residential and industrial neighborhoods between which the employee must drive.

Obviously I think it's a mistake to understand the segmentation of agency as a sort of super-policy. But notice that even if it could be construed that way, for most of us it would not count as a Bratmanian policy; most people are unaware that their lives are segmentable, and in their case, the impossible super-policy would violate Bratman's transparency condition.

<sup>36</sup>Kuhn (1970).

are being jettisoned, it has been hard for both philosophers and historians to see how the outcomes of scientific revolutions can count as rational. Consequently, it has also been hard to see why scientists should get credit for them: should an Einstein get a Nobel Prize if his achievement was really just a bit of lucky craziness? But when Kuhnians treat paradigm shifts as irrational episodes sandwiched between the longer stretches of scientific rationality, rather than as among the highest intellectual achievements of the scientific tradition, achievements that stand out as the strongest examples of rational thought that we have available, that is evidently a theoretical failure on the Kuhnians' part. It is a philosophy of science suitable for a plodding and intellectually crippled version of the enterprise that might be conducted by a more limited species, one incapable of invention that was both genuine and thoughtful.<sup>37</sup>

It takes a big man to admit he's wrong, or that's what popular wisdom says. The problem we were homing in on is that, on a view like Bratman's, you can't ever be that big a man, because you can't admit you *really are* wrong, wrong all the way down: not wrong in ways for which your policies leave room, not wrong in ways that are made out through the reflexive application of self-governing, reason-determining policies, but completely wrong, wrong even about what would be a good reason to change your mind. Or rather, the problem is that coming to that point can't be something *you do*—though it can *happen* that you come to see your former attitudes as wrong, all the way down.<sup>38</sup>

Once the point is conceded for the dramatic illustrations, it should be allowed in more mundane cases of niche-switching as well. When we show, in a laboratory experiment, that certain birds are sensitive to the Earth's magnetic field, we correctly assume that that sensitivity is deployed outside the laboratory; if we convince ourselves, by considering artificial or extreme circumstances, that human beings are able to manage choice that is not policy driven, we should similarly assume that they will deploy that ability wherever there is an advantage in doing so.<sup>39</sup> A philosopher may decide that the business of philosophy is just not for him (and we philosophers can all

---

<sup>37</sup>Friedman (2001) suggests that thinking outside the box is *philosophy*, and that's why awareness of philosophical developments has an important role in scientific revolutions. While there is certainly something to the suggestion, not all thinking outside the box is philosophy; when a lurch in your career puts you in the position of having to figure out how you're going to reinvent yourself, your deliberations may well not count as philosophy by anybody's lights.

<sup>38</sup>Here Bratman is, again, representing mainstream commitments that are largely implicit, but occasionally explicitly acknowledged. For instance, Watson (2004, p. 26) remarks: "The important feature of one's evaluational system is that one cannot coherently dissociate oneself from it *in its entirety*. . . . One can dissociate oneself from one set of ends and principles only from the standpoint of another such set that one does not disclaim." However, a version of the worry I am pressing has recently surfaced; see Fisch and Benbaji (2011, p. 243).

<sup>39</sup>And indeed, as Jenann Ismael has reminded me, the phenomenon I'm identifying is built into the completely routine formal structures of our ordinary early lives. First you go to elementary school; then to high school; then to college; then perhaps to professional or graduate school; finally you emerge



think of people we know who did decide that, and decided on the basis of reasons for which their reason-governing policies had left no room). When he does, he will slough off the Bratmanian policies that guided him as a philosopher—perhaps that system of standards taken over from the Leiter Report—and doing so, without a Bratmanian policy to guide his deliberations about the matter, can be the most difficult, demanding, and admirable form of self-government at work.<sup>40</sup>

Bratman allows that there may be psychic structures which are not policies, but which play much the same role that his Master Move assigns to policies; he calls them quasi-policies, and mentions ideals as an example. So let's just check that the point we're making sticks even when we have Bratman's quasi-policies in the mix. One very striking aspect of the human predicament is that ideals often have a shorter shelf-life than people: call this the *Shelf Life Problem*. (Think of the ideological politics of the twentieth century, during which communists outlived the relevance of communism, and Germany's "revolutionary nationalists" outlived the respectability of their movement.<sup>41</sup>) It is unusual that such an ideal contains within itself the intellectual and emotional resources for recognizing that its time has passed; on the contrary, well-developed ideologies generally provide catch-all methods of dismissing the reasons adduced against them.<sup>42</sup> When someone (as many, perhaps most

---

from this series of cocoons and get a real job. At each stage, your policies, goals, preferences, and self-image have to be reworked pretty much from scratch. It's not just that an elementary school pupil who tried to formulate a way of getting through the world intended to guide him through the subsequent stages would be very peculiar; even if he tried, he would be badly served later on by sticking with the plans and policies he'd made up as a child. Later life is in fact no different from childhood; as you mature, you have to rethink your Bratmanian policies and other such components of your personality and—we can hope—you're never done maturing.

<sup>40</sup>Let me add a qualification to my endorsement of Bratman's account as a satisfactory reconstruction of how we manage within a stable niche. Because human beings are opportunistic when it comes to exploiting available resources, cognitive and otherwise, we should expect that when we focus tightly enough even on activities within stable niches, we will also find action not fully controlled by Bratmanian policies. Let's revisit the very low-key example used by Bratman, long ago, to introduce plans, and which we mentioned in footnote 17. Once again, you are about to drive up from Palo Alto to San Francisco, and you could take either of two freeways, 101 or 280; but now, let's imagine that you have adopted a Bratmanian policy covering such decisions, namely, you will choose the route based on the time you expect it to take, with stop-and-go traffic as a tie-breaker. Let's further imagine that, in this case, 101 and 280 are on a par, as far as your policy is concerned; if you simply decide to take one or the other, it will have been full-fledgedly your own choice. But now, suppose that, in these circumstances, you form an intention to take 280 because it strikes you that the scenery along 280 is much nicer. This is introducing a further consideration, one not covered by your Bratmanian policy. So is the ensuing drive up superlatively *your* action? My own sense is that, in circumstances like these, the conceptual apparatus should be thought of as delivering a mixed answer: it is your decision, up to a point. And such mixed outcomes need not arise only in trivial circumstances: a physician, for instance, is only too likely to have to make choices that, while they conform to his niche-specific policies, also outrun them, and in just this way. (I'm grateful to Paulina Sliwa for this last example.)

<sup>41</sup>The label comes from Rose (1990).

<sup>42</sup>Unusual, but nevertheless, some ideals exhibit the phenomenon Nietzsche calls "self-overcoming," in which the rigorously applied ideal is turned against itself. What matters for the present

of us, will have to) lets go of an ideal that has reached its expiration date, that is, when it is done right, their choice, in the fullest-fledged sense, and autonomy at its best; but it is not, usually, a choice that is guided by the ideal itself—or, for that matter, by a distinct and previously formulated policy for dealing with one's ideals.

Recapping, and spelling out the problem a little more slowly, allow that it is a frequent enough requirement in the lives of segmented agents that they have to admit they were wrong. Allow also that this sort of reconsideration can be understood as autonomous, and the attitudes involved in it, as superlatively your own. If the only way we now have on board to understand superlative attribution and autonomy is made out in terms of Bratmanian policies, then to accept the Bratmanian account as adequate is to presume, for practical purposes, that such reconsideration can always be managed by one's Bratmanian policies. (If you didn't accept that, you'd make room in your account for full-fledged choice of other sorts.) That in turn is to presume that we never have to reconsider Bratmanian policies in ways those policies themselves don't license, which is to presume that there are certain things we can't be wrong about: as I was phrasing it earlier, that we can't be wrong all the way down.

Once again, the view we are considering is being taken up as a best representative of a large and mainstream family of views. And you might be wondering whyever a philosopher would get himself into the position of assuming, even implicitly, that you can't be wrong all the way down: the supposition is strongly belied by ordinary experience.<sup>43</sup> Rather than take up this diagnostic question, let me turn briefly to the task of replacing obsessive preoccupations with useful devices, and to proposing a real function for the distinction between attitudes that are *merely* attributable to you and attitudes (and choices) that are *superlatively* attributable to you.

---

point is, first, that self-overcoming is not (as Nietzsche seems to have thought) the fate of every ideal, but only of some of them. And second, the criticisms an ideal is able to direct against itself are often not the ones that matter: the ideal of fairness may motivate a policy that you are later able to understand to be unfair; but perhaps the readjustment required in the face of devastatingly unfair damage to your prospects is accepting that sometimes you have to let go of the preoccupation with what is fair, and get on with what is left of your life.

<sup>43</sup>While it's easy to wonder whether one or another version of noncognitivism is in the background, Bratman (2007, pp. 174, n. 29, 226, 297, n. 26) intends his treatment to be metaethically neutral.

I am insisting that you can be wrong, completely wrong, wrong all the way down—and that, for human beings, failure to acknowledge this has the distinctive look and feel of self-deception. And isn't this just skepticism, toward which we should also be considering a diagnostic posture?

Not really. The iconic skeptical suggestion is that you *might* be wrong about everything (not that you really are), and if you were, you would never know. Philosophers' skepticism, the modern version anyway, is an intellectual enterprise where the live-but-not-actual possibilities are used to reassess the epistemic status of your beliefs. Whereas I am reminding you of something that grownups already know—that you *are* wrong about a very great deal and that you're all-too-likely to find out, probably in ways that amount to hardship. The epistemic status of your beliefs and other attitudes is beside the point.

## 10.6

A segmented agent can come to realize that its policies (and a fortiori, its Bratmanian policies) are thoroughly wrong, wrong all the way down. That means that it must be cognitively equipped to tell when its policies are bumping up against a recalcitrant reality, but we can be pretty sure that a segmented agent's equipment for telling that things have gone wrong in a way that requires massive revision *won't* be a policy. Things go wrong in unanticipated ways, and there are no criteria for telling when *that's* happened, and no rules for responding to unanticipated circumstances. Policies just set criteria or trigger conditions for rules, and consequently you can't have a policy for handling the unanticipated.

In those circumstances, a person will have to rely on *himself*, rather than his policies: from which it follows that the distinction between himself and his policies must make sense, and make sense to him. Returning for a moment to Bratman's Master Move, this is a reason to resist the assumption that when your policies speak, *you're* speaking. The need for that distinction should not be taken for an occasion to revert to the untenable picture of an immaterial self that is different from and concealed behind the individual's psychology. I will presently take up the question of what the third way here has to be.

I earlier recommended reopening the question of why we care (and whether we should care) about superlative attribution—and, by implication, the distinction between autonomous and nonautonomous action. However, before making my own suggestion on this score, I need to register a caveat. These are starting to seem like useful concepts (as construed by Bratman, but, yet again, I take him to be giving the best available rendering of a widely shared view), when applied *within* the local framework of a niche. *Outside* those niches, it's less obvious that these are the most interesting or important contrasts and distinctions to pursue. This is perhaps especially striking in over-dramatic examples of the sort we were considering. On the one hand, even a self-aware agent is likely to be overwhelmed by such circumstances; autonomy is a concept used to assess the quality of one's choices, and when one is swept away, one often has, as we say, no choice. On the other hand, when people rise to the occasion, even in such circumstances, we say that, then especially, they are thinking for themselves. Evidently, the problem with the sorts of account we are considering is not just that they don't provide the wherewithal to explain how the German-Jewish professor's decision to ditch his prior Bratmanian policies can be *his* choice, but that they require that attitudes, choices, and actions fall on one side or the other of distinctions that they rather seem to straddle when agents are performing well in unstructured and challenging conditions.

So the point of the argument we have been constructing is not that we have the view that certain responses to real or imagined circumstances are

autonomous; that Bratman's account does not classify them as autonomous; and that that is an objection to the account. It is rather that we *need* intellectual equipment with which we can assess performance in a hiatus between agential segments, and that the notion of autonomy as construed by Bratman (and the other philosophers for whom I am using him as a representative) does not meet that need.

Now, suppose a segmented agent has realized that its former policies, Bratmanian and otherwise, have to be replaced. Since we are using Bratman's account as our foil, we can assume that its psychology mostly consists of a mass of ingrained and habitual policies, attitudes that are endorsed and underwritten by such policies, and other psychological structure that has accreted around the relatively stable Bratmanian policies that managed activity during a previous segment of its life. All of that won't go away in a moment, and at the onset of the changeover a newly invented replacement policy will be no more than a tiny, not-very-well-entrenched wiggle in the psychic stew. Successful changeovers will require self-monitoring, to allow a segmented agent to catch itself when it goes on doing things the old, habitual way. So a segmented agent needs to be able to say, in what we can think of as a proleptic or anticipatory register: No, that's not *my* policy anymore. This is *a* function that the distinction between merely mine and superlatively mine serves; I strongly suspect that it is not its only function (and so, this is only a first step in addressing the question of what our practices of superlative attribution are really for), but it is at any rate a job that has to get done.<sup>44</sup>

Notice that the proleptic form of the superlatively-mine/merely-mine distinction is deployed only when one is, in the sense we introduced earlier, *dissatisfied*, whereas recall that Bratman imposed satisfaction as a precondition for making the distinction at all. Here's what's at stake (returning to and upgrading the Romantic objection): for segmented agents, dissatisfaction is a *normal* component of a *successful* life. Consequently, and leaving aside the technical sense of "satisfaction" for a moment, it is not something about which they should be dissatisfied; when you rise to the occasion, and the occasion requires you to switch gears in unanticipated ways, you may well feel, not dissatisfaction, but appropriate pride in taking a stand against older and perhaps deeply entrenched attitudes. To switch gears in this way is to come to contain competing and typically independent structures of agency. These structures—whether construed as Bratmanian policies, or as quasi-policies, or as "constitutions," or as hierarchies of higher-order desires—are *normally* conflicting and mutually undercutting. Therefore, during the period in which

---

<sup>44</sup> A good way to see that it's not the only use to which we put such distinctions is that even when one correctly says to oneself, No, that's not what I think! one may later not get to say, to someone else, of the very same attitude embedded in the very same episode, No, that's not what I thought—or anyway, not with the same definitiveness. (I'm grateful to Tom Pink for the observation.)

they overlap, the agent is (reverting now to the technical sense) dissatisfied. But that's (often) a good thing, and the agent can be entirely aware that it is.

## 10.7

We have been criticizing Bratman's view as a best representative of a philosophical approach that consists in looking for psychic structures with this feature: when they guide, you govern.<sup>45</sup> If I am right, we need to be taking a very different approach: that of making philosophical sense of forms of receptivity to the world such that, when the *world* guides, via those forms of receptivity, you govern. I do not have the account I am pointing toward in my pocket. But I do want briefly to address two sorts of incredulity that are likely to meet the proposal. How can it be *self*-government when the guidance is coming from outside of you? And how can the *world* be providing guidance? Why isn't that suggestion a return to superstition and the pre-Enlightenment magical thinking that takes decision-making to consist in the search for omens?

You can't generally understand or assess creatures outside of the environment in which they function, and if that is true of humans, then assessments of agency that consist just in looking *inside* a person, at his internal psychic structures, will get off on the wrong foot. If that is true generally, we should expect it to be true of assessments of autonomy and of the superlative attribution of attitudes and actions. Philosophers with the family of views that I am resisting sometimes analogize the person and the state, so perhaps it will help to remind ourselves that my recommendation *is* our practice with respect to states: we judge that Neville Chamberlain's England "slept," not because its internal structures of governance had changed, but rather because the British political system failed to respond intelligently to an unprecedented external challenge. That near-fatal lapse in Britain's collective autonomy was a matter of failure of receptivity: an inability to acknowledge and act on observations for which the prevailing policies left no room.

Returning now to individuals, examining organisms together with their environments allows one to pick out signaling pathways that serve to guide action; in the design solution such an organism implements, the content of these signals is *practical*.<sup>46</sup> Humans are equipped to inhabit much more variable environments than some other species, and so the contents of those signals are correspondingly flexible: some of them mark actions one has performed as successes, and others as failures; some mark changes in circumstances as improvements, and others as deterioration; some mark attainments or

---

<sup>45</sup>This way of phrasing the objective comes from Bratman (2009, p. 430).

<sup>46</sup>For discussion of such signals in some nonhuman organisms, see Sterelny (2003).

acquisitions as desirable, and others as undesirable; some mark avenues of exploration as promising, or the contrary. These signals are functionally, in all respects, practical observations; in humans, they are not used raw, but typically serve as inputs to inference and practical theorizing. A philosopher's example: one might experience a series of straight readings of Nietzsche as disappointing, and conclude that one should not spend any more time on them, but look instead to other ways of reading Nietzsche.<sup>47</sup>

Moral philosophers who have been around the block a few times will be quick to leap to conclusions: that I must be endorsing a form of "moral realism," on which we perceive—using sense modalities unheard of by science—what we might as well call *helicopter values*. (Like helicopter parents, they *hover*—in this case, over the physically existing objects in our world.) So notice that the signaling pathways I am gesturing at include the responses currently being investigated by hedonic psychologists. And while this is not the place to lay out a metaethical position, please do not assume that I hold moral realist views; on the contrary.<sup>48</sup>

Serial hyperspecializers are explorers of their environment, and their philosophers owe them an understanding of self-government on which it makes sense of autonomous exploration. When someone goes looking for a shortcut to Asia, and instead discovers America, or embarks on a surveying and bartering expedition, and instead conquers an empire, previous plans have become irrelevant, and a successful response (whether by that person or his successors in the field) consists in large part in being willing to cast off the previous standards that had guided the enterprise and to adopt novel reconceptions of success—and with them, new hopes and ambitions—rendered appropriate by serendipity.<sup>49</sup> Our schoolchildren are taught to admire the heroes of the age of

---

<sup>47</sup>By a "straight reading," I mean one that treats Nietzsche as a philosopher engaged in the first place in producing a theory that we aim to reconstruct, and that ignores or only cursorily acknowledges the elephant in the room, that Nietzsche *writes funny*. My own test for whether my students see them as disappointing illustrates how we exploit the signaling systems we have available. After presenting someone's straight reading, I have them vote on whether an academic now advancing this view would be worth their attention. After a few rounds of this, the students generalize from their own votes and conclude that straight readings, of this particular philosopher, are not worth their attention. Since the students have almost uniformly been brought up to think of the history of philosophy as the pursuit of such readings, we have a slightly exotic example of practical observation correcting policies in unplanned-for ways; this is autonomy at work in philosophy.

