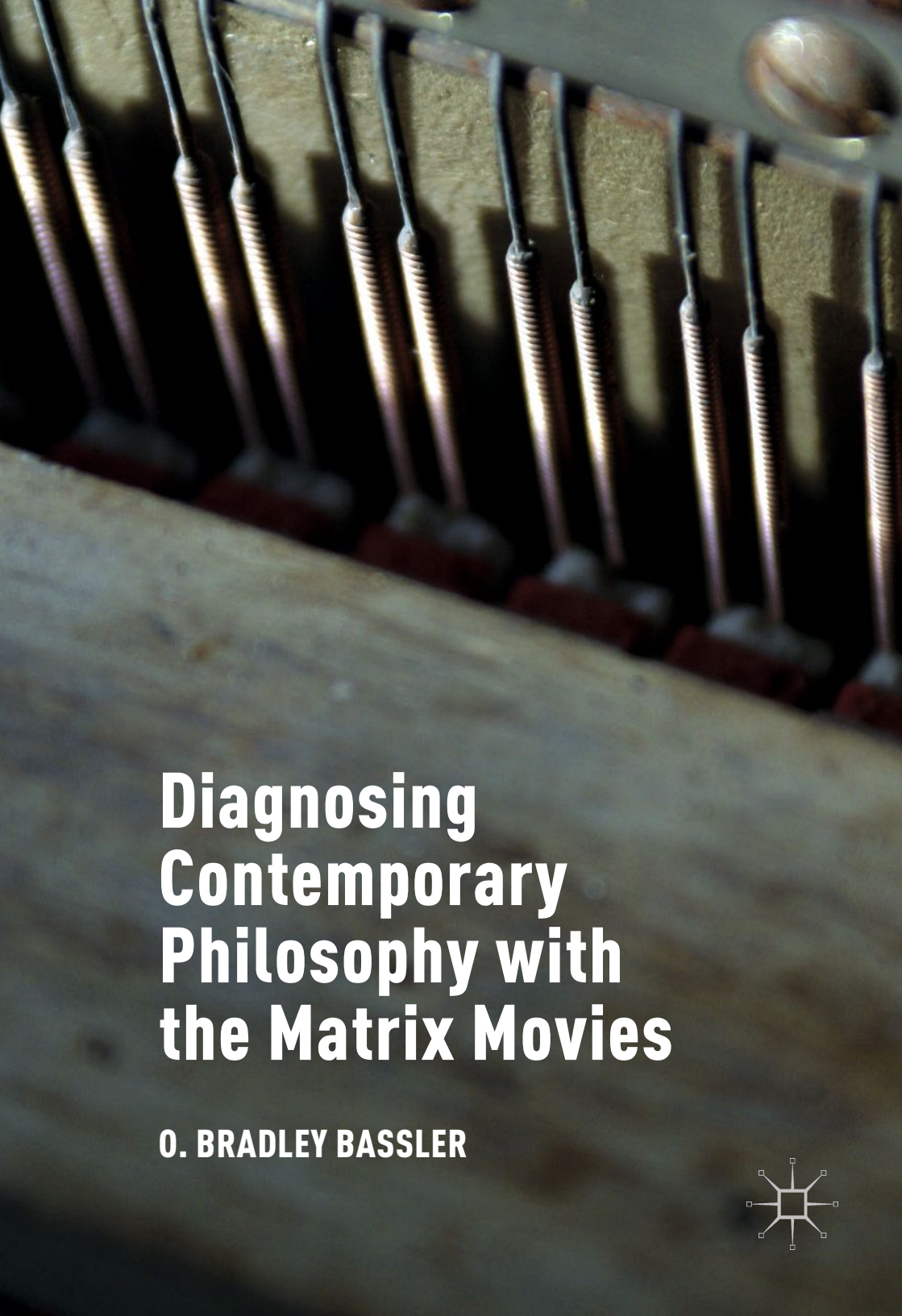

Diagnosing Contemporary Philosophy with the Matrix Movies

O. BRADLEY BASSLER

Diagnosing Contemporary Philosophy with the Matrix Movies

O. Bradley Bassler

Diagnosing Contemporary Philosophy with the Matrix Movies

palgrave
macmillan

O. Bradley Bassler
Athens, Georgia, USA

ISBN 978-1-137-57888-4 ISBN 978-1-137-57889-1 (eBook)
DOI 10.1057/978-1-137-57889-1