<sup>48</sup>For more of the picture and for supporting argument, see Millgram (1997, 2005a, Chapter 1, 2004a—and since I have had readers take *Practical Induction* for a moral realist tract, a belated reading instruction: it is carefully phrased throughout to avoid any such commitments). It does seem to me that, like other entries that make up the standard menu of metaethical positions today, moral realism is a metaethics unsuited to serial hyperspecializers; for further discussion, see Chapters 5 and 6 in the present book.

<sup>49</sup>The first of those cases shows that the adjustments can proceed quite unevenly. Columbus never abandoned the conviction that he had reached the Indies, and almost until the end of his life continued to look for access to China and Japan; this despite his encountering populations that were not plausibly the periphery of the advanced manufacturing and trading economies that he sought.

exploration as models of autonomous agency, and for once the schoolchildren are being taught properly. Our philosophical theorizing should at least live up to what every schoolchild knows.

## 10.8

Recall that Bratman's Master Move appealed to what is the mainstream view of personal identity among analytic philosophers today: what makes you the same person you used to be is psychological continuity, typically glossed as remembering your past, acting on your former intentions, having a similar character, and so on. However, it should be obvious that identity concepts need to be tailored to species.<sup>50</sup> A butterfly is not psychologically continuous with the caterpillar it was. But it is still the same animal, because metamorphoses are what butterflies *do*: that's their *ergon*, or, in a more modern locution, metamorphoses are part of the design solution they implement.

We metamorphose psychologically: that's what it is to be a serial hyperspecializer and a segmented agent. In moving from niche to niche, you can quite correctly throw out your old goals, standards, preferences, intentions, and policies wholesale; you can also, quite correctly, throw out almost all of your memories. When this happens, you are still someone in whom your former self has a prudential stake (this being what really matters, in this philosophical subject area, about being the same person)—at any rate, provided that the transition to your new psychology was managed using the modes of rationality appropriate to segmented agency. For instance, when cognitive signals such as boredom and frustration prompt you to move on to a new niche, to forget your former life as thoroughly as possible, and to take up entirely new activities that you can find interesting, and in which you feel yourself competent and at home, that is a benefit to *you*.<sup>51</sup>

---

On the other hand, he very rapidly formulated and implemented colonial policies that—however repugnant—served as the basis for the Spanish colonial enterprise: policies that were effective against the background of the cultures he did encounter, and which it would have made no sense to adopt in the immediate vicinity of a great Asian power. For an overview, see Morison (1992, especially pp. 278f, 290f, 355f, 380f, 464–467, 553f).

Prescott (2000, p. 137) summarizes the initial scope of the mandate given to Cortés, and pp. 614f look back on one of history's most astounding cases of mission creep.

We should not, however, assume that exploration is always well-managed; again, it is not as though whatever cognitive mechanisms we have to guide us are infallible. In the literature of the age of exploration, *La relación* (Cabeza de Vaca, 1542/2002) is a suitable corrective.

<sup>50</sup>This is a Wiggins-like observation, but I wouldn't want to buy into all of the details of his treatment. For the latest revision, see Wiggins (2001).

<sup>51</sup>Instrumentalist theories of practical rationality are unable to account for this truism; see Williams (1973b, Chapter 6).

These days, sophisticated neo-Lockean accounts take time out to define psychological continuity in such a way that there may in fact be very little psychologically in common between two temporally widely separated stages of a person: you will count as psychologically continuous with your future

Neo-Lockean accounts of personal identity are just about right for a creature that, on the one hand, is like us in having a mental life, but on the other, is designed (or rather, “designed”) to live out its life in a single, stable niche. Such a creature can be imagined as starting life with a program that will govern its activities until the end of its days—a design approach that could be effective within the confines of a stable niche. The psychological states that the program deposits and uses—records in memory of the creature’s progress and the state of its environment, overarching goals, subgoals, and the like—may (and ought to be) updated constantly. But they will never need to be deleted wholesale, and if they *are* deleted wholesale, an individual creature of this kind will not normally survive the operation. So treating *such* a creature’s life as coextensive with a continuous psychology of this kind is entirely reasonable. We are not such creatures; a neo-Lockean account of personal identity is inappropriate for us.<sup>52</sup>

---

self if all of the adjacent pairs of intermediate stages share *some* of their psychology, and the relation that either is taken for or taken to replace personal identity is then introduced as the ancestral of that relation between adjacent stages. Put more concretely: you may remember nothing of your five-year-old self, but you are still that person because you remember some of your ten-year-old experiences, and your ten-year-old self remembered some of his five-year-old experiences; see, for instance, Parfit (1987, pp. 205f). That much psychological continuity will often be found in segmented agents; for instance, they may remember why they gave up on a previous niche.

But how is one to motivate the psychological continuity approach? Surely by way of the thought that the degree of connectedness is what matters, and even if it’s not *always* true that the more in the way of connections, the better, what matters in neo-Lockean personal identity would be getting short shrift in a human animal whose stages (*properly*, on some occasions) shared only a tiny handful of psychological states with their predecessor and successor stages. For segmented agents, it will often be the case that the *less* in the way of continuity, the better. The more efficient your garbage collection algorithms, the more effectively you can devote your cognitive resources to mastering your new niche.

<sup>52</sup>A less popular but still respectable position on personal identity prefers bodily continuity as the criterion of sameness—you are the very same person you once were if you have the same body—and it might seem that my account is committed to this alternative. I am not at all certain, for two reasons. First of all, I take seriously Bernard Williams’s arguments to the effect that we do not have a philosophically satisfactory account of the body, and that we have not thought through what our distinction between body and mind comes to (Williams, 1973b, Chapters 1–5, and especially pp. 11f, 68ff)—though, oddly enough, Williams himself went on to endorse a bodily-continuity account of identity, which strikes me as an uncharacteristic failure of philosophical nerve. “Body” (and, more recently, “organism”), in these discussions, is just a placeholder, a we-know-not-what. Second, we do not have an explanation for bodily continuity being the basis of an identity concept suitable for segmented agents, and, without that explanation, we should not just accept whatever looks like the leftover theory.

The other side of a philosophical theory of personal identity is a philosophical theory of death; after all, you are dead once there is no one who is identical to you. (For one expression of this insight, see Parfit 1987, pp. 281f.) I have been gesturing at an account of what it is to be a human being, one on which the fact that you are going to die is not an essential part of the design, but rather on the order of a manufacturing flaw. Death is a form of planned obsolescence suitable for creatures that—conformably to the mainstream model of agency—cannot reprogram themselves to adjust to deeply different environments. When a creature is self-reprogramming, throwing out the hardware platform at one-generation intervals looks *wasteful*.

That might suggest that something is deeply amiss in the account. (I’m grateful to Havi Carel for pressing me on this point.) It is a widely held view that your death is the frame in which the elements



Return to the rough-but-widely-shared picture of unified agency with which we began. That's a pretty good picture of a creature for whom a neo-Lockean theory of personal identity would be appropriate. But it's not at all a satisfactory rendering of a segmented agent. Segmented agents are not completely fragmented; we will see a great deal of top-down, policy-governed coordination of activity within each segment. But we should expect to see substantial agential disunity induced by fault lines between niches and by the cognitive devices that facilitate niche exploration and niche jumping. Like theorizing about personal identity, recent theoretical work on agency has been rich, subtle, and interestingly argued, but also philosophizing suited to a species not our own.

That disunity will be both diachronic and synchronic. I mentioned in passing that segmented agents will not infrequently pursue activities that go on in different niches. (The picture of segmented agency, recall, was a first approximation, and this is one important qualification to add to it.) Because the evaluative maps of these niches will typically amount to incommensurable systems of standards, the various activities in which segmented agents engage will often fail to be governed by a unified hierarchy of evaluative judgments and goals. They may even be at cross-purposes, and although a segmented agent is to some degree equipped to manage resource competition between projects and activities, we should not expect these to be regulated by higher level goals, desires, or standards. That would be to commit the error of thinking that an engineering solution which works within niches can be scaled up to solve problems that span niches, as though the world were simply a larger niche, and could be navigated by constructing a system of standards—or a policy—suitable to a much more constrained environment, only *bigger*.

I suggested that segmented agents must be cognitively equipped with prompts that tell them when it is time to switch niches, and that frustration and boredom are probably representative prompts of this sort. Let's conclude by considering what this means for the familiar philosophical interest in superlative attribution. When they respond to such prompts, segmented agents can surprise themselves by taking steps that contravene their current projects: steps they do not endorse, steps that do not ensue on their policies for weighing reasons, steps that lead them to say, "I didn't really choose to do it . . . it just kind of happened." When they act, they may not know what they are doing, or why. ("I don't know why I did that: I just *found* myself doing it.

---

of your life are meaningful, and that to lose sight of your death is to live "inauthentically." A suspicious reader might even wonder whether the present account is not just a way of avoiding the confrontation with one's own mortality. Now, certainly there is no point in pretending that one is not going to die. But in the segmented form of agency at which I have been gesturing, the frames that make activity meaningful may well be much shorter than an entire life. To face up to one's death does not mean: to think about it in terms that would be suitable only for the much simpler sort of animal implicitly presupposed by mainstream theory of agency.

I really surprised myself.”) And when they behave in this way, they may well be evincing, not their irrationality, and not their lack of self-government, as the theoretical mainstream has it, but what is precisely practical rationality and autonomy for creatures of this kind.

Mainstream theory of agency is very good at articulating the phenomenology of acting on the basis of an override—a signal that is independent of your system of standards, of your complex of goals and desires (in one famous bit of terminology, of your “subjective motivational set”), of your practical identity, and so on. (The accompanying utterances at which I just gestured are typical, and it can feel as though it wasn’t *you* who did it. Although of course you can often *explain* why you did it; for instance, you were bored out of your mind.) Because mainstream theorists identify the disposable personae you happen to be projecting at the moment with your *self*, they treat your responding to such signals as a disaster: as the unraveling of your agency, and thus of your personhood. But if you are a segmented agent, it is not necessarily a disaster at all; this is just how you cast off devices that are suitable, temporarily, for coping with one environment, in favor of other devices that are suitable for other environments. If you like, and granting, just for a moment and for the sake of the argument, that the mainstream is giving a successful analysis of what it is to be an agent, you are not a single agent, but rather the substrate of a series of them. Agents are interfaces you conjure up to meet the needs of the moment. Do not make the mistake of thinking that one or another of them is *you*.

## 10.9 Postscript

Michael Bratman has been kind enough to respond to the preceding essay, and I’d like to take the opportunity to follow up on his remarks. He writes:

[I]t does seem to me that if you are rationally self-governed and faced with such a fundamental change in the world—change that was “impossible-to-anticipate”—you would normally draw on certain more-or-less methodological policies, as in “stop, look, listen”; “be attuned to what bores or frustrates you”; or “rehearse alternatives in your imagination and pay close attention to how you respond,” and so on. This suggests that in many cases we can understand such revolutionary change as a special case of Neurath’s boat plan-theoretic self-governance where you primarily “stand on” such methodological policies.

Further, if you don’t draw on anything at all like such methodological policies, and yet you still end up with a kind of solution, it seems to me that your case will be more like a matter of grace or good luck than self-governed activity. This suggests that for a radical shift to be self-governed it needs to be grounded in *some* stable, plan-like commitment,

even if it is only a methodological commitment to treat certain kinds of “signals” as relevant. Once we take seriously the potential roles of such methodological policies, we should be skeptical about Millgram’s claim that “you can’t have a policy for handling the unanticipated.” . . . And the role of such methodological policies can help us distinguish self-governance in such extreme cases from something that is more like a conversion experience.<sup>53</sup>

I hope I’m not caricaturing his take on the argument by peeling out three thoughts that seem to be at work in it. First, you’ve got to be drawing on *something* when you respond to unexpected circumstances in a way that’s not just flailing. Whatever you draw on can be redescribed as a policy—and *is* a policy, because what *else* could it be? So you have policies for responding to unexpected circumstances.

Second, when that’s not the case, it’s time to be honest and admit that, sometimes, everyone thrashes. Faced with circumstances that aren’t covered by the policies they have reasonably, deliberately, intelligently, choicefully adopted, some people are plain and simple lucky, and when we say, after the fact, that they made thoughtful, self-controlled, autonomous decisions, we’re just complimenting success. In fact, there’s no discrimination to be made having to do with the nature of the responses themselves, even though some of them subsequently turn out to meet challenges and others come to merely being swept away by them, being oblivious to the demands they pose and so on.

Third, and here I need to repair an expository lapse of my own, when illustrations invoke the National Socialist regime of 1930s Germany, or abrupt transitions between life stages, it’s easy to come to think that the question at issue is whether situations can arise in which you have to drop *everything at once*. If a well-trained philosopher takes that to be a premise of the argument, he absolutely ought to wonder whether such cases are possible, and whether, even allowing they’re possible, they aren’t *rare*. If they don’t come up, we can ignore them, and if they come up only rarely, surely we can treat them as outliers.

Call the claim that one can meet up with novel circumstances that one has never had the chance to think through the *Little Claim*, and call the claim that situations can arise in which one has to drop everything at once the *Big Claim*. We don’t want to conflate them, but they are closely connected: if the Little Claim is true, then the Big-Claim cases can’t be dismissed as outliers; rather, they’re what you’d expect to see as you look at ever larger instances of the Little Claim, and they ought to be useful in displaying how *much* of a difference the ideas we’re developing make to how we should comport ourselves.

---

<sup>53</sup>Bratman (2014, p. 325).

It would be a mistake to dismiss the scaled-up version of the Little Claim as something you outgrow with childhood, or as a matter of rare encounters with mythologized evildoers. As I write this, it is becoming apparent that academia, anyway in the United States, is about to go through an abrupt transition, one that will plausibly be analogous to the deluge that has recently submerged the world of newspaper journalism. When it's over, it's likely that most instructors in higher education will have been replaced by web sites, and the jobs of those who remain will look very different, though it's hard now to anticipate exactly how. Everyone who works in higher education is rapidly moving toward a series of choices which it's difficult to imagine getting through on the basis of anything formally like the sort of policy that guided our imaginary department meeting (in Section 10.1). Tenured professors have been leading very sheltered lives, and as they confront this future and decide what to do, they had better get ready to make decisions very differently.

The assumption that whatever decisions they make will nonetheless be directed by policies raises the question of whether the concepts that Bratman has provided us are to be thought of as part of a theory, that is, as an organized body of assertions, or rather as a medium or rendering tool—conceptual paint, as it were, meant to allow philosophers (and perhaps nonphilosophers also) to say whatever it is they want to about a given range of topics (e.g., what moves people to act). This choice is one that only Bratman, as owner of the toolkit, can make. Once it's made, if talk of plans and policies excludes no substantive or contentful position (including under that heading methodological positions), then I'm happy to allow that whatever allows us to meet unanticipated circumstances will be a plan or policy. But if not *just anything* is a plan or a policy, then I see no reason to think that they are what will prove to be the resources we need when we are thinking outside of the box.

The choice doesn't have to be all one or all the other, and indeed I mentioned above that Bratman has used the vocabulary of plans and policies to reappropriate promising ideas from other philosophers while avoiding the pointilist effects induced by their descriptive apparatus of desires; when he did so, he was further developing his own substantive view. So is there a path between the horns of the dilemma? Certainly it's possible to draw on policies even when the execution of the policy does not explain everything one is doing. *Eureka!* moments are no doubt the upshot of a mixture of prior policies and luck, and perhaps the moment counts as your own, in the superlative sense we have been discussing, if the admixture of luck is not too great. Aristotle thought that when you put a drop of wine in a bucket of water, the wine *turns into* water; maybe when you put a drop of luck into policy-governed activity, the luck becomes *yours*.<sup>54</sup> But it would be an error to decide that

---

<sup>54</sup> Aristotle (1984, 328a25ff).

whenever there is a policy in the background, no matter how thin its contribution, we thereby have the sort of policy-governed activity that is Bratman's alternative to getting by just on dumb luck and triumphalism.<sup>55</sup>

We do owe an answer to the rhetorical "What *else* could it be?" that I read into Bratman's comments, and by way of seeing what an alternative *could* be, let's return to a small-scale example from one of the earlier footnotes.<sup>56</sup> Suppose you decide that it doesn't make sense to read Nietzsche as you read most philosophers—just for the theories. Suppose further that you have a backup metapolicy: to ask what the author is trying to *do*, where that means construing the text as a product of coherent, unified authorial decision-making.<sup>57</sup> Here the fallback approach (not all that atypically, when it comes to fallbacks) is also unmistakably inadequate to these particular texts, which indeed seems somehow to be their point: the author of Nietzsche's corpus is evidently not a unified agent and is unable to sustain coherent authorial intentions. Your policies for interpreting texts have run out, you need a different way of reading Nietzsche, and if you're like my own students, it's not like you have a metapolicy for adjusting your exegetical strategies. Rather, you were just taught one or two ways of approaching texts. Take it from me: my students do not come to class with a plan-like methodological commitment to treating their own responses to philosophical views as triggers for any sort of reconsideration; on the contrary, twelve years of primary and secondary education have trained them to ignore those signals and do what they're told. For purposes of talking through the example, let's suppose that's true of you, too.

In that case, you'll have to come up with a different way of reading Nietzsche's writings, one that isn't governed by a previously available Bratmanian policy. I hope I have conveyed that when my students decide they need a different way of reading Nietzsche's texts, they're being perceptive, and I actually

---

<sup>55</sup>Bernard Williams once made a structurally very similar error, which I spell out in Chapter 4, footnote 4, and here's an illustration of what it would look like to repeat the mistake, only with policies substituted for motivations. Robert Nozick once taught a course he called "Philosophical Heuristics." A member of Harvard's faculty, at a time when its philosophy department's all-star lineup included Nelson Goodman, W. V. Quine, Hilary Putnam, and so on—philosophers whose ideas had, one after another, changed the field—Nozick's program was to work through the writings in which they had proposed those ideas, develop plausible hypotheses as to the methods, techniques, and heuristics they had used to arrive at them, and subsequently have his subjects visit the classroom, to confirm or disconfirm those speculations in person.

The outcomes were unfortunately disappointing: it was easy to locate policies behind the revolutionary ideas, but they were thin: on the order of "Try denying shared assumptions." Even if one does draw on such policies, when field-changing ideas are the outcome, the policies make up too little of the explanation to allow us to credit the views to those philosophers on the basis of the policies. (Everyone has policies like these; why isn't everyone as original as Harvard's once-upon-a-time all-star lineup?) But of course Nozick didn't for a moment assume that *these* policies were the all-stars' secret sauce.

<sup>56</sup>See above, footnote 47.

<sup>57</sup>For a sophisticated presentation of such a policy, see Nehamas (1981), and Nehamas (1987); for a reading of Nietzsche in this vein, see Nehamas (1985). In the case of an author who is trying to present theories, that metapolicy endorses your initial first-level exegetical strategy.

mean that literally: they're *observing* that the policies they have just won't do.<sup>58</sup> And while what they come up with in the way of replacements is a bit uneven, it does look to me like they're doing much better than chance. As far as I can tell, they're managing that in part by trying out alternatives and *seeing* how well they do. And if you are going to do better than the analytic Nietzsche scholars who largely stick with the exegetical policies and metapolicies on which they were raised, you will also have to do what my students are doing.

Of course you can have policies—I mean, the sort of thing you decide on and then execute—meant to control, direct, or even just modulate experimentation and observation.<sup>59</sup> But the important point here is that perception can change your mind even when you *haven't* decided that it's going to. (You may have a policy, as Bratman suggests, of being attuned to what bores or frustrates you; but your frustration may well derail you, and correctly so, even when you have no such policy.<sup>60</sup>) When it does, that's not luck, and it's not grace, but rather the world catching your attention. *Observing* is not a policy; if perceiving what is going on around you were merely execution of policy, expectation really would determine perception—not just a little bit, but entirely. Whereas what makes observation so important is that you can come face to face with something that you have no room for in your understanding of the world, something that your policies overlook or tell you to ignore, and nonetheless face up to it.<sup>61</sup>

It is worth pointing out the similarities between the present back and forth and a disagreement about the place of observation in theoretical reasoning that occupied philosophers during the twentieth century. On the one hand, some of them insisted that experience must always involve subsumption under concepts; after all, how could unconceptualized experience initiate an inference? On the other hand, others pointed out that it is politically very important—and, I should add myself, if this is true, it is not just *politically*

---

<sup>58</sup>When Bratman suggests “conversion” as the way that luck and grace can figure into a successful readjustment, he is repeating a move made in the internalism debate some two decades back, and one that was, in my view, a misstep. At the time, John McDowell corrected Bernard Williams’s claim that a reason for doing something has to be tied by argument to antecedently available motivations, pointing out that someone could come to have reasons to do something by undergoing a conversion experience. Perhaps that can happen, but the appeal to arational transitions obscures the central way of coming to have reasons without having previously been in a position to argue your way to them, namely, being convinced by observation.

For internalism, see Williams (2001); for the correction, McDowell (1998, Chapter 5); Millgram (1996) registers the misstep.

<sup>59</sup>Recall Bratman’s short list of plausible suggestions. However, I don’t endorse that last proposed policy, the one focused on imagining responses to imagined alternatives. Why? See above, Section 3.4.

<sup>60</sup>Millgram (2004a).

<sup>61</sup>It is perhaps a Murdochian thought that when you do, you are resisting heteronomy; in making the effort to see things as they are, one sidesteps the urges and fantasies and, yes, the elaborately articulated policies that allow one to avoid governing oneself in demanding circumstances (Murdoch, 1970; Millgram, 2005a, Chapter 5).

important—that we can become aware that the concepts we have fail to do justice to our experience and that we need to think up new concepts to capture it.<sup>62</sup> Bratman looks to be assuming that unless the content of practical experience—say, the feeling of boredom or of interest—activates a policy you already have, any uptake that feeling gets can’t be properly attributable to *you*. But it seems to me to be very important that experience can overflow and escape our policies; we can become aware of that, and revise our policies, in very much the way we can fashion new conceptual tools to accommodate experiences we notice are slipping through our already present net of concepts—and when we do so, *we* do so.<sup>63</sup>

---

<sup>62</sup>See, e.g., McDowell (1994), MacKinnon (1989, Chapter 5), and Lovibond (2002).

<sup>63</sup>I’m grateful to Tom Blackburn, Luca Ferrero, Alisa Garcia, Karen Mottola, Tom Pink, and Margaret Stock for helpful discussion; to Chrisoula Andreou, Michael Bratman, Teresa Blankmeyer Burke, Sarah Buss, Christine Chwaszcza, Kelly Heuer, Jenann Ismael, Kimberly Johnston, Matt Mosdell, Stephanie Shiver, and Ben Wolfson for comments on earlier drafts; to audiences at Tel Aviv University, the University of Cologne, the University of Arizona, the University of Bern, the University of the Saarland, the University of Modena, and the University of Utah’s Kaffeeklatsch; to the participants of the Templeton Foundation Summer Retreat on Love and Human Agency; and, for reading through *Structures of Agency* with me, to my Fall 2008 seminar on Practical Reasoning.

## Afterword: A Call to Arms

Early on in this book, I argued that we face a crisis, which I called the *Great Endarkenment*. And over the course of the intervening chapters, I have been attempting to convince you that at the bottom of this crisis lie a series of specifically philosophical problems. So I want to conclude by considering what agendas the Endarkenment sets for philosophers. These will have to do both with content—what questions do philosophers need to take up?—and with institutional structure.

To speak to the question of what it will take for philosophers to address the problems we have identified, I must first face up to a skeptical objection. Is it reasonable to expect philosophers to find the problems posed by serial hyperspecialization tractable? While the exercises in cognitive-function analysis that we have seen are *some* evidence that this is something we know how to do, nonetheless a further doubt, which we should take very seriously, is motivated by the institutional structures of philosophy as we now have it: why think that philosophers are going to be able to resist the Endarkenment themselves? So I am going to start out by speaking for the skeptic and giving his complaint the best run for his money that I can.

### 11.1

Philosophers have in the past occasionally presented themselves as the Department of Interdisciplinary Studies, and there is in any case a longish history of philosophers offering to sort out and fix the foundational problems of the sciences.<sup>1</sup> Someone who inherits the view that this is the primary business of

---

<sup>1</sup>For instance, Dewey (1958, pp. 398, 403) famously introduces philosophy as “criticism,” “having its distinctive position among various modes of criticism in its generality; a criticism of criticisms,



philosophy might well think that “knowledge integration” is just the sort of job to justify philosophy department budgets.<sup>2</sup> Is he right?

If the problem is that no one can see the Big Picture, because no one can see how the pieces of the cross-disciplinary intellectual puzzle fit together, then to arrive at any single solution to the problem, you would have to see lots of Big Pictures. After all, no matter what technique or procedure you came up with to fit those pieces together, and no matter what story you came up with to make that technique or procedure out to be plausible, unless you were able to verify its outputs in a wide range of cases, you would have very little reason to trust it. And to verify its outputs—to see that it *does* work—is to see one or another cross-disciplinary Big Picture. So philosophers are going to be able to address this series of problems only if they’re *very* interdisciplinary. Are they?

On the contrary. It is not just that, for most of the twentieth century, philosophers were notable for their disciplinary narrowness; over the last decade or so, analytic philosophers have for the most part remade themselves into *subdisciplinary* specialists. Not that long ago, philosophers might have had a topic or subject on which they especially focused, but it was assumed that, modulo perhaps ancient languages, mathematical logic, and quantum field theory, any competent philosopher was competent at *all* of philosophy. The most prestigious journals in the field bore titles that reaffirmed the presumption of the availability of work in the discipline to any of its members.<sup>3</sup> In particular, it was expected that any philosopher would be able to read around in, make quality judgments of, and use in his own research the work of any other philosopher.

---

as it were”; he argues that “philosophy is and can be nothing but this critical operation and function become aware of itself and its implications, pursued deliberately and systematically.”

What especially makes necessary a generalized instrument of criticism, is the tendency of objects to seek rigid non-communicating compartments . . . a distribution of emphasis such as are designated by the adjectives scientific, industrial, political, religious, artistic, educational, moral and so on.

But however natural from the standpoint of causation may be the institutionalizing of these trends, their separation effects an isolation which is unnatural. (Dewey, 1958, p. 409)

His train of thought seems to run something like this (see, e.g., Dewey 2008b): Specialized fields arise as ways of handling classes of similar situations. But any concrete fact has indefinitely many effects. So the treatment given to a fact by a specialized field will overlook some of the effects of the fact. Some of those effects will turn out to be too important to be overlooked; they will turn out to disrupt other activities in which we’re engaged. So we need to deliberate about the problems that specialized methods have, and especially about the clashes that arise between them. Deliberating about the problems that specialized methods have is criticism of criticism, i.e., philosophy. And that is why we need philosophy.

Here the focus is on resolving clashes between specialized fields. Dewey seems less concerned with managing commerce between such fields, which is our own primary problem.

<sup>2</sup>The phrase is now the title of a department at the University of Waterloo.

<sup>3</sup>For instance, “*The Journal of Philosophy*”; “*The Philosophical Review*.”

Those days are long behind us. We have philosophers who work only on freedom of the will; philosophers who specialize in a particular historical figure; philosophers who are knowledgeable only about the philosophy of physics, or biology, or chemistry; philosophers who are no longer even, say, epistemologists, but who work only on the epistemology of testimony. They no longer read outside their own subspecialization; they go to conferences populated entirely by their own subspecialization; they publish only in specialty journals. In institutional settings, when they have to make decisions that turn on out-of-specialty assessments, these philosophers defer to referees, rather than relying on their own judgment. (For example, when they are making hiring decisions, they will rely on letters of recommendation rather than reading the candidates' work for themselves; they will accept the assessment of the letters over the evidence of their own eyes.) And philosophy is now a field most of whose members feel comfortable talking only to a handful of people who do the very narrow thing that they themselves do.

On the face of it, it appears that academic philosophy is undergoing the very fragmentation that gives rise to the problems we have identified and flagged as so very urgent. In particular, if specialization threatens autonomy, both intellectual and practical, philosophers should no longer be expected to think for themselves, even *within* philosophy. But if the philosophers cannot think for themselves, how are they going to retrieve for the rest of us the possibility of doing so from the pressures of hyperspecialization?

## 11.2

That dismal expectation seems to be confirmed by a remarkable recent icon of the abdication of intellectual autonomy within analytic philosophy. In the early 1990s, it came to the attention of American philosophers that someone was circulating a Top Ten list (actually, a Top Fifty list) of American philosophy departments. Brian Leiter, the author of the list, was at the time relatively unknown, and not regarded by anyone as an authority figure in the profession. What is more, I recollect that the almost uniform response to the list was that its judgments were highly idiosyncratic: no one—or hardly anyone—else agreed with them.

Twenty years on, Leiter has been for some time an authority figure in the profession, primarily on the basis of the list. Its judgments are no longer idiosyncratic. Amazingly, the list became a self-fulfilling prophecy: the ranking has been fully internalized, and across the country, philosophy departments make policy in meetings that start off by asking what the department can do to move up Leiter's list.<sup>4</sup>

---

<sup>4</sup>Richard Heck, in an online criticism of Leiter's list, has pointed this out, and I can confirm it on the basis of my own experience.

How did this happen? Part of the effect is likely to have to do with the list being presented as a way of choosing a graduate program. Philosophy undergraduates looking around for guidance in assessing providers of graduate education are unable to assess the merits of whatever ranking they encounter; thus the Leiter Report serves to anchor their judgments of philosophy programs and of kinds of philosophizing more generally. Eventually, these undergraduates grow up into assistant professors, into associate professors—and that first generation of Leiter-anchored students will shortly be full professors as well. As we know from the anchoring literature, normally one does not move far from one's initial anchor. However, this cannot be nearly all of the story: the shift in professional judgment took place very rapidly, thus among the already acculturated grownups as well.<sup>5</sup>

Although members of the community of philosophers who have internalized the Leiter rankings tend to insist that the assessments are widely shared because they are *correct*, and that the Leiter Report merely reflects an existing consensus within that community, the order of events makes that explanation implausible. There is now a biannual survey that is used to produce the Report; it no longer is simply an individual's Top Fifty list. However, that survey was added to the process only after the mainstreaming of Leiter's own judgments; by the time anyone thought of instituting a poll, the formerly idiosyncratic list had generated a community that shared them. Briefly: at its inception, Leiter's ranking was idiosyncratic; the later addition of a survey did not change the overall rankings significantly; therefore, the antecedently existing ranking must have shaped the survey results. There are various explanations at higher levels of granularity of how that might have happened, but, with one exception, we don't need to decide amongst them: we do need to exclude the possibility that this was a process of collective reevaluation on the basis of careful, closely reasoned argument; however, those of us who were around at the time will remember that that was not how it went. Thus the way in which the Leiter Report seems to reflect the professional sensibilities of a community should not be mistaken for any sort of independent confirmation: what is instructive is the manner in which those sensibilities arose, namely, as an effect of the Report.

---

<sup>5</sup>Is it possible that the prophecy was self-fulfilling not just in that it came to be accepted, but by bringing about the facts it purported to describe? On the one hand, because prospective graduate students see the list, higher-ranked departments have first pick of graduate student applicants. Because freshly minted PhDs see the list, higher-ranked departments have a much easier time recruiting junior faculty. Budgets are affected, albeit less directly. And so you would expect higher-ranked departments to *become* better departments, and lower-ranked departments, for the complementary reasons, to *become* worse departments.

On the other hand, institutions selected by such a ranking will tend to reward the quite specific talents that allow one to do its sort of philosophy well, and they will tend to shape the philosophical personalities of their students to favor its sort of philosophy. That need not be philosophy done well by prior lights.

I will call the general phenomenon *Leiterism*, and I'm going to review a few candidate explanations for Leiterism in philosophy. Then I will return to the question of whether philosophers are in a position to stave off the Great Endarkenment.

The first hypothesis is that, as analytic philosophy fractured into subdisciplines, it lived through the dynamics of the Great Endarkenment in just the way the culture as a whole did. As the sheer quantity of philosophical information grew, and as specialized techniques for manipulating it proliferated, analytic philosophers became progressively less able to assess the intellectual outputs of other areas of specialization within philosophy. As they coped with the deluge of information by subdividing the discipline into increasingly narrow areas of expertise, they found themselves delegating the larger and comparative assessments to an outside source.<sup>6</sup> So our first candidate explanation for Leiterism is that the philosophers did not override the new assessments with which they were being presented because that would have required solving the problem of intersubdisciplinary evaluation, for which the now subspecialized philosophers—like everyone else in our culture—have no formula or recipe or even worked-out theory.

I trust that we can dismiss this hypothesis out of hand. Although philosophy as a discipline has indeed fractured into subspecializations, what drives the institutional process must be very different from the ongoing causes of specialization in the surrounding society. In the sciences, for instance, there is simply too much to know for anyone to be able to assimilate it all; division of intellectual labor is inevitable. In philosophy, however, there is not all that much to know: it is not an information-rich field, and the necessary skills are largely shared by all of its practitioners.<sup>7</sup>

Proceeding now to the second hypothesis, recall my earlier suggestion that the ecological strategy of serial hyperspecialization is now the species form of *Homo sapiens*. Of course, organisms don't always manage to do everything that their overall strategy calls for, but asking what it does call for can nevertheless be a good guide to investigating how an organism gets by. So consider what would motivate serial hyperspecialization, seen as a design response to a challenging environment.

Serial hyperspecialization—adapting first to one ecological niche and then to another, and then to still another—makes sense because it's not efficient

---

<sup>6</sup>You will recall that the drivers for this sort of delegation, in the case of academic administrators, are surveyed in Chapter 2, Appendix A.

<sup>7</sup>This is not the conception that many younger analytic philosophers have of their discipline; their experience is that too much is constantly being published to keep up with it all. I will shortly explain the pace of publication, and in the meantime I suggest the following corrective to the misimpression that what those publications contain is *information*: make a list (that's right, go ahead and do it) of the uncontroversial results that philosophy has established.

Have you done the exercise? Surprise: you're working in a field in which there *are* no results.

to throw out your capital when you don't have to. Remember the simpler version of human beings, able to specialize only once in a lifetime, which I invited you to imagine in earlier chapters.<sup>8</sup> When such a creature's skills, desires, and so on no longer meet the demands its environment makes on it, it becomes unable to cope, and that is that. Serial hyperspecializers improve on such simpler specialists in that they economize by reusing the biological hardware: the trick is to load new software onto it. When there is less demand for electrical engineers, a serial hyperspecializer might become an expert in reconfiguring abandoned mines as thermal storage facilities.

But not throwing out your capital investment is just as good an idea when it comes to the software as it is for the hardware. Why squander the results of previous efforts? Why reinvent the wheel when you can just copy it? It would be good design to equip serial hyperspecializers with the wherewithal to identify successful strategies for getting by in a new niche and to mimic them. So you should expect to find serial hyperspecializers to be preadapted for appropriating their fellows' adaptations to different ecological niches—which isn't to say that the expectation can't be disappointed.