Library of Congress Control Number: 2016962056

Printed on acid-free paper

This Palgrave Macmillan imprint is published by Springer Nature
The registered company is Macmillan Publishers Ltd.
The registered company address is: The Campus, 4 Crinan Street, London, N1 9XW, United Kingdom

for Zoe Lalene Brient, and the future

Acknowledgments

I would like to thank the many students and former students who indicated interest in this project along the way. I owe especial thanks to Ricardo Abend Van Dalen, who helped me through a difficult transition; without his encouragement, this project might not have seen the light of day. Thanks also to my editors at Palgrave Macmillan, Esme Chapman and Grace Jackson, for all their help. My family has, as ever, been a constant source of support. This book is for my daughter, Zoe.

Contents

1

Introduction

1 What This Book Is and Isn't

This book is intended as a manifest for a larger, still outstanding project called Matrix Philosophy, which will be written according to the specifications introduced here. Hence, my shorthand for referring to the current book in the course of preparing the larger book will be to call the former the Manifest. The larger project, Matrix Philosophy, which mirrors the progressive levels presented in this Manifest, will consist of four parts, currently called "An Invitation to the Parafinite," "Paraphysics," "Philosophical Praxis," and "Philosophia Perennis."

A manifest, in addition to being a list of cargo or passengers, is a fast freight train, usually carrying perishable goods. Both definitions fit: the content of this Manifest (freight) is a manifest (roster) whose content, in turn, will lapse at the production of the completed project. In this Manifest, I conduct a diagnosis of contemporary philosophy, framed by the *Matrix* movies. The diagnostic work is itself philosophical, and so it yields philosophical fruit in its own right. Hence, this Manifest, oriented by its diagnostic intention, delivers a full specification for the contents of Matrix Philosophy to come. Accordingly, the diagnosis of contemporary

O.B. Bassler, *Diagnosing Contemporary Philosophy with the Matrix Movies*, DOI 10.1057/978-1-137-57889-1_1

philosophy conducted in the Manifest generates only the rudiments of the Matrix Philosophy project. Matrix Philosophy is a "machine" still under construction.

2 What Matrix Philosophy Is and Isn't: A First Descent Down the Rabbit Hole

The Matrix Philosophy project is oriented toward answering the question "What is the Matrix?," which is asked in a different way and a progressively deeper form at each diagnostic level of the Manifest. Although this is a guiding question for the trilogy of *Matrix* movies, Matrix Philosophy itself is not a philosophy "of" these movies, nor is it an attempt to specify what the answer to this question is according to these movies. My project is not to give a reading of these movies, in that sense. But I draw upon them as a source of inspiration. I will, for example, be saying what I find right and wrong (helpful and unhelpful) in their approach to the Matrix, and more relevantly perhaps, what limits are imposed on a treatment of the Matrix by the sort of blockbuster Hollywood movies that comprise the *Matrix* trilogy. This is not intended to denigrate what these movies accomplish, however; in my view, what they achieve is indeed something rather special. In fact, my treatment here is intended to indicate how contemporary popular cultural resources can serve philosophical work. The point is simply this: the *Matrix* movies are about the Matrix, and so they shed light on Matrix Philosophy, but Matrix Philosophy is *not* about the *Matrix* movies. In short, the *Matrix* movies are an important example for Matrix Philosophy to consider. Although the movies are our constant companions in this Matrix enterprise, I give individual diagnosis of contemporary philosophy and provide my own answer to the question "What is the Matrix?" Part of the challenge of dealing with these movies within the context of Matrix Philosophy is that, while the movies accompany us along the way, their most interesting contribution to Matrix Philosophy occurs at a rather deep level—that of myth. While plenty can be said about the treatment of technology in these movies, their more fundamental contribution is to the mythology they cultivate.

However, this mythology is syncretic, and so it needs to be handled with care; hence, this issue is deferred to Chap. 7 of the Manifest.