What sort of heuristics are serial hyperspecializers likely to use to locate and appropriate a successful strategy? As I've remarked, activities tailored to different ecological niches will typically incorporate structurally different standards and methods of assessment; these won't generally be accessible to a prospective newcomer, someone who is still trying to determine which up-front costs to pay to enter a given niche. So how is a diligent hyperspecializer supposed to tell whether another hyperspecializer, in another niche, is doing well and to be mimicked? If it applies the standards of its own niche, we can expect it to come up with an answer that is crude, or wrong, or just plain irrelevant. It does not yet control the standards of the prospective niche; the ability to deploy those standards is normally acquired in the course of adopting the niche-occupying strategy, and can't be relied on when it is looking in from the outside, or just dabbling. (Even if it could do both, how would it square the different answers?) And developing a system of standards on the basis of its own explorations is likely to be a painfully slow and resource-intensive process, one that recapitulates the efforts of the first colonizers of the niche.<sup>9</sup>

---

<sup>8</sup>See, for example, Chapters 3 and 4 in the present book.

<sup>9</sup>The mimicking hyperspecializer might use one or another signal to tell it when other specialists are doing well (or poorly) with their strategies. Bernard Williams once invoked the "ethological standard of the bright eye and the gleaming coat" (Williams, 1985, p. 46), but that is not the only such signal our implementation of the hyperspecializing strategy has, and I'm just going to gesture at another with a label: we notice and mimic what's *cool*. Another very prominent technique tracks social status in the prospective niche; here one is relying on others who are competent with the standards internal to the niche-occupying activity. And we also try to piggyback on others' judgments of success more generally. Fashion and celebrity are such prominent elements of human social life because they are side effects of a solution to problems posed by the hyperspecializing strategy.

Many of the ideas in this note are anticipated by Boyd and Richerson (1985), and Richerson and Boyd (2005).

One detour around this dilemma is also a shortcut: to rapidly identify *whatever* the prevalent standards in an already occupied niche are, and to adopt them, pretty much *regardless* of their merits. (At the outset, at the level of control available to novices; thus, not with the command of those standards that lends itself to critically reassessing them.) Assessing the standards and developing alternatives is an expensive option, and so it is one which will be deployed sparingly. If humans are serial hyperspecializers, we should not be surprised to find the disposition to take that fast-and-dirty approach in our own lives.

On our second candidate hypothesis, that's Leiterism. Human beings are serial hyperspecializers, and serial hyperspecializers should be expected to come equipped with a disposition to identify niche-specific systems of standards, and to adopt them, without being all that discriminating about whether the standards are any good. Here we have a system of standards for a niche (professional philosophy in the United States). The niche is one whose occupants claim to prize intellectual autonomy.<sup>10</sup> At the inception of the Leiter Report, a great many philosophers thought that the standards were out of line with their own. Nonetheless, it was a visible system of standards, and the only visible system of standards.<sup>11</sup> So the philosophers collectively and severally adopted the standards. It is as though we were built to implement the version of the design I was just describing, and the built-in tendency is powerful enough for an ideology of intellectual autonomy to be no match for it at all.

Things don't look so good for the philosophers. They haven't been immune to the very sort of specialization that puts thinking for oneself out of bounds. How much faith should we have that philosophy can solve the problems of the Great Endarkenment, when philosophers are visibly unable to solve their scaled-down version of those very problems for themselves? We tell physicians to heal themselves; before we turn to philosophy, won't we insist that the philosophers fix their own dysfunctional discipline first? But more pointedly, in their own professional regulation they have adopted practices that amount to giving up on thinking for themselves. If our second candidate explanation of

---

<sup>10</sup> An indication of the entrenchment of that priority is the treatment given in philosophy to arguments from authority. In other disciplines, it is fine to cite someone else as an authority in support of your view. In philosophy, it isn't; you normally cite someone as an example of the view you intend to attack. And philosophy students are taught that arguments from authority are always a fallacy.

<sup>11</sup> Visibility is important. There had of course been previous professional standards and rankings, but they were largely tacit, and you had to have been a member of the professional elite to know what they were.

If I'm right about how we arrived at our current institutional situation, it's an interesting exercise to imagine how the process would have played out had a different system of rankings been deployed: say, one that reflected the former tacit standards, or alternatively, one expressing the view I have been exploring in this book, on which philosophy done right is mostly, as I have been putting it, intellectual ergonomics.

Leiterism is correct, how likely is it that restoring the possibilities of practical and intellectual autonomy, and making thinking for oneself once again a live ideal, is going to be the accomplishment of specialists who, within their own discipline, have surrendered to the Endarkenment?

### 11.3

There is, however, a third hypothesis that could account for the predicament in which we now find ourselves. And if it is correct, the prospects for a philosophical response to the Great Endarkenment are not nearly as bad as they might seem.

Recall the need for the institutions that employ philosophers and other academics to assess the work being done for them. For reasons we have discussed in Appendix A to Chapter 2, academic institutions are unable to assess their employees' performance directly themselves, and so they look for proxies that they can see and, more importantly, count. I suppose it needn't have turned out this way, but academic "productivity" is now benchmarked by counting publications in professional venues. Since institutions provide performance incentives for their staff, the effect has been to put a premium on publishing, and it is now routine for philosophy graduate students to be urged to publish as soon as possible. And of course any professor who hopes ever to get a raise publishes as much as he can. No human being could keep up with the tidal wave of books and journal articles that was the inevitable consequence of this incentive structure, yet academics are supposed to stay on top of the literature in their field.

Keeping one's head above water, in circumstances like these, requires finding a legitimate way of exempting oneself from reading everything. An obvious way to do this is to define a very small subspecialization, one that has, ideally, a population of twenty or so, and make it one's own. So subspecialization within philosophy is partly driven by the need to keep one's reading load manageable. However, publications, when they are generated for these sorts of reasons, don't imply that there was anything that needed to be said in print. The recent wave of specialization within philosophy was not a way of coping, as in the sciences, with a surfeit of information, but rather a response to a perverse institutional incentive structure.

It wasn't all about reading load; those universities also rely heavily on letters of recommendation and similar assessments provided by specialists outside the institution. Because the institutions cannot make sense of nuanced or contentful assessments—recall that those require field-specific expertise, and the problem is that at the level of an academic institution's administration, the requisite expertise is not available—the assessments are read for, roughly, how much work was involved in producing them, as well as for a sense of enthusiastic praise. The upshot is that letters of recommendation and

similar professional evaluations have become ever more labor intensive. How can an aspiring junior faculty member assure that when tenure time rolls around, whoever his university asks for an assessment will be willing to go to bat for him?

Once again, the solution is to create a (very small) subspecialty.<sup>12</sup> If you choose to become one of the twenty people who work on Whatever Studies, if you go to an annual conference in your field, if you edit an anthology in which you publish all of your cospecialists, and if you favorably review any article or book manuscript in your subfield, you can put yourself in a position where anyone your university can legitimately ask for an assessment is a known quantity, and indeed, someone both beholden to you and against whom you can retaliate if he fails to reciprocate.

People generally, and academic philosophers in particular, respond to the incentive structures they are given. And so professional philosophy in the English speaking world has become a landscape of very, very small subspecializations. But, to reiterate, the reasons this happened have very little to do with the content of philosophy, or with what it takes to think through a philosophical problem successfully. This suggests that it is an institutional accident that philosophy as a field has undergone the same sort of articulation into very specialized subfields as the surrounding society. Philosophers who can extricate themselves from this incentive structure—perhaps by finding nonacademic employers—should in principle still be able to think for themselves, at least as long as the trains of thought they are developing are internal to philosophy, and do not rely overmuch on outside expertise. The situation in philosophy is bad, but it doesn't *have* to be as bad as it is; in a somewhat different institutional environment, in which the underlying problems played out slightly differently, the philosophers, at any rate, would be able to think through new tools for managing the flow of information across the boundaries of niches populated by serial hyperspecializers.

## 11.4

I haven't given a direct argument against that second hypothesis. But it seems to me that the best way to find out if it is true—or to what extent it is true—is to act on the optimistic alternative and see how it plays out. Furthermore, if

---

<sup>12</sup>Of course, philosophy is not the only discipline to experience these sorts of pressures (Culligan and Peña-Mora, 2010, p. 153).

I entirely agree with Frodeman (2014, pp. 87f, 92, 96) that analytic philosophers have failed to understand their own disciplinary institutionalization to be a philosophical problem and that uncritically accepting the notion that philosophy is a profession amounts to an unphilosophical dogmatism. On the other hand, his own proposal, that philosophy should be “dedisciplined,” disregards, among other things, its distinctive modes of argumentation, the enormous importance for successful philosophizing of mastering the philosophical tradition, and the cohesiveness of that tradition's problem set.



heteronomy is deeply a part of our species form, that would be something I am not sure we know how to fix. So we might as well start off by construing our problem so that we are able to try addressing it now. So, proceeding on the assumption that the grip of the Great Endarkenment within philosophy is at least to a large extent the effect of perverse institutional structures which we can reasonably hope to change, what changes do we need to make, both to loosen that grip locally and to start ourselves off on a viable approach to the problems of serial hyperspecialization?

Let's first point out a mistake made by our predecessors. Analytic philosophers of science, especially, have long struggled with a special case of the problem centrally underlying the Great Endarkenment. Scientists produce theories with very different vocabularies, very different ontologies, and very different bodies of associated technique. All of these theories describe the same world. And so philosophers committed to the unity of science repeatedly attempted to show how to convert all those apparently heterogenous theories into a single theory; indeed, descendants of that approach can still be found in the current literature.<sup>13</sup>

We can pinpoint their misstep from the perspective of the present project. While the problem of transferring information across the boundaries of the different sciences is very important, and while it is one to which philosophers have important contributions to make, it is a mistake to assume that the problems of communication within the scientific enterprise are to be solved by embedding the various sciences in a uniform and homogeneous *theory*. There is both an insuperable obstacle to that enterprise, and a misapprehension built into it. Theories are held together by inferences, and so a theory presupposes that a single logic—a single view of what counts as correct inference—spans that theory in its entirety. I have observed that different specializations (and different sciences, in particular) typically develop proprietary views as to what counts as a good argument: put dramatically, scientists in different fields are logical aliens with respect to one another. Thus their respective theories cannot be unified into a shared theory. More importantly, the notion that science is to be unified by producing a single theory to serve as a shared point of reference conflates a family of practical problems—of how-to problems—with a very elaborate question of fact.

---

<sup>13</sup>For instance, Ladyman and Ross (2007) has it that the point of doing metaphysics is to show science to be unified by embedding the theories of the various sciences within, or against the background of, a unified metaphysical theory. For what it is worth, however, the historical track record suggests that their program is a dead end. When it is successful, unification of this kind has almost always been managed by the scientists themselves, and not by philosophical onlookers. Compare, for instance, the discovery of the conservation of energy, which was part of the conversion of mechanics, electromagnetism, and a number of other distinct fields of enquiry into the science we now call “physics,” with the sort of unifying theories that metaphysicians have produced—for example, the insistence, by these particular authors, that everything is patterns.

I recommend taking as our shared point of reference a repurposed and re-focused philosophy of logic. Once we acknowledge that, from the perspective of anyone in our society, many of the other members of the society on whom we have to rely are logical aliens, philosophy of logic can be directed to a handful of related problems; these have appeared repeatedly in the preceding chapters, and by now should come as no surprise. We need ways of assessing the various modes of argumentation that have taken root in different disciplinary specializations. We need ways of assessing arguments constructed in an alien logic; since it is hard to believe that, except in cases where a logic can be made completely formal, one can learn to do this as it were remotely, this will be largely a matter of developing principled techniques for successfully identifying appropriate expertise. We need ways of managing the interfaces between disciplines; under this heading, we need techniques appropriate to reasoning in which one is unable to reassess and revise one's premises for oneself. We need ways of managing arguments that traverse disciplines, and in particular and especially, we need to figure out techniques for monitoring and controlling the quality of defeasible argumentation.

I have tried to make it plausible through my illustrations that the practice of this sort of philosophy of logic ought to emphasize cognitive function analyses of available logical devices, and the design of improved replacements for them. Although such a practice diverges from the day-to-day routine of today's analytic philosophers, while constructing my illustrations I pointedly drew on our shared training and tradition. I hope to have persuaded readers that my recommendation means putting available skills and an extant resource base to new uses.

## 11.5

That said, the way we currently train philosophers nonetheless needs to be re-organized thoroughly, if we are going to pursue the problems which I have been claiming have to be at the very top of our collective agenda. I earlier observed that no one is in a position to develop a single top-down response to the Great Endarkenment, because the problem is precisely that no one can be competent in all disciplines. That means that solutions must be developed bottom up and relatively locally: by figuring out techniques to manage the transfer of information and advice across the interface between two particular disciplinary specializations . . . and in parallel, figuring out techniques for the analogous problems facing another pair of disciplines . . . and only after a great deal of this, attempting to extract from such already implemented techniques a generally characterizable common denominator. This approach does not presuppose superhuman abilities: a philosopher, besides coming up to speed in philosophy, can be trained in another specialized field up to the point where

he understands its representational vocabulary, its modes of argumentation, and its other standards of success; this is already common enough in such specializations as philosophy of biology.<sup>14</sup>

For every alien logic, some philosophers can learn to be at home in it. Specialization in philosophy is in any case artificial. Philosophers are in any case made more aware of argumentation as itself a topic of investigation and inquiry than the practitioners of just about any other discipline. Moreover, we can collectively make an emphasis on philosophy of logic as I have been reimagining it common ground among philosophers, by turning it into a central and explicitly required component of graduate training. So philosophers can communicate with each other, and whether they are trained as they now are, or as I am proposing they get trained in the future, they can be good at talking with each other about how specialists in their respective secondary disciplines run their arguments. And so we can expect such philosophers to do as well as anyone at figuring out how pairs of specialists who are logical aliens with respect to one another can manage shared argumentation.<sup>15</sup>

Philosophers will only be able to effect improvements in interface management if they can explain what they are doing—their proposals, the rationales for them, and so on—to their clients. So part of a philosopher's job is writing intelligibly for outsiders, in something like the way that part of an agronomist's job is knowing how to talk to farmers. Moreover, while I am skeptical about the general claim that experts must be able to explain themselves to nonexperts, it is part of the disciplinary practice of philosophers in particular to shoulder this burden. This is not the time or place to advance an account of philosophy that would show that aspect of the practice to be not merely accidental, and I'll suffice with a reminder as to what the practice is. If you are teaching an introductory physics class and a student doesn't understand, that's *his* problem. But if you are teaching an introductory philosophy class and a student doesn't understand, that's *your* problem: especially if the student is bright, and if the failure to understand can't be cleared up, the failure itself constitutes an objection to the view you were presenting. Let's not forget that the logical positivists, that is, the founders of our very own philosophical tradition, treated their own pointed failure to understand their predecessors as the decisive refutation of the predecessors' views.

---

<sup>14</sup>However, we cannot afford to neglect the issue of institutional incentives. As things stand, philosophers with such training are rewarded for interacting only with philosophers who have the very *same* training, rather than with philosophers who know about anything else. (For example, philosophers of biology do best for themselves, professionally, by publishing in philosophy of biology journals, by going to philosophy of biology conferences, and, in general, by talking to other philosophers of biology.) If making headway on the problems of the Endarkenment requires having philosophers with different training collaborating with one another, these incentives will have to change.

<sup>15</sup>A helpful example of the sort of thing I have in mind: Tabery (2014) diagnoses and recommends a fix for the crosstalk between two different methodological traditions within biology.

Philosophers have the mechanics of argumentation at the center of their subject matter. So they are equipped to make their own arguments explicit to others. And if philosophers are further focused on the way their non-philosophical clients' arguments work, they are doubly so equipped. And so over the course of this volume I have been trying to live up to these responsibilities (partly so that you will be convinced, in retrospect, that it can be done). I am aware that some of the chapters have been harder for outsiders to follow than others—in particular, those in which I have been trying to demonstrate that what seem to me to be misguided philosophical enterprises fail in their own terms. Nonetheless, even where some readers have had to proceed more slowly, I have tried throughout to keep the discussion accessible to an intelligent general audience.

Perhaps eventually an overall Big Picture will emerge—and perhaps not: Hegel thought that the Owl of Minerva would take wing only at dusk (i.e., that we will only achieve understanding in retrospect, after it's all over), but maybe the Owl's wings have been broken by hyperspecialization, and it will never take to the air at all. What we can reasonably anticipate in the short term is a patchwork of inference management techniques, along with intellectual devices constructed to support them. One final observation: in the Introduction, I gave a number of reasons for thinking that our response to the Great Endarkenment is something that we can start working on now, but that it would be a mistake at this point to try to produce a magic bullet meant to fix its problems. That turns out to be correct for yet a further reason. Because the approach has to be bottom-up and piecemeal, at present we have to suffice with characterizing the problem and with taking first steps; we couldn't possibly be in a position to know what the right answers are.

Thus far our institutional manifesto. Analytic philosophy has bequeathed to us a set of highly refined skills. The analytic tradition is visibly at the end of its run. But those skills can now be redirected and put in the service of a new philosophical agenda. In order for this to take place, we will have to reshape our philosophical pedagogy—and, very importantly, the institutions that currently have such a distorting effect on the work of the philosophers who live inside them. However, as many observers have noticed, academia is on the verge of a period of great institutional fluidity, and flux of this kind is an opportunity to introduce new procedures and incentives. We had better take full advantage of it.<sup>16</sup>

---

<sup>16</sup>I'm grateful to Chrisoula Andreou and Jim Tabery for comments on an earlier draft of this Conclusion.