An approach to Matrix Philosophy can be organized heuristically, as a multitiered enterprise. At each level, the contemporary philosophical condition is diagnosed, with the question "What is the Matrix?" serving as our guide. Yet, even this progressive diagnosis cannot be comprehensive, just as a physical examination by a physician can never really be truly comprehensive. That's because the nature of a diagnosis is simply to target specific functions by identifying symptoms of an underlying condition. It is in this sense that I refer to this work as a diagnosis. Here are the levels we will encounter:

Level 0. (Level of Cognitive Science) At the surface, we have the stratum where philosophy meets cognitive science. Here, the question is: "Could the Matrix (now in the sense, or roughly in the sense, of the *Matrix* movies) be implemented?" When the movies were first released, a philosopher wrote an article in which he gave as the answer to this question that it was not only feasible but, under certain assumptions, even overwhelmingly likely. Such "optimism," we must note, is the very propaganda that keeps the cognitive science community afloat, giving it its metaphysical "swooshes" (Bostrom 2003).[1] The first section of the *Manifest*, then, outlines why what we might call the "theology of cognitive science" (not cognitive science itself) is misguided. As an excursus, I consider what I call, following Edgar Wind, the *problem of art and anarchy*, allowing for a complementary perspective opened up by the arts. This complementary consideration serves several purposes, among them to show that there are balancing questions in the domain of the arts, but also that these balancing questions can ultimately be subsumed under the same diagnostic condition.

Level 1. (Level of Logic) A level down, but still fairly close to the surface, is the *logical* level. It is at this level that the debates within the philosophical community about brains-in-vats have transpired. Considering these debates allows us to diagnose the role played by appeals to logic in one canonical contemporary philosophical context: "Could I really be a brain in a vat?" Here, the question is not whether such a scenario is *feasible* but whether it is (logically) *possible*. It is the possibility of such a scenario that makes it logical. The underlying questions that this

level of investigation inspires move us beyond the feasibility arguments of the sci-fi–inspired cognitive scientist. Although I will not do so, an analogous presentation could be made about another area of discussion from another philosophical tradition—one which apparently inspired the Wachowskis, the creators of the *Matrix* movies. This would center on concepts of *simulacrum* and *simulation,* and the canonical work here is Jean Baudrillard's book. According to Baudrillard, the Wachowskis didn't really understand his book at all (Watson 2003, 162). (Since I am not providing a philosophical reading of the *Matrix* movies, that doesn't matter.) The similarity of the being a brain-in-a-vat question and that of simulation (in Baudrillard's sense, not in the cognitive scientist's sense) may be obvious enough to the reader, but if it isn't, no matter. I merely suggest that the two inquiries are parallel and that both belong at this level; further, I argue that the former belongs at this level. And neither is it my goal to "do justice to" or "solve" the brains-in-vats "problem." I use it only to diagnose a particular level of Matrix Philosophy. (I am saving my philosophical strength for the heavy lifting that goes on deeper down this rabbit hole.)

At this point, the Manifest divides into two versions: the Theatrical Version and the Director's Cut. The Theatrical Version is recommended for a first reading, and so presents a quick and dirty (axiomatic) argument that takes us, in one step, from Level 1 to Level 2. In the Director's Cut, the long and arduous path between these two levels is spelled out in detail. It's a transition that occurs at what I call Level 1.5, or the Level of Transit. This Level 1.5 is sequestered in the Appendix, but I subsume a schematic description of Level 1.5 within the description of Level 2.

Level 2. (Level of the Parafinite) The debate about brains-in-vats is supported by a result that straddles the Level of Logic and the Level of the Parafinite. Much twentieth-century philosophy was oriented toward a revolution in the foundations of mathematics that occurred in the late nineteenth and early twentieth centuries, generating a result that focuses both the strengths and the weaknesses of this revolution. This is typically known as the Löwenheim-Skolem Theorem (hereafter, "L-S Theorem"), residing at Level 1.5 (Level of Transit). Hilary Putnam, in an article on this theorem, says his thinking about it inspired his argument about brains-in-vats. I focus on this article to show how the theorem is

connected to the debate about brains-in-vats. Most important, however, is the transportation this theorem gives us to move from the Level of Logic to the Level of the Parafinite. The L-S Theorem is typically interpreted as demonstrating a relativity in the distinctions between various levels of the mathematical infinite. If so, there is something indefinite in the mathematical concept of the infinite: what level of the infinite we "see" depends on our point of view.