## BIBLIOGRAPHY

- Ainslie, G., 1992. *Picoeconomics*. Cambridge University Press, Cambridge.
- Albee, E., 1957. *A History of English Utilitarianism*. George Allen and Unwin, London.
- Anderson, R. L., 2004. It adds up after all: Kant's philosophy of arithmetic in light of the traditional logic. *Philosophy and Phenomenological Research*, 69(3), 501–540.
- Andreou, C., 2006. Might intentions be the only source of practical imperatives? *Ethical Theory and Moral Practice*, 9, 311–325.
- Anscombe, G. E. M., 1997. Modern moral philosophy. In Crisp, R., and Slote, M., editors, *Virtue Ethics*, pages 26–44. Oxford University Press, Oxford.
- Aristotle, 1984. *Complete Works*. Princeton University Press, Princeton, NJ. Edited by Jonathan Barnes.
- Arora, S., Barak, B., Brunnermeier, M., and Ge, R., 2011. Computational complexity and information asymmetry in financial products. *Communications of the ACM*, 54(5), 101–107.
- Arthur, W., 1987. *The Niche in Competition and Evolution*. John Wiley and Sons, New York.
- Asimov, I., 2001. Nightfall. In *The Complete Stories*, pages 334–362. Broadway Books, New York.
- Auerbach, E., 2003. *Mimesis*, 2nd edition. Princeton University Press, Princeton, NJ. Translated by Willard R. Trask.
- Austin, J. L., 1962. *Sense and Sensibilia*. Oxford University Press, Oxford. Edited by G. J. Warnock.
- Axelrod, R., 1984. *The Evolution of Cooperation*. Basic Books, New York.
- Ayer, A. J., 1951. *Language, Truth and Logic*. Victor Gollancz, London.
- Bain, A., 1882. *John Stuart Mill: A Criticism with Personal Recollections*. Longman, Green, and Co., London.
- Bastiat, F., 1995. What is seen and what is not seen. In *Selected Essays on Political Economy*, pages 1–50. Foundation for Economic Education, Irvington-on-Hudson, NY. Translated by Seymour Cain. Edited by G. de Huszar.
- Blackburn, S., 1998. *Ruling Passions*. Oxford University Press, Oxford.
- Borges, J. L., 1999. The library of Babel. In *Collected Fictions*, pages 112–118. Penguin, New York. Translated by Andrew Hurley.
- Bowman, M., 2012. *Are Our Goals Really What We're After?* PhD thesis, University of Utah.
- Boyd, R., 1988. How to be a moral realist. In Sayre-McCord, G., editor, *Essays on Moral Realism*, pages 181–228. Cornell University Press, Ithaca, NY.
- Boyd, R., and Richerson, P., 1985. *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.
- Brandom, R., 1994. *Making It Explicit*. Harvard University Press, Cambridge, MA.
- Brandom, R., 2001. Action, norms, and practical reasoning. In Millgram, E., editor, *Varieties of Practical Reasoning*. MIT Press, Cambridge, MA.
- Bratman, M., 1987. *Intention, Plans and Practical Reason*. Harvard University Press, Cambridge, MA.

- Bratman, M., 1999. *Faces of Intention*. Cambridge University Press, Cambridge.
- Bratman, M., 2001. Taking plans seriously. In Millgram, E., editor, *Varieties of Practical Reasoning*. MIT Press, Cambridge, MA.
- Bratman, M., 2007. *Structures of Agency*. Oxford University Press, Oxford.
- Bratman, M., 2009. Intention, practical rationality, and self-governance. *Ethics*, 119(3), 411–443.
- Bratman, M., 2014. Rational and social agency: Reflections and replies. In Vargas, M., and Yaffe, G., editors, *Rational and Social Agency*, pages 294–343. Oxford University Press, New York.
- Bremer, M., 2003. Is there an analytic limit of genuine modal realism? *Mind*, 112(445), 79–82.
- Brink, D., 1989. *Moral Realism and the Foundations of Ethics*. Cambridge University Press, Cambridge.
- Broome, J., 2013. *Rationality through Reasoning*. Wiley-Blackwell, Oxford.
- Cabeza de Vaca, A. N., 1542/2002. *Chronicle of the Narváez Expedition*. Penguin, New York. Translated by Fanny Bandelier, revised by Harold Augenbraum.
- Carroll, L., 1895. What the Tortoise said to Achilles. *Mind (New Series)*, 4(14), 278–280.
- Chang, R., 1997. *Incommensurability, Incomparability, and Practical Reason*. Harvard University Press, Cambridge.
- Chesnow, Z., 2008. *Kol ha-Neharot Zormim la-Yarden*. Philobiblion, Jerusalem.
- Chihara, C., 1998. *The Worlds of Possibility*. Oxford University Press, Oxford.
- Chisholm, R., 1948. The problem of empiricism. *Journal of Philosophy*, 45(19), 512–517.
- Christman, J., editor, 1989. *The Inner Citadel*. Oxford University Press, New York.
- Clark, D., Ranney, M., and Felipe, J., 2013. Knowledge helps: Mechanistic information and numeric evidence as cognitive levers to overcome stasis and build public consensus on climate change. In Knauff, M., Pauen, M., Sebanz, N., and Wachsmuth, I., editors, *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, pages 2070–2075. Cognitive Science Society, Austin, TX.
- Converse, P., 1964. The nature of belief systems in mass publics. In Apter, D., editor, *Ideology and Discontent*, pages 206–261. The Free Press of Glencoe, New York.
- Coulton, J. 2011. Still alive. *Artificial Heart*. Jocoserious Records. Sung by Sara Quin.
- Craig, E., 1990. *Knowledge and the State of Nature*. Clarendon Press, Oxford.
- Crary, A., 2007. *Beyond Moral Judgment*. Harvard University Press, Cambridge, MA.
- Culligan, P., and Peña-Mora, F., 2010. Engineering. In Frodeman, R., Klein, J. T., and Mitcham, C., editors, *The Oxford Handbook of Interdisciplinarity*, pages 147–160. Oxford University Press, Oxford.
- Dancy, J., 1993. *Moral Reasons*. Blackwell, Oxford.
- Davidson, D., 1984. *Inquiries into Truth and Interpretation*. Clarendon Press, Oxford.
- de Groot, A. D., 1965. *Thought and Choice in Chess*. Mouton and Co., The Hague.
- Debats, M., Cluzard, J., and Perrin, J. 2001. Winged migration. Produced by Christophe Barratier and Jacques Perrin. Columbia Tristar Hom.
- Dennett, D., 1991. *Consciousness Explained*. Little, Brown, Boston.
- Dewey, J., 1958. *Experience and Nature*. Dover, New York.
- Dewey, J., 2008a. *The Quest for Certainty*. Southern Illinois University Press, Carbondale, IL. Edited by J. A. Boydston.
- Dewey, J., 2008b. Theory of valuation. In Boydston, J. A., and Levine, B., editors, *John Dewey: The Later Works, 1925–1953*, vol. 13, pages 191–251. Southern Illinois University Press, Carbondale, IL.

- Diamond, C., 1988. The dog that gave himself the moral law. *Midwest Studies in Philosophy*, 13, 161–179.
- Diamond, C., 1991. *The Realistic Spirit*. MIT Press, Cambridge, MA.
- Divers, J., 1997. The analysis of possibility and the possibility of analysis. *Proceedings of the Aristotelian Society*, 97, 141–160.
- Divers, J., 1999. A genuine realist theory of advanced modalizing. *Mind*, 108(430), 217–240.
- Divers, J., 2002. *Possible Worlds*. Routledge, London.
- Divers, J., and Melia, J., 2002. The analytic limit of genuine modal realism. *Mind*, 111(441), 15–36.
- Divers, J., and Melia, J., 2003. Genuine modal realism limited. *Mind*, 112(445), 83–86.
- Duden, B., 1991. *The Woman Beneath the Skin*. Harvard University Press, Cambridge, MA. Translated by Thomas Dunlap.
- Duhem, P., 1985. *To Save the Phenomena*. University of Chicago Press, Chicago. Translated by Edmund Soland and Chaninah Maschler.
- Dummett, M., 1978. Wang's paradox. In *Truth and Other Enigmas*, pages 248–268. Harvard University Press, Cambridge, MA.
- Dworkin, G., 1988. *The Theory and Practice of Autonomy*. Cambridge University Press, New York.
- Dworkin, R., 1996. Objectivity and truth: You'd better believe it. *Philosophy and Public Affairs*, 25(2), 87–139.
- Enoch, D., 2005. Why idealize? *Ethics*, 115, 759–787.
- Enoch, D., 2006. Agency, shmagency: Why normativity won't come from what is constitutive of action. *Philosophical Review*, 115(2), 169–198.
- Fara, M., and Williamson, T., 2005. Counterparts and actuality. *Mind*, 114(453), 1–30.
- Ferrero, L., 2009. Constitutivism and the inescapability of agency. *Oxford Studies in Metaethics*, 4, 303–333.
- Ferrero, L., 2010. Decisions, diachronic autonomy, and the division of deliberative labor. *Philosophers' Imprint*, 10(2), 1–23.
- Fine, K., 1975. Critical notice: D. Lewis, *Counterfactuals*. *Mind*, 84(335), 451–458.
- Fisch, M., 1991. *William Whewell: Philosopher of Science*. Oxford University Press, Oxford.
- Fisch, M., and Benbaji, Y., 2011. *The View from Within*. University of Notre Dame Press, Notre Dame, IN.
- Fleischacker, S., 1999. *A Third Concept of Liberty*. Princeton University Press, Princeton, NJ.
- Fleischacker, S., 2004. Review of Bernard Williams, *Truth and Truthfulness*. *Ethics*, 114(2), 380–385.
- Fleischacker, S., 2013. *What Is Enlightenment?* Routledge, London.
- Forbes, G., 1985. *The Metaphysics of Modality*. Clarendon Press, Oxford.
- Foucault, M., 1984. What is an author? In Rabinow, P., editor, *The Foucault Reader*, pages 101–120. Pantheon, New York.
- Foucault, M., 1988. *Madness and Civilization: A History of Insanity in the Age of Reason*. Random House, New York. Translated by Richard Howard.
- Foucault, M., 1995. *Discipline and Punish: The Birth of the Prison*. Random House, New York. Translated by Alan Sheridan.
- Frankfurt, H., 1988. *The Importance of What We Care About*. Cambridge University Press, Cambridge.
- Frankfurt, H., 1999. *Necessity, Volition, and Love*. Cambridge University Press, Cambridge.



- Frede, M., 1987. Individuals in Aristotle. In *Essays in Ancient Philosophy*, pages 49–71. University of Minnesota Press, Minneapolis, MN.
- Frege, G., 1978. *The Foundations of Arithmetic*, 2nd revised edition. Basil Blackwell, Oxford. Translated by J. L. Austin.
- Friedman, M., 2001. *Dynamics of Reason*. CSLI Publications, Stanford, CA.
- Frodeman, R., 2014. *Sustainable Knowledge: A Theory of Interdisciplinarity*. Palgrave Macmillan, New York.
- Furth, M., 1998. *Substance, Form and Psyche: An Aristotelean Metaphysics*. Cambridge University Press, Cambridge.
- Galison, P., 1997. *Image and Logic*. Chicago University Press, Chicago.
- Gasking, D., 1953. Mathematics and the world. In Flew, A., editor, *Essays in Logic and Language*, second series. Blackwell, Oxford.
- Geertz, C., 1973. *The Interpretation of Cultures*. Basic Books, New York.
- Gendler, T. S., 2000. The puzzle of imaginative resistance. *Journal of Philosophy*, 97(2), 55–81.
- Gentner, D., and Rattermann, M. J., 1991. Language and the career of similarity. In Gelman, S. A., and Byrnes, J. P., editors, *Perspectives on Language and Thought: Interrelations in Development*, pages 225–277. Cambridge University Press, Cambridge.
- Gerber, S., Mayerik, V., Colan, G., Brunner, F., Buscema, J., and Infantino, C., 2008. *Howard The Duck Omnibus*. Marvel Comics, New York.
- Gibbard, A., 1990. *Wise Choices, Apt Feelings: A Theory of Normative Judgment*. Harvard University Press, Cambridge, MA.
- Gibbard, A., 2003. *Thinking How to Live*. Harvard University Press, Cambridge, MA.
- Ginsberg, B., 2011. *The Fall of the Faculty: The Rise of the All-Administrative University and Why It Matters*. Oxford University Press, Oxford.
- Gombrich, E. H., 1984. *Art and Illusion*. Princeton University Press, Princeton, NJ.
- Gordon, D., 1986. The dynamics of the daily round of the harvester ant colony (*Pogonomyrmex barbatus*). *Animal Behavior*, 34, 1402–1419.
- Gordon, D., 1989. Caste and change in social insects. In Harvey, P., and Partridge, L., editors, *Oxford Surveys in Evolutionary Biology*, vol. 6, pages 55–72. Oxford University Press, Oxford.
- Grice, P., 1975. Method in philosophical psychology (from the banal to the bizarre). *Proceedings and Addresses of the American Philosophical Association*, 48, 23–53.
- Grover, D., 1992. *A Prosentential Theory of Truth*. Princeton University Press, Princeton, NJ.
- Hacking, I., 1975. *The Emergence of Probability*. Cambridge University Press, Cambridge.
- Hacking, I., 1995. *Rewriting the Soul*. Princeton University Press, Princeton, NJ.
- Hamblin, C. L., 1958. Questions. *Australasian Journal of Philosophy*, 36(3), 159–168.
- Hardwig, J., 1985. Epistemic dependence. *Journal of Philosophy*, 82(7), 335–349.
- Hardwig, J., 1988. Evidence, testimony, and the problem of individualism—a response to Schmitt. *Social Epistemology*, 2(4), 309–321.
- Hare, R. M., 1961. *The Language of Morals*. Clarendon Press, Oxford.
- Haukioja, J., 2006. Why the new missing explanation argument fails, too. *Erkenntnis*, 64, 169–175.
- Hayek, F. A., 1948. *Individualism and Economic Order*. University of Chicago Press, Chicago.
- Hayek, F. A., 1973. *Law, Legislation and Liberty*, vol. I. University of Chicago Press, Chicago.

- Hayek, F. A., 1994. *The Road to Serfdom*. University of Chicago Press, Chicago.
- Hearne, V., 1986. *Adam's Task: Calling Animals by Name*. Knopf, New York.
- Herman, B., 1993. *The Practice of Moral Judgment*. Harvard University Press, Cambridge, MA.
- Hill, T., 1991. *Autonomy and Self-Respect*. Cambridge University Press, New York.
- Hilton, B., 2006. *A Mad, Bad, and Dangerous People? England 1783–1846*. Oxford University Press, Oxford.
- Hmura, M., 1996. *Mountain View Ranch*. Leaning Vine, Lockport, IL.
- Hölldobler, B., and Wilson, E. O., 2009. *The Superorganism*. W. W. Norton, New York.
- Hooker, B. and Little, M., 2000. *Moral Particularism*. Clarendon Press, Oxford.
- Housman, A. E., 1965. *Collected Poems*. Holt, Rinehart and Winston, New York.
- Huemer, M., 2005. *Ethical Intuitionism*. Palgrave Macmillan, New York.
- Hughes, G. E., and Cresswell, M. J., 1996. *A New Introduction to Modal Logic*. Routledge, New York.
- Hume, D., 1888/1978. *A Treatise of Human Nature*, 2nd edition. Clarendon Press, Oxford. Edited by L. A. Selby-Bigge and P. H. Nidditch.
- Hurley, S., 1989. *Natural Reasons: Personality and Polity*. Oxford University Press, Oxford.
- Hussain, N., 2012. A problem for ambitious metanormative constructivism. In Lenman, J. and Shemmer, Y., *Constructivism in Practical Philosophy*, pages 180–194. Oxford University Press, Oxford.
- Hussain, N., and Shah, N., 2006. Misunderstanding metaethics: Korsgaard's rejection of realism. *Oxford Studies in Metaethics*, 1, 265–294.
- Hussain, N., and Shah, N., 2013. Meta-ethics and its discontents: A case study of Korsgaard. In Bagnoli, C., editor, *Constructivism in Ethics*, pages 82–107. Cambridge University Press, Cambridge.
- Ismael, J., 2007. *The Situated Self*. Oxford University Press, Oxford.
- Ismael, J., 2013. *Essays on Symmetry*. Routledge, New York.
- Jackson, F., 1998. *From Metaphysics to Ethics*. Oxford University Press, Oxford.
- Jacquet, L. 2005. March of the penguins. Produced by Ilann Girard. Warner Independent Pictures.
- Jaffe, A. and Quinn, F., 1993. "Theoretical mathematics": Toward a cultural synthesis of mathematics and theoretical physics. *Bulletin of the American Mathematical Society, New Series*, 29, 1–13.
- James, W., 1977. *The Meaning of Truth*. Prometheus Books, Amherst, MA.
- Johnston, M. 1991. Explanation, response-dependence and judgement-dependence. Australian National University Working Papers in Philosophy, no. 1.
- Johnston, M., 1998. Are manifest qualities response-dependent? *Monist*, 81(1), 3–44.
- Kahneman, D., Slovic, P., and Tversky, A., 1982. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge.
- Kant, I., 1902–. *Kants gesammelte Schriften*. de Gruyter, Berlin. 29 volumes, edited by the German Academy of Sciences (formerly the Royal Prussian Academy of Sciences).
- Kant, I., 1991. An answer to the question: "What is Enlightenment?" In Reiss, H., and Nisbet, H. B., editors, *Kant: Political Writings*, 2nd ed., pages 54–60. Cambridge University Press, Cambridge.

- Kant, I., 1992. *Lectures on Logic*. Cambridge University Press, Cambridge. Translated and edited by J. Michael Young.
- Kaplan, D., 1979. Transworld heir lines. In Loux, M., editor, *The Possible and the Actual*, pages 88–109. Cornell University Press, Ithaca, NY.
- Kaufman, Y., 1953/1956. *Toldot Haemuna Hayisraelit: Memai Kedem ad Sof Bayit Shaini*. Bialik Institute/Dvir, Jerusalem/Tel Aviv. 8 volumes.
- Kavka, G., 1983. The toxin puzzle. *Analysis*, 43, 33–36.
- Kazan, E. 1954. On the waterfront. Produced by Sam Spiegel. Horizon Pictures/Columbia Pictures.
- Keefe, R. and Smith, P., 1996. *Vagueness: A Reader*. MIT Press, Cambridge, MA.
- Kirzner, I., 1963. *Market Theory and the Price System*. Van Nostrand, Princeton, NJ.
- Kolodny, N., and MacFarlane, J., 2010. If and oughts. *Journal of Philosophy*, 107(3), 115–143.
- Korsgaard, C., 1996a. *Creating the Kingdom of Ends*. Cambridge University Press, New York.
- Korsgaard, C., 1996b. *The Sources of Normativity*. Cambridge University Press, Cambridge.
- Korsgaard, C., 2008. *The Constitution of Agency*. Oxford University Press, Oxford.
- Korsgaard, C., 2009. *Self-Constitution*. Oxford University Press, New York.
- Krantz, S., 1989. Fractal geometry. *Mathematical Intelligencer*, 11(4), 12–16.
- Kripke, S., 1980. *Naming and Necessity*. Harvard University Press, Cambridge, MA.
- Kubrick, S. 1968. 2001: A space odyssey. Produced by Stanley Kubrick and Victor Lyndon. Metro-Goldwyn-Mayer.
- Kuhn, T., 1970. *The Structure of Scientific Revolutions*, 2nd ed. University of Chicago Press, Chicago.
- Kusch, M., 1995. *Psychologism: A Case Study in the Sociology of Philosophical Knowledge*. Routledge, London.
- LaBarge, S., 1997. Socrates and the recognition of experts. *Apeiron*, 30(4), 51–62.
- Ladyman, J., and Ross, D., 2007. *Every Thing Must Go*. Oxford University Press, Oxford. With David Spurrett and John Collier.
- Laloux, R. 1973. *La planète sauvage*. Screenplay by Roland Topor. Based on novels by Stefan Wul. Anchor Bay Entertainment.
- Latour, B., 1987. *Science in Action*. Harvard University Press, Cambridge, MA.
- Laudan, L., 1981. A confutation of convergent realism. *Philosophy of Science*, 48, 19–49.
- Laxness, H., 1997. *Independent People*. Vintage, New York. Translated by J. A. Thompson.
- Lear, J., 1990. *Love and Its Place in Nature*. Farrar, Straus and Giroux, New York.
- Lee, M.-K., 2005. *Epistemology after Protagoras: Responses to Relativism in Plato, Aristotle, and Democritus*. Oxford University Press, Oxford.
- Lenman, J., and Shemmer, Y., 2012. *Constructivism in Practical Philosophy*. Oxford University Press, Oxford.
- Levi, P., 1996. *Survival in Auschwitz*. Simon and Schuster, New York. Translated by Giulio Einaudi.
- Lewis, C. I., 1946. *An Analysis of Knowledge and Valuation*. Open Court, La Salle.
- Lewis, C. I., 1956. *Mind and the World-Order*. Dover, New York.
- Lewis, D., 1973. *Counterfactuals*. Harvard University Press, Cambridge, MA.
- Lewis, D., 1983–1986. *Philosophical Papers*. Oxford University Press, Oxford.
- Lewis, D., 1986. *On the Plurality of Worlds*. Blackwell, Oxford.
- Lewis, D., 1999. *Papers in Metaphysics and Epistemology*. Cambridge University Press, Cambridge.

- Liao, S. M., 2010. The buck-passing account of value: Lessons from Crisp. *Philosophical Studies*, 151(3), 421–432.
- Lomborg, B., 2001. *The Skeptical Environmentalist*. Cambridge University Press, Cambridge.
- Lovibond, S., 2002. *Ethical Formation*. Harvard University Press, Cambridge, MA.
- Lycan, W., and Shapiro, S., 1986. Actuality and essence. *Midwest Studies in Philosophy*, 11, 343–377.
- MacBride, F., 2001. Can the property boom last? *Proceedings of the Aristotelian Society*, 101(3), 225–246.
- Machery, E., Mallon, R., Nichols, S., and Stich, S., 2004. Semantics, cross-cultural style. *Cognition*, 92, B1–B12.
- MacIntyre, A., 1988. *Whose Justice? Which Rationality?* University of Notre Dame Press, Notre Dame, IN.
- MacIntyre, A., 1990. *Three Rival Versions of Moral Enquiry*. University of Notre Dame Press, Notre Dame, IN.
- MacIntyre, A., 1997. *After Virtue*, 2nd edition. University of Notre Dame Press, Notre Dame, IN.
- MacIntyre, A., 2006. *Ethics and Politics*. Cambridge University Press, Cambridge.
- MacIntyre, A., 2009. *Alasdair MacIntyre's Engagement with Marxism: Selected Writings 1953–1974*. Haymarket Books, Chicago.
- Mackie, J. L., 1977. *Ethics: Inventing Right and Wrong*. Penguin, New York.
- MacKinnon, C., 1989. *Toward a Feminist Theory of the State*. Harvard University Press, Cambridge, MA.
- Mandler, M., 2001. A difficult choice in preference theory. In Millgram, E., editor, *Varieties of Practical Reasoning*. MIT Press, Cambridge, MA.
- Mann, W.-R., 2000. *The Discovery of Things*. Princeton University Press, Princeton, NJ.
- Mares, M., Ojeda, R., Borghi, C., Giannoni, S., Diaz, G., and Braun, J., 1997. How desert rodents overcome halophytic plant defenses. *BioScience*, 47(10), 699–704.
- Martin, S., 1998. Changes in the memory after fifty. In *Pure Drivel*, pages 18–21. Hyperion, New York.
- Mayr, E., 1984. Species concepts and their application. In Sober, E., editor, *Conceptual Issues in Evolutionary Biology*, pages 531–540. MIT Press, Cambridge, MA.
- McDowell, J., 1994. *Mind and World*. Harvard University Press, Cambridge, MA.
- McDowell, J., 1998. *Mind, Value, and Reality*. Harvard University Press, Cambridge, MA.
- Medin, D., Goldstone, R., and Gentner, D., 1993. Respects for similarity. *Psychological Review*, 100(2), 254–278.
- Melia, J., 1992. Against modalism. *Philosophical Studies*, 68(1), 35–56.
- Melia, J., 2003. *Modality*. McGill-Queen's University Press, Montreal and Kingston.
- Menzies, P., and Pettit, P., 1993. Found: The missing explanation. *Analysis*, 53(2), 100–109.
- Mill, J. S., 1967–1989. *Collected Works of John Stuart Mill*. University of Toronto Press/Routledge and Kegan Paul, Toronto/London.
- Millar, R., 1972. *The Piltdown Men*. Victor Gollancz, London.
- Miller, A., 2001. The missing explanation argument revisited. *Analysis*, 61(1), 76–86.
- Millgram, E., 1995. Inhaltsreiche ethische Begriffe und die Unterscheidung zwischen Tatsachen und Werten. In Fehige, C. and Meggle, G., editors, *Zum moralischen Denken*, pages 354–388. Surhkamp, Frankfurt a.M.

- Millgram, E., 1996. Williams' argument against external reasons. *Noûs*, 30(2), 197–220.
- Millgram, E., 1997. *Practical Induction*. Harvard University Press, Cambridge, MA.
- Millgram, E., 1999. Moral values and secondary qualities. *American Philosophical Quarterly*, 36(3), 253–255.
- Millgram, E., editor, 2001. *Varieties of Practical Reasoning*. MIT Press, Cambridge, MA.
- Millgram, E., 2004a. On being bored out of your mind. *Proceedings of the Aristotelian Society*, 104(2), 163–184.
- Millgram, E., 2004b. The ontological meta-argument (and the ontological argument for the actuality of the world). *Croatian Journal of Philosophy*, 4(12), 331–334.
- Millgram, E., 2005a. *Ethics Done Right: Practical Reasoning as a Foundation for Moral Theory*. Cambridge University Press, Cambridge.
- Millgram, E., 2005b. Practical reason and the structure of actions. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/sum2012/entries/practical-reason-action/>.
- Millgram, E., 2006a. Review of Bas van Fraassen, *The Empirical Stance*. *Philosophical Review*, 115(3), 404–408.
- Millgram, E., 2006b. Review of Candace Vogler, *Reasonably Vicious*. *European Journal of Philosophy*, 14(3), 430–434.
- Millgram, E., 2008. Specificationism. In Adler, J. and Rips, L., editors, *Reasoning: Studies of Human Inference and its Foundations*, pages 731–747. Cambridge University Press, Cambridge.
- Millgram, E., 2009a. *Hard Truths*. Wiley-Blackwell, Oxford.
- Millgram, E., 2009b. John Stuart Mill, determinism, and the problem of induction. *Australasian Journal of Philosophy*, 87(2), 183–199.
- Millgram, E., 2009c. Liberty, the higher pleasures, and Mill's missing science of ethnic jokes. *Social Philosophy and Policy*, 26(1), 326–353.
- Millgram, E., 2009d. The persistence of moral skepticism and the limits of moral education. In Siegel, H., editor, *Oxford Handbook of Philosophy of Education*, pages 245–259. Oxford University Press, New York.
- Millgram, E., 2009e. Review of Michael Thompson, *Life and Action*. *Analysis Reviews*, 69(3), 1–7.
- Millgram, E., 2011a. Critical notice: Christine Korsgaard, *Self-Constitution and The Constitution of Agency*. *Australasian Journal of Philosophy*, 89(3), 549–556.
- Millgram, E., 2011b. Ismael's Anscombian and Dennettian selves. *Philosophy and Phenomenological Research*, 82(3), 759–762.
- Millgram, E., 2013. Relativism, coherence, and the problems of philosophy. In O'Rourke, F., editor, *What Happened in and to Moral Philosophy in the Twentieth Century?*, pages 392–422. Notre Dame University Press, Notre Dame, IN.
- Millgram, E., 2014a. Mill's and Whewell's competing visions of logic. In Loizides, A., *Mill's A System of Logic: Critical Appraisals*, pages 101–121. Routledge, New York.
- Millgram, E., 2014b. Private persons and minimal persons. *Journal of Social Philosophy*, 45(3), 323–347.
- Millgram, E., forthcoming. Mill's epiphanies. In Macleod, C., and Miller, D., editors, *A Companion to John Stuart Mill*. Wiley-Blackwell, Oxford.
- Mitchell, J., 1999. *Eccentric Lives and Peculiar Notions*. Black Dog and Leventhal Publishers, New York.
- Moore, A., Gibbons, D., and Higgins, J., 2005. *Watchmen*. DC Comics, New York.

- Moran, R., 1994. The expression of feeling in imagination. *Philosophical Review*, 103(1), 75–106.
- Morison, S. E., 1992. *Admiral of the Ocean Sea: A Life of Christopher Columbus*. Book-of-the-Month Club, New York.
- Morton, A., 1991. *Disasters and Dilemmas*. Blackwell, Oxford.
- Mosse, D., 2005. *Cultivating Development*. Pluto Press, London.
- Munger, C., 2007. *Poor Charlie's Almanack*, 2nd edition. Walsworth, Marcelline. Edited by Peter Kaufman.
- Murdoch, I., 1970. *The Sovereignty of Good*. Routledge Kegan Paul, London.
- Nagel, T., 1978. *The Possibility of Altruism*. Princeton University Press, Princeton, NJ.
- Nagel, T., 2007. The value of inviolability. In Bloomfield, P., editor, *Morality and Self-Interest*, pages 102–113. Oxford University Press, Oxford.
- Nehamas, A., 1981. The postulated author: Critical monism as a regulative ideal. *Critical Inquiry*, 8, 133–149.
- Nehamas, A., 1985. *Nietzsche: Life as Literature*. Harvard University Press, Cambridge, MA.
- Nehamas, A., 1987. Writer, text, work, author. In Cascardi, A. J., editor, *Literature and the Question of Philosophy*, pages 265–291. Johns Hopkins, Baltimore.
- Neurath, O., 1959. Protocol sentences. In Ayer, A. J., editor, *Logical Positivism*, pages 199–208. MacMillan, New York.
- Nichols, S., 2006. Imaginative blocks and impossibility: An essay in modal psychology. In Nichols, S., editor, *The Architecture of the Imagination*, pages 237–255. Oxford University Press, Oxford.
- Nichols, S., and Stich, S., 2003. *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Clarendon Press, Oxford.
- Nietzsche, F., 1988. *Sämtliche Werke*. Deutscher Taschenbuch Verlag/de Gruyter, Berlin. 15 volumes. Edited by G. Colli and M. Montinari.
- Nietzsche, F., 1989. *On the Genealogy of Morals and Ecce Homo*. Vintage Books, New York. Translated by Walter Kaufmann.
- Nozick, R., 1981. *Philosophical Explanations*. Harvard University Press, Cambridge, MA.
- Nozick, R., 1997. *Socratic Puzzles*. Harvard University Press, Cambridge, MA.
- Odling-Smee, F. J., Laland, K., and Feldman, M., 2003. *Niche Construction*. Princeton University Press, Princeton, NJ.
- Oreskes, N. and Conway, E., 2010. *Merchants of Doubt*. Bloomsbury Press, New York.
- Parfit, D., 1987. *Reasons and Persons*. Oxford University Press, Oxford.
- Paseau, A., 2006. Genuine modal realism and completeness. *Mind*, 115(459), 721–729.
- Pitcher, G., 1970. Pain perception. *Philosophical Review*, 79(3), 368–393.
- Plantiga, A., 1987. Two concepts of modality. *Philosophical Perspectives*, 1, 189–231.
- Plato, 1997. *Complete Works*. Hackett, Indianapolis, IN. Edited by J. Cooper and D. S. Hutchinson.
- Polya, G., 1990. *Mathematics and Plausible Reasoning*. Princeton University Press, Princeton, NJ.
- Prescott, W., 2000. *History of the Conquest of Mexico and History of the Conquest of Peru*. Cooper Square Press, New York.
- Prichard, H. A., 1912. Does moral philosophy rest on a mistake? *Mind*, 21(81), 21–37.
- Priest, G., 2001. *An Introduction to Non-Classical Logic*. Cambridge University Press, Cambridge.

- Pruss, A., 2007. Conjunctions, disjunctions and Lewisian semantics for counterfactuals. *Synthese*, 156, 33–52.
- Putnam, H., 1975a. Brains and behavior. In *Mind, Language and Reality*, pages 325–341. Cambridge University Press, Cambridge.
- Putnam, H., 1975b. The meaning of “meaning”. In *Mind, Language and Reality*, pages 215–271. Cambridge University Press, Cambridge.
- Putnam, H., 1981. *Reason, Truth and History*. Cambridge University Press, Cambridge.
- Putnam, H., 1990a. Is water necessarily H<sub>2</sub>O? In Conant, J., editor, *Realism with a Human Face*, pages 54–79. Harvard University Press, Cambridge, MA.
- Putnam, H., 1990b. Objectivity and the science/ethics distinction. In Conant, J., editor, *Realism with a Human Face*, pages 163–178. Harvard University Press, Cambridge, MA.
- Quine, W. V., 1966. Carnap and logical truth. In *The Ways of Paradox*, pages 100–125. Random House, New York.
- Rabin, M., 1976. Probabilistic algorithms. In Traub, J. F., editor, *Algorithms and Complexity*, pages 21–40. Academic Press, New York.
- Rabin, M., 1980. A probabilistic algorithm for testing primality. *Journal of Number Theory*, 12(1), 128–138.
- Ranney, M. 2013. How global warming works. Education videos: <http://www.howglobalwarmingworks.org/>.
- Ravetz, J., 1979. *Scientific Knowledge and Its Social Problems*. Oxford University Press, New York.
- Rawls, J., 1989. Themes in Kant’s moral philosophy. In Förster, E., editor, *Kant’s Transcendental Deductions*, pages 81–113. Stanford University Press, Stanford, CA.
- Rayo, A., 2013. *The Construction of Logical Space*. Oxford University Press, Oxford.
- Raz, J., 1990. Authority and justification. In Raz, J., editor, *Authority*, pages 115–141. New York University Press, New York.
- Reutlinger, A., Schurz, G., and Hüttemann, A., 2011. Ceteris paribus laws. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. <http://plato.stanford.edu/archives/spr2011/entries/ceteris-paribus>.
- Richard, M., 2008. *When Truth Gives Out*. Oxford University Press, Oxford.
- Richerson, P., and Boyd, R., 2005. *Not By Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press, Chicago.
- Richardson, H., 2013. Revising moral norms. In Bagnoli, C., editor, *Constructivism in Ethics*, pages 221–242. Cambridge University Press, Cambridge.
- Ringer, F., 1990. *The Decline of the German Mandarins*. Wesleyan University Press, Hanover, NH.
- Robertson, J., 2001. Internalism, practical reason, and motivation. In Millgram, E., editor, *Varieties of Practical Reasoning*, pages 127–151. MIT Press, Cambridge, MA.
- Rose, P. L., 1990. *German Question/Jewish Question*. Princeton University Press, Princeton, NJ.
- Rovane, C., 1990. Branching self-consciousness. *Philosophical Review*, 99(3), 355–395.
- Rowe, T., 1987. Definition and diagnosis in the phylogenetic system. *Systematic Zoology*, 36(2), 208–211.
- Russell, B., 1999. William James’s conception of truth. In Blackburn, S. and Simmons, K., editors, *Truth*, pages 69–82. Oxford University Press, Oxford.
- Salmon, W., 1974. The concept of inductive evidence. In Swinburne, R., editor, *The Justification of Induction*, pages 48–57. Oxford University Press, Oxford.