Following the consequences of the L-S Theorem thus prepares us for the task of overhauling the very idea of supplying mathematics with a "foundation"—indeed, this is the major focus of Level 2 proper. The key result is that mathematics, viewed as living within the Matrix, cannot be grounded in any collection of fundamental objects, but only in a Web of Proof. This Web of Proof has the curious feature of being not a typically finite or infinite sort of thing but, rather, somehow in between the two, and so it is what I call *parafinite*. This is where the Director's Cut joins back up with the Theatrical Version. And it is here that we encounter the first controlled experience of the "Matrix feeling," which deepens progressively at each successive level.

Level 3. (Level of Paraphysics) At this third level down, we assess the more general philosophical implications of revising one of the basic conceptual distinctions traditionally undergirding the philosophical enterprise: that between the finite and the infinite. Work at Level 2 introduced the idea that the distinction between the finite and the infinite could be reconsidered in terms of the parafinite. Here, at this level, a whole host of questions present themselves—so many, literally, that at first it is not clear how to proceed. But we may at least indicate the dominant concern through analogy. If, at the Level of Cognitive Science (Level 0), the question "What is the Matrix?" takes the form "Could we feasibly live in the Matrix?"; and at the Level of Logic (Level 1) it is "Could we possibly live in the Matrix?"; while at the Level of the Parafinite (Level 2), it is "Could we even secure mathematics in the Matrix?"; then the question at the Level of Paraphysics (Level 3) is: "Is the Matrix parafinite?" The answer, suitably construed, is yes. In particular, the Matrix, as in the special case of the Web of Proof previously considered, is neither finite nor infinite in the traditional senses of these terms (Bassler 2015).

Saying what this question and response mean will, of course, take some work, but already the reader may notice that the new question takes a different form from the previous versions of the underlying question of "What is the Matrix?" This one is indicative of a philosophical movement taking place as we move deeper down the rabbit hole. The answer to the question at Level 0—"not likely"—indicated that we hadn't really begun to consider the question. The answer at Level 1—"If this is just a logical possibility, who really cares? And if it isn't a logical possibility, who really cares?"— indicated that we had still not touched what really concerns us. At Level 2, if mathematics can only be secured in the Matrix as a Web of Proof, then this answer remained pretty esoteric. At Level 3, however, the problem starts to grow some teeth: the Matrix is neither finite nor infinite. *It is not a typical sort of thing.* At this level, Matrix Philosophy starts to get interesting. And this level of Matrix Philosophy I have called *paraphysics*.

Level 4. (Level of Philosophical Praxis) The move to the next level down is much more difficult (we might consider the issue of sacrifice) because it involves shifting from philosophical theory to praxis, and in particular, we must confront the question what it means to *do* philosophy. At this level, I assess the praxical results of opening the perspective at Level 3. In the *Matrix* movies, Levels 1 through 3 figure Neo's taking the red pill, "waking up," coming unjacked from his pod, and sliding down the tube; Level 4 is the point at which Neo then lands in the soup. *What the hell do I do now*? The first answer is that you had better hope there are some helpful people around. The move down to Level 4 in the *Matrix* movie is brought about by the *deus ex machina* of the red pill. No real explanation is given in the *Matrix* movies, and because it is a movie maybe none can be given. Much of the first *Matrix* movie involves the question "Did I make the right choice in choosing the red pill?" It's a reflection on the philosophical conundrum of free will versus determinism. Later in the *Matrix* trilogy, the debate is set in terms more influenced by the human/machine dichotomy and the question of control. But at this fourth level, the question for Matrix Philosophy is "How may I, and how ought I, act in the Matrix?" Notice that this question presupposes that we *do* inhabit the Matrix. This may come as a surprise to the reader, especially after my insistence that we are not living

in a simulation. As we descend deeper down the rabbit hole, though, the conception of the Matrix changes and deepens, too—well beyond the notion of simulation that we find in either philosophy or the *Matrix* movies. The problem is ultimately not to escape the Matrix, but to learn to inhabit it *mindfully*. As we do so progressively, the referent of the term "Matrix" changes accordingly. This fourth level of Matrix Philosophy I have called *philosophical praxis