- Scanlon, T. M., 1998. *What We Owe to Each Other*. Harvard University Press, Cambridge.
- Schaffer, J., 2009. On what grounds what. In Chalmers, D., Manley, D., and Wasserman, R., editors, *Metametaphysics*, pages 347–383. Oxford University Press, Oxford.
- Schapiro, T., 2006. Kantian rigorism and mitigating circumstances. *Ethics*, 117, 32–57.
- Schmidtz, D., 2009. Because it's right. In Black, S. and Tiffany, E., editors, *Reasons to Be Moral Revisited*, pages 63–95. University of Calgary Press, Calgary.
- Seuss, D., 1949. *Bartholomew and the Oobleck*. Random House, New York.
- Shafir, E., 1993. Choosing versus rejecting: Why some options are both better and worse than others. *Memory and Cognition*, 21(4), 546–556.
- Shalkowski, S., 1994. The ontological ground of the alethic modality. *Philosophical Review*, 103(4), 669–688.
- Shope, R., 1978. The conditional fallacy in contemporary philosophy. *Journal of Philosophy*, 75(8), 397–413.
- Shope, R., 1983. *The Analysis of Knowing*. Princeton University Press, Princeton, NJ.
- Shore, D. 1991. *16 Most Requested Songs*. Sony.
- Sider, T., 2003. Reductive theories of modality. In Loux, M., and Zimmerman, D., editors, *The Oxford Handbook of Metaphysics*, pages 180–208. Oxford University Press, Oxford.
- Sidgwick, H., 1907/1981. *The Methods of Ethics*. Hackett, Indianapolis, IN.
- Skyrms, B., 1976. Possible worlds, physics and metaphysics. *Philosophical Studies*, 30(5), 323–332.
- Smith, A., 1981. *An Inquiry into the Nature and Causes of the Wealth of Nations*. Liberty Fund, Indianapolis, IN.
- Smith, M., 1987. The Humean theory of motivation. *Mind*, 96(381), 36–61.
- Snyder, L., 2011. *The Philosophical Breakfast Club*. Random House, New York.
- Stallman, R., Weinreb, D., and Moon, D., 1984. Lisp machine manual, 6th edition. MIT Artificial Intelligence Laboratory.
- Stalnaker, R., 2012. *Mere Possibilities*. Princeton University Press, Princeton, NJ.
- Stearns, S., 2004. *The Evolution of Life Histories*. Oxford University Press, Oxford.
- Sterelny, K., 2003. *Thought in a Hostile World: The Evolution of Human Cognition*. Blackwell, Oxford.
- Sterelny, K., 2010. Mind: Extended or scaffolded? *Phenomenology and the Cognitive Sciences*, 9(4), 465–481.
- Stevenson, C. L., 1944. *Ethics and Language*. Yale University Press, New Haven.
- Stock, K., 2005. Resisting imaginative resistance. *Philosophical Quarterly*, 55(221), 607–624.
- Stokes, D., 2006. The imaginative character of evaluative resistance. *British Journal of Aesthetics*, 46(4), 387–405.
- Stoljar, D., 2010. *Physicalism*. Routledge, London.
- Stratton-Lake, P. and Hooker, B., 2006. Scanlon versus Moore on goodness. In Horgan, T., and Timmons, M., editors, *Metaethics after Moore*, pages 149–168. Oxford University Press, Oxford.
- Strawson, P. F., 1971. *Individuals*. Methuen, London.
- Tabery, J., 2014. *Beyond Versus: The Struggle to Understand the Interaction of Nature and Nurture*. MIT Press, Cambridge, MA.
- Tal, E., 2014. Making time: A study in the epistemology of measurement. *British Journal for the Philosophy of Science*, <http://doi.org/10.1093/bjps/axu037>.
- Taleb, N., 2010. *The Black Swan*, 2nd edition. Random House, New York.
- Taleb, N., 2012. *Antifragile: Things That Gain from Disorder*. Random House, New York.



- Taylor, F. W., 1997. *The Principles of Scientific Management*. Dover, New York.
- Tenn, W., 2001. *Of Men and Monsters*. Orion, London.
- Thompson, M., 2008. *Life and Action*. Harvard University Press, Cambridge, MA.
- Tubert, A., 2010. Constitutive arguments. *Philosophy Compass*, 5, 656–666.
- Urmson, J. O., 1968. *The Emotive Theory of Ethics*. Hutchinson, London.
- Velleman, J. D., 2000. *The Possibility of Practical Reason*. Oxford University Press, Oxford.
- Vines, T., Albert, A., Andrew, R., Débarre, F., Bock, D., Franklin, M., Gilbert, K., Moore, J.-S., Renaut, S., and Rennison, D., 2013. The availability of research data declines rapidly with article age. *Current Biology*, 24, 1–4, 94–97.
- Vogler, C., 2001. *John Stuart Mill's Deliberative Landscape*. Garland Publishing, New York.
- Vogler, C., 2002. *Reasonably Vicious*. Harvard University Press, Cambridge.
- Waits, T., 1976. *Small Change*. Asylum Records, Los Angeles.
- Watson, G., 2004. *Agency and Answerability*. Oxford University Press, Oxford.
- Watterson, B., 1996. *It's a Magical World*. Andrews and McMeel, Kansas City.
- Wedgwood, R., 2007. *The Nature of Normativity*. Oxford University Press, Oxford.
- Weinberg, J., 2013. The prospects for an experimentalist rationalism. In Casullo, A., and Thurow, J., editors, *The A Priori in Philosophy*, pages 92–108. Oxford University Press, Oxford.
- Weinberg, J., Nichols, S., and Stich, S., 2001. Normativity and epistemic intuitions. *Philosophical Topics*, 29(1 and 2), 429–460.
- West-Eberhard, M. J., 2003. *Developmental Plasticity and Evolution*. Oxford University Press, Oxford.
- Whewell, W., 1847. *The Philosophy of the Inductive Sciences*. John W. Parker, London.
- Whewell, W., 1857. *History of the Inductive Sciences*, 3rd edition. John W. Parker and Son, London.
- Whewell, W., 1858. *Novum Organon Renovatum*. John W. Parker and Son, London.
- White, M., 1963. *Toward Reunion in Philosophy*. Atheneum, New York.
- Wiggins, D., 1991. A sensible subjectivism? In *Needs, Values, Truth*, 2nd edition, pages 185–214. Blackwell, Oxford.
- Wiggins, D., 2001. *Sameness and Substance Renewed*. Cambridge University Press, Cambridge.
- Williams, B., 1973a. A critique of utilitarianism. In Smart, J. J. C., and Williams, B., editors, *Utilitarianism: For and Against*, pages 75–150. Cambridge University Press, Cambridge.
- Williams, B., 1973b. *Problems of the Self*. Cambridge University Press, Cambridge.
- Williams, B., 1978. *Descartes: The Project of Pure Enquiry*. Penguin, New York.
- Williams, B., 1981a. *Moral Luck*. Cambridge University Press, Cambridge.
- Williams, B., 1981b. *Obscenity and Film Censorship: An Abridgement of the Williams Report*. Cambridge University Press, Cambridge.
- Williams, B., 1985. *Ethics and the Limits of Philosophy*. Harvard University Press, Cambridge, MA.
- Williams, B., 1993a. *Morality*, 2nd edition. Cambridge University Press, Cambridge.
- Williams, B., 1993b. *Shame and Necessity*. University of California Press, Berkeley.
- Williams, B., 1995a. *Making Sense of Humanity*. Cambridge University Press, Cambridge.
- Williams, B., 1995b. Truth in ethics. *Ratio*, 8(3), 227–242.
- Williams, B., 2001. Internal and external reasons (with postscript). In Millgram, E., editor, *Varieties of Practical Reasoning*, MIT Press, Cambridge, MA.

- Williams, B., 2002. *Truth and Truthfulness*. Princeton University Press, Princeton, NJ.
- Williams, B., 2006a. *Philosophy as a Humanistic Discipline*. Princeton University Press, Princeton, NJ. Edited by A. W. Moore.
- Williams, B., 2006b. *The Sense of the Past*. Princeton University Press, Princeton, NJ. Edited by Myles Burnyeat.
- Wilson, M., 2006. *Wandering Significance*. Oxford University Press, Oxford.
- Wimsatt, W., 2007. *Re-Engineering Philosophy for Limited Beings*. Harvard University Press, Cambridge, MA.
- Wittgenstein, L., 1983. *Remarks on the Foundations of Mathematics*, revised edition. MIT Press, Cambridge, MA. Translated by G. E. M. Anscombe. Edited by G. H. von Wright, R. Rhees, and G. E. M. Anscombe.
- Wittgenstein, L., 1998. *Philosophical Investigations*, 3rd edition. Blackwell, Oxford. Translated by G. E. M. Anscombe.
- Wodehouse, P. G., 1984. *Carry On, Jeeves*. Penguin, Harmondsworth.
- Wootton, D., 2007. *Bad Medicine*. Oxford University Press, Oxford.
- Yagisawa, T., 1988. Beyond possible worlds. *Philosophical Studies*, 53(2), 175–204.
- Zahavi, A., and Zahavi, A., 1999. *The Handicap Principle*. Oxford University Press, Oxford. Translated by Naama Zahavi-Ely and Melvin Patrick Ely.
- Zaller, J., 1992. *The Nature and Origins of Mass Opinion*. Cambridge University Press, Cambridge.



# INDEX

- absolutism, 128, 130
- action, 203 n. 31, 234–235, 240
- action, superlative attribution of, *See* identification
- actual world, 159, 161–162, 176
- agency, 8, 115, 234–260, 262–264
  - model vs. theory of, 237, 250
  - nonhomuncular, 237, 256
  - segmented, 235–236, 247, 250–251, 255–257, 260, 262–263
  - self-governing, 237, 240, 258, 259
  - unity of, 234–235, 262, 266
- agglomeration, 82–85, 89–92, 94–96, 98, 150
- akrasia*, *See* weakness of will
- akrasia*, epistemic, 148
- analyticity, 190, 212, 215–218
- analytic-synthetic distinction, 98, 212, 216
- anchoring, 272
- Anscombe, G. E. M., 79 n. 2, 83 n. 11, 103
- ants, 6, 16, 19, 25
- a posteriori, 5, 7, 212, 214, 217
- approximation, *See* truth, partial
- apriority, 4–5, 7, 9, 11–12, 13 n. 19, 17, 190, 212–219
- Aquinas, Thomas, 63 n. 17
- argument(s), 105–106, 113–120, 122 n. 37, 131–135, 140 n. 30
  - quantifier over, 111, 117 n. 30, 118 n. 32, 119, 120 n. 35, 121, 131–135, 140, 143 n. 34, 147
- Aristotle, 15 n. 22, 16, 18, 27, 43, 54, 70, 73–74, 113 n. 23, 118 n. 32, 119 n. 33, 132, 203 n. 31, 208, 218
- Aristotelian categoricals, 74 n. 35
- Asimov, Isaac, 127
- attention management, 61, 189, 195–201, 206–207, 210–216, 219–220
- Austin, J. L., 196 n. 17
- autonomy, 15, 22–25, 28–31, 33–35, 41, 44, 229 n. 16, 236–237, 240–241, 243–244, 250, 255–258, 260, 263–264
  - in philosophy, 259 n. 47, 271, 275
- Babel, Tower of, 1, 3, 5, 15
- bees, 6, 19, 84
- belief, 68, 82, 89, 115 n. 28, 130 n. 8, 175
- benchmarking, 46, 48 n. 47, 276
- Bentham, Jeremy, 21 n. 1, 24 n. 4, 34, 63, 75
- Big Claim, 264–265
- birds, 84
  - migrating, 55–56, 59, 73
- black box, 160–162, 170, 172, 176, 178
- Blackburn, Simon, 117 n. 31
- bodily continuity, 261 n. 52
- Bond, James 242 n. 20
- boredom, 61, 72, 225, 232, 260, 262–263, 267–268
- Bowman, Margaret, 123
- brakes, 40
- Brandom, Robert, 116 n. 29
- Bratman, Michael, 235–268
- Bratman's Master Move, 241–242, 244, 256
- businessman, 28, 226
- Calvinball, 118 n. 32
- Canberra Plan, 152 n. 51, 160, 174, 176–177, 179–182, 184–185
- Cantor, Georg, 158, 210
- Carnap, Rudolph, 11 n. 13, 22, 205 n. 34
- Carroll, Lewis, 119, 160
- cats, 164, 208 n. 39
- causation, 114, 191, 217
- ceteris paribus clause, *See* defeasibility
- Chisholm, Roderick, 163–164
- CI-procedure, 24 n. 7, 73, 115, 118 n. 32, 133 n. 15, 222
- climate change, 37 n. 28, 38, 97 n. 2
- coffee, 17
- cognitive function analysis, 5, 8, 13–15, 87, 126–127, 141–143, 151–152, 190, 193, 206, 210, 212, 219–220, 269, 279
- coherence, 10, 74–75, 92 n. 27, 118 n. 32
- color, 84
- Columbus, Christopher, 259
- computational complexity, 39, 41 n. 40, 115 n. 28
- Comte, Auguste, 101–102
- conceptual analysis. *See* semantic analysis
- conditional fallacy, 65 n. 22, 110, 176 n. 35, 190 n. 4
- consciousness, 238 n. 15
- consistency, 92
- constructivism, 105 n. 8, 118 n. 32, 120 n. 35
- context-sensitivity, 170 n. 24, 179 n. 36
- Cortés, Hernán, 259–260

- counterfactual, 66, 68, 144, 157–160, 163–191  
     Psychic Hotline, 191–192  
 countermoral, 144  
 counterpart, 159–160, 165–168, 177, 187  
 Craig, Edward, 87–88  
 creature construction, 91–92, 126, 194 n. 12  
 cross-disciplinary interfaces, 15, 18, 28, 32, 35, 40,  
     42, 96–97, 205, 277–280
- Davidson, Donald, 88 n. 22, 169 n. 21  
 death, 56 n. 5, 70, 81, 136, 231 n. 17, 261 n. 52  
 Decided Preference Criterion, 109–112, 116,  
     138, 143  
 defeasibility, 31–32, 39 n. 36, 42, 49, 51–52, 111 n.  
     19, 131–135, 164–165, 167–168, 179–182,  
     203–205, 218 n. 55, 279  
 definition, 138–140, 217, 218 n. 55  
 desire, 2, 7–8, 17, 59–60, 62–63, 65, 67, 72–73,  
     79–82, 89, 92–93, 106, 109, 111, 113,  
     115 n. 28, 175, 222, 228 n. 15, 235, 238,  
     240–241, 263, 265  
     categorical, 81, 234  
     informed, 65–69, 110, 116, 138, 143, 161 n. 9,  
         250 n. 31  
     occurrent vs. dispositional, 68  
     second-order, 235 n. 4, 237  
     strength of, 63  
 desire, superlative attribution of, *See*  
     identification  
 determinism, 180 n. 38  
 Dewey, John, 13 n. 19, 269 n. 1  
 Diamond, Cora, 122 n. 37  
 Dijkstra, Edsger, 141 n. 31  
 disposition, 68–69, 157, 189, 190–191, 248 n. 29  
 division of labor, *See* hyperspecialization  
 ducks, 103  
 duty, 103
- education, quality control in, 51–52  
 Eichmann, Adolf, 231  
 emotivism, 107, 131 n. 8, 198 n. 25  
 Empiricism, British, 114  
 Endarkenment, 22, 36, 39, 41, 43–44, 269, 273,  
     275–276, 278–281  
 ends, *See* desire  
     deliberation of, 60, 62  
     incommensurable, 63–65, 73 n. 32, 262  
     specification of, 60 n. 13, 79 n. 3, 93 n. 29,  
         115, 133  
 Enlightenment, 21–22, 24–25, 27, 31, 33–39, 42,  
     75, 112  
 epicycle, 166  
 epistemology, 15, 19–20, 22, 25, 87, 152 n. 51, 271  
     Bayesian, 33 n. 22  
     moral, 129–130
- ergon*, 10, 18–20, 25, 27, 54, 73–74, 252 n. 34, 260,  
     273, 278  
 error theory, 108 n. 13  
 essence, 165, 167, 186 n. 45, 190, 208, 210  
 ethics, 10, 230  
     applied, 83 n. 14, 221, 226 n. 10, 229 n. 16, 233  
     as generic social regulation, 6, 149,  
         228–230, 232  
     revisable, 6–7, 231  
     species-specific, 19, 224  
 exploration, 60–61, 75, 259–260  
*eudaemonia*, 73–75, 93 n. 29, 222, 228 n. 15,  
     252 n. 34  
*Euthyphro* argument, 216–217  
 evolution, 54–55, 71, 91 n. 25, 96 n. 2, 223 n. 5  
     -ary just-so story, 72 n. 30, 193 n. 10
- fact-value distinction, 82–83, 85–86, 89, 98–99  
 Ferrero, Luca, 245 n. 25  
 fiction writers, 68–69, 232  
 Finkelstein, David, 191  
 Foot, Philippa, 116 n. 29  
 Foucault, Michel, 21 n. 1, 242 n. 22  
 Frankfurt, Harry 81 n. 6, 235 n. 4, 237, 243–244  
 freedom, 40–41, 109 n. 16  
 Frege, Gottlob, 213 n. 48  
 Friedman, Michael, 214 n. 49, 253 n. 37  
 frustration, 62, 72, 260, 262, 267  
 functionalism, analytic, 152 n. 51, 174–175  
 funny, 85, 217–218
- Gauguin, Paul, 85 n. 17, 99  
 genealogy, 88 n. 22  
 ghost, 107–108, 115, 117, 167, 237  
 Ginsberg, Benjamin, 47 n. 47  
 goal, *See* desire  
 gold standard, 14 n. 21, 120 n. 35  
 Gombrich, Ernst, 11 n. 13  
 good, 103, 107, 119 n. 33, 149  
     buck-passing account of, 106 n. 10, 145 n. 37  
 Goodman, Nelson, 181, 266 n. 55  
 Great Recession, 39, 231  
 Grice, Paul, 91 n. 25  
 grounding, 11 n. 13  
 Grover, Dorothy, 142
- Halting Problem, 115 n. 28  
 Harris, Sidney, 24  
 Hayek, Friedrich, 49  
 hedonic response, 23 n. 3, 60–63, 72–73, 225,  
     232, 259  
 Hegel, G. W. F., 281  
 Homer, 9 n. 10  
 Hopper, Edward, 133  
 Housman, A. E., 166

- Hume, David, 114, 134, 194 n. 11, 211  
 hybrid, 159  
 hyperspecialization, 2, 7, 9, 28, 32–35, 70–72,  
   148, 236  
   cognitive repertoire for, 6, 15–19, 51 n. 51,  
     60–65, 69, 73, 93, 221, 225, 232  
   necessary for survival, 3, 71  
   parallel, 64–65, 262  
   serial, explication of, 6, 16, 26, 58–59, 98, 147,  
     201, 223, 245–246, 274  
  
 ideal, 240 n. 18, 254–255  
 ideal advisor theory, 66, 69 n. 27  
 idealization, *See* truth, partial  
 identification, 8, 234, 236–237, 240–244, 254 n.  
   40, 255–258, 262–263  
 imagination, 67, 69, 145 n. 38, 192 n. 8, 212 n. 46,  
   263, 267 n. 59  
   simulation theory of, 67, 69, 192 n. 8  
 imaginative resistance, 144–145  
 imperative, 82, 89  
 impossible worlds, 184  
 induction, 22, 104 n. 6, 114, 132  
   pessimistic meta-, 97 n. 33, 199  
   practical, 7, 60–61, 63, 71–72, 113–114, 132, 144  
     n. 35, 249 n. 30  
   problem of, 114  
 infinity, 158, 204 n. 32, 210  
 instrumental reasoning, *See* means-end  
   reasoning  
 instrumentalism, 7, 17, 59–60, 62–63, 65, 75–76,  
   79, 109, 111, 113, 115, 121, 144 n. 35, 206  
   n. 35, 260 n. 51  
 intellectual ergonomics, 5, 13–15, 275 n. 11  
 intention, 24 n. 7, 83 n. 11, 119 n. 33, 238,  
   240–241, 260  
 interest, 61, 72, 225, 232, 260, 267–268  
 internalism, 77, 79–81, 83, 92–93, 98–99, 144, 245  
   n. 25, 267 n. 58  
 intuition, 5 n. 5, 11, 117 n. 30, 128, 130, 152 n. 51,  
   190, 208–210  
 intuitionism, 129–130, 131 n. 8  
 Ion, 9 n. 10  
  
 James, William, 139 n. 28, 216 n. 51  
 Johnson, Amanda, 197  
 Johnston, Mark, 190, 216–219  
 just-really view, 128, 144  
  
 Kahneman, Daniel, 176 n. 32  
 Kant, Immanuel, 18–20, 22–24, 27, 34 n. 24, 73,  
   213 n. 47, 228 n. 15  
 Kaufmann, Yehezkiel, 21 n. 1  
 Kierkegaard, Søren, 218  
 Korsgaard, Christine, 133 n. 15, 235 n. 4, 242 n. 22  
  
 knowledge, 20, 87–88, 152 n. 51, 214  
   ethical, 85  
 Kriesche, Johannes, 133  
 Kripke, Saul, 158 n. 3  
 Kuhn, Thomas, 199, 253  
  
 Latour, Bruno, 39  
 law, 224  
 Leiter, Brian, 271  
 Leiterism, 273, 275–276  
 Leiter Report, 145, 225, 246, 254, 271–272, 275  
 Lewis, Clarence Irving, 158 n. 3, 164  
 Lewis, David, 16, 155–179, 181–186  
 lie, 226 n. 11  
 Lisp Machine, 30 n. 15  
 literary studies, 9 n. 10, 33 n. 22, 224  
 Little Claim, 264–265  
 Locke, John, 240  
 logical alien, 32–34, 278–280  
 logic,  
   deductive, 22–24, 31, 104 n. 6, 108 n. 14, 122 n.  
     37, 126, 158  
   modal, 158, 188, 190, 210–212  
   philosophy of, 22, 24, 42–44, 114, 279–280  
  
 MacIntyre, Alasdair, 21 n. 1, 131 n. 8  
 markets, 21, 36 n. 27, 49–50, 52  
 material object, 164  
 mathematics, 33 n. 22, 70, 189  
   counterfactuals in, 185 n. 44,  
   philosophy of, 108  
   standards for argumentation in, 33 n. 22  
 Mayr, Ernst, 26 n. 10  
 McDowell, John, 267 n. 58  
 means-end reasoning, 68–69, 75–76, 79, 92, 113,  
   132–133, 248 n. 48  
 medical decision-making, 35 n. 25, 37, 50 n. 49,  
   254 n. 40  
 Meinong, Alexius, 159  
 memory, 240, 260  
   loss of, 12, 213–214, 220, 260–261  
 mental states, 115 n. 28, 152 n. 51, 174–175,  
   243–244, 261  
   occurrent vs. dispositional, 68, 122 n. 37, 167,  
     176–178  
   noninstrumental, 239  
 mental states, superlative attribution of, *See*  
   identification  
 metaethics, 3, 8, 10, 101, 104–105, 107–108,  
   110–119, 124–126, 150 n. 49, 151,  
   153–154, 233, 237, 259  
 metaphysics, 5, 10–16, 86–87, 94, 96, 100, 102,  
   104, 106–107, 108 n. 13, 153, 156, 168,  
   174, 205, 229  
   as intellectual ergonomics, 5, 13, 19, 153, 189

- metaphysics (*continued*)  
 caveman, 15 n. 22  
 old-school, 5, 10, 11 n. 13, 13 n. 17, 14, 16,  
 112, 114, 118, 120 n. 35, 122 n. 37, 153,  
 198, 205–206, 209, 220  
 Mill, John Stuart, 25 n. 8, 101–112, 114, 122, 138,  
 190, 192, 194, 213 n. 48  
 mind,  
   philosophy of, 174  
   small-finite, 171  
 Minsky, Marvin, 191 n. 7  
 Missing Explanation Argument, 190, 216–219  
 Mithridates, 157, 165–166, 168–169, 172 n. 26,  
 176, 186  
 modal  
   cognition, 14, 188, 190–192  
   operators, iterated, 158, 212  
   realism, 16, 155, 162 n. 16, 186  
   similarity judgments, 158, 161 n. 9, 168–173,  
   175–185  
 modality, 10, 16, 156–157, 188  
   epistemic, 160 n. 8, 197 n. 22  
 Moore, G. E., 101, 107, 128 n. 3  
 morality. *See* ethics.  
 moral  
   expertise, 150 n. 49, 229 n. 16, 232 n. 20  
   luck, 85  
   psychology, 234  
   realism, 107, 115 n. 28, 130, 137, 259  
   theory, 3, 7–8, 72–75, 78, 81–82, 85–86, 93 n.  
   29, 103, 108, 117 n. 31, 128, 230, 234  
     Aristotelian, 73–75, 94 n. 29  
     Kantian, 24 n. 7, 35, 73, 75, 81, 94 n. 29, 133  
     n. 15, 227 n. 14, 228 n. 15  
 Munger, Charles, 226 n. 12  
 Murdoch, Iris, 267 n. 61  
  
 Nagel, Thomas, 245 n. 25  
 naturalism, 21 n. 1, 101, 107–108, 155, 198 n. 25, 237  
 necessity, 14, 104, 114, 157–158, 163, 188–210,  
 214–216, 218, 220  
   expiration of, 207  
   logical, 102–104, 192, 196, 200, 204, 208 n. 39  
   mathematical, 189, 195–196, 204  
   metaphysical, 197 n. 22, 200, 202, 205,  
   208–210  
   progressive, 190, 198–199, 206, 220  
   specialized, 189, 201–202, 214–215, 220  
 neo-Taylorism, 48 n. 47  
 Neurath, Otto, 248, 250  
 Newcomb's Problem, 121 n. 36  
 Newton, Isaac, 198  
 niche, 16–17, 25–26, 48 n. 47, 55–58, 60, 62–69,  
 73, 98, 146, 200–202, 207, 223–224,  
 228, 230–231, 233, 245–247, 251, 253,  
 260–262, 275  
  
 philosophy as a, 45, 46 n. 46, 66, 224–226,  
 230, 246, 253–254, 275  
 novel, 227, 274  
 -specific representation, 2, 13, 27–29, 95, 148,  
 200, 249 n. 30  
 world taken for a, 17, 262  
 Nichols, Shaun, 188 n. 2, 192 n. 8, 207 n. 38  
 Nietzsche, Friedrich, 9 n. 10, 218, 222 n. 3,  
 254 n. 42  
   straight readings of, 259, 266  
 Nixon, Richard, 226 n. 11  
 noncognitivism, 101, 104 n. 5, 107–108, 117 n. 31,  
 255 n. 43  
 normativity, 10, 16, 103–104, 107, 110, 112, 114,  
 117–121, 125 n. 44, 153  
 non-natural property, 106–107, 118 n. 32  
 Norwich terrier, 59  
 Nozick, Robert, 119 n. 36, 131 n. 9, 266 n. 55  
 NP-completeness, 115 n. 28  
  
 obscenity, 83  
 observation, practical, 80 n. 4, 258–259, 267  
 oobleck, 200–201  
 opacity, 140 n. 29  
 Open Question Argument, 107 n. 12  
 ought, 10, 14–15, 66, 82, 85 n. 17, 89, 103–107, 111,  
 113–114, 116–119, 121, 122 n. 37, 123–124,  
 126–131, 134, 137–139, 145–151, 153, 211  
   all things considered, 111 n. 19  
   central vs. penumbral, 105, 110, 112, 117, 131,  
   138, 140, 151–152  
   cognitive function of, 141–143, 146–151  
   epistemic, 103 n. 3, 124 n. 43, 127 n. 1  
   *just*, 105, 110, 117 n. 30, 126, 128, 131 n. 9,  
   140, 143  
   objective vs. subjective, 111 n. 19  
   sovereign, 138–140, 143  
  
 pain, 164–165  
 Parfit, Derek, 128 n. 3, 218  
 particularism, 117 n. 30, 128  
 penguin, 223  
 Perp Theory, 242 n. 21  
 personal identity, 236, 240, 242, 245 n. 25, 251,  
 260–261  
 phenomenology, 123–124, 129, 144, 199, 205,  
 207–208, 263  
 philosopher-king, 229–230  
 philosophy  
   analytic, 3, 5, 9–10, 21 n. 1, 59, 78, 99–100, 105,  
   126, 139, 153 n. 53, 184, 188, 209, 218 n.  
   55, 225, 267, 270, 279, 281  
   apriorist, 9, 11–12, 17  
   experimental, 5 n. 5, 33 n. 22, 87 n. 21  
   history of, 116

- ordinary language, 152 n. 53, 196 n. 17
- problems of, 19
- subspecialized, 3–4, 270–271, 273, 276–277, 280
- physicalism, 21 n. 1, 96–97, 168, 171–172, 198
- Pitldown Man, 57–59, 65, 69–73, 75–76, 92–94, 99–100, 200–201, 220, 274
- plans, 123, 235, 238, 247
  - stability of, 238, 240, 245, 248 n. 29, 251
- Plato, 9, 19, 52 n. 51, 54, 122 n. 37, 190, 216, 221, 229–230
- pleasure, *See* hedonic response
- policy, 199, n. 25, 206, 208, 220, 235, 238–241, 247, 262, 264–266
  - Bratmanian, 234, 238–241, 243–244, 246, 248, 250–252, 254–257, 266–267
  - covering all bases, 249
  - fantasy, 241–242
  - higher-order, 238, 242 n. 21
  - reason-determining, 238
  - reflexive, 238, 241, 248, 250, 253
- politician, 33 n. 22, 226
- positivism, 101–119, 153 n. 54, 192
  - logical, 22, 102, 115 n. 28, 152 n. 52, 280
- possibility, 118, 157–163, 167, 186 n. 45, 188–189, 191, 192 n. 8
- possible worlds, 10, 14 n. 19, 155–163, 186–188, 191, 209 n. 42
  - ersatz, 169 n. 22
  - nearby, 159, 165–167, 177, 180
- practical identity, 234, 263
- pragmatism, 43 n. 43, 139 n. 28
- preference, *See* desire
  - lexical, 109 n. 16, 238 n. 13
- prescriptivism, 107
- procedural utopia, 23, 25
- productivity, 276
- project, 79–81, 92, 234
  - ground, *See* desire, categorical
- projectivism, 107, 108 n. 13
- proposition, 24, 122 n. 37, 130 n. 8, 140 n. 29
- protoarithmetic, 189, 194–195, 202–205, 212 n. 45, 213–215
- psychologism, 24–25, 33, 42, 43 n. 43, 79 n. 2
- psychological continuity, 240, 260–261
- public arena, 29–30
- publication, 276
- Putnam, Hilary, 153 n. 53, 162 n. 16, 164–165, 174, 266 n. 55
- Puzzle of Positive Metaethics, 101, 105, 107, 109, 117
- quasi-policy, *See* ideal
- quietism, 112, 114, 125 n. 44
- Quine, W. V. O., 11 n. 13, 98, 266 n. 55
- rationality
  - bounded, 65 n. 21, 106 n. 10, 131, 136, 252
  - theory of, 7, 43, 54
- Rawls, John, 118 n. 32
- reasoning, 24
  - practical, 7–8, 54, 59–60, 63, 65, 72, 82, 101, 105–106, 109–110, 112–113, 117, 124, 128, 132–137, 144 n. 35, 153, 234, 248
  - theoretical, 13 n. 18, 80 n. 4, 82, 116 n. 29, 119 n. 33, 267–268
- reduction, 55, 140 n. 30
  - behaviorist, 155, 164–165, 174
  - modal, 155–163, 165–168, 182, 184–185, 197
  - phenomenalist, 155, 158 n. 3, 162 n. 16, 164, 167, 190
- reflective equilibrium, 10–12
- relativism, 42, 77 n. 1, 89, 96 n. 32, 111 n. 19
- Richard of St. Victor, 147 n. 42
- right, 103, 106 n. 10
- rule-based systems, 249
- Russell, Bertrand, 139 n. 28
- sand rat, 146–147
- satisfaction, 239, 242–244, 248
  - genetic objection to, 244
  - motivation for, 242–243
  - Romantic objection to, 244, 257–258
- science, 27, 39
  - ethics distinction, 77, 84–86, 95, 99
  - stages of, 102–104, 106
  - unity of, 84, 278
- scientific revolution, 252–253
- scientism, 96–97, 98 n. 34
- screening, 142–146, 148, 152–153
- secondary quality, 84, 216–217, 218 n. 55
- selection,
  - group, 53 n. 52
  - meme, 57, 60
  - natural, 55–57, 70–72, 141 n. 31, 244; *See also* evolution
- semantic analysis, 5, 10–12, 87, 115–116, 126, 138–141, 143 n. 34, 151–152, 156, 188, 193, 195 n. 16, 206, 220
- sense data, 115 n. 28
- Shelf Life Problem, 254
- Shore, Dinah, 93
- should, *See* ought
- singer-songwriter, 67
- skepticism, 45, 87 n. 21, 114, 120 n. 35, 163 n. 16, 255 n. 43
  - moral, 221–222, 225–228, 230
- slack, 239 n. 17, 254 n. 40
- slide rule, 71
- Smith, Adam, 9 n. 10
- Socrates, 9 n. 10, 43



- Sokol, Alan, 32
- species, 26, 55 n. 2, 74 n. 35, 208, 218, 260  
     specialized vs. weedy, 55–59, 222–223, 251
- species form, *See ergon*
- squirrel, 74–75
- standards, 27–29, 60, 62–63, 65, 71–72, 85 n. 17, 148, 222, 233, 260  
     for argumentation, 22–23, 25, 28, 32–34, 42–43, 95, 148, 224, 226, 278  
     niche-specific, 2, 27, 45–47, 50–51, 66–69, 75, 224–227, 228 n. 15, 229–231, 245–246, 249 n. 30, 274–275, 279, 280  
     distinguished from priorities, guidelines, 224  
     military, 226–227
- state-of-nature argument, 87–88, 92, 126–127, 141 n. 31, 148, 151 n. 51, 189, 194–195, 202–205, 213–215
- Stoicism, 81 n. 6
- subjective motivational set, 79, 83–85, 91–93, 263
- substitution license, 139, 140 n. 29, 153 n. 53, 219 n. 60
- supervenience, 96, 156, 170–171, 182–185
- Taylor, Frederick Winslow, 48 n. 47
- Tenn, William, 59
- testimony, 30, 33, 36, 39, 42, 44–45, 271  
     practical, 39
- theater, 225
- thick ethical concept, 83, 85–86, 89, 91, 93, 95, 103 n. 3, 106 n. 10
- Thompson, Michael, 74 n. 35
- traffic management, 12 n. 16, 149, 206–207, 210, 212, 231
- tragedy, 82, 85, 111 n. 19, 150 n. 47
- transcendental deduction, 115, 148, 209 n. 42
- truth, 20, 87–89, 96, 119 n. 33, 139 n. 28, 140 n. 30, 146, 150, 214
- deflationist theory of, 141–142  
     partial, 2, 13 n. 18, 20, 28, 88, 90, 95, 98, 122–123, 218 n. 55  
     prosentential theory of, 142–143  
     redundancy theory of, 130 n. 8, 143
- universal, 168–170, 185
- universalizability, *See* CI-procedure
- university administration, 44–48, 276–277
- utilitarianism, 23, 25, 63, 72–73, 75, 78, 80–82, 86, 94 n. 29, 101–104, 108 n. 13, 111–112, 135 n. 20
- utility, 111  
     expected, 176  
     function, 93, 175–176, 178
- value, 65, 73, 115 n. 28, 237, 239 n. 17  
     incommensurable, *See* ends,  
         incommensurable
- valuing, 240–241, 244
- VaR (value-at-risk), 28 n. 14
- Velleman, J. David, 81 n. 6, 235 n. 5, 237
- virtue, 3, 88, 93
- Vogler, Candace, 251 n. 34
- Voltaire, 21 n. 1
- Watson, Gary, 237
- weakness of will, 119 n. 33
- Wedgwood, Ralph, 116 n. 29
- Wellhausen, Julius, 21 n. 1
- Whewell, William, 2, 27, 139 n. 28, 198–199, 219
- Wilde, Oscar, 218
- Williams, Bernard, 8, 15, 77–100, 105 n. 9, 132–133, 134 n. 17, 150, 226 n. 10, 261 n. 52, 266 n. 55, 267 n. 58, 274 n. 9
- Wittgenstein, Ludwig, 19, 153 n. 53, 205 n. 34, 206
- Wooster, Bertram, 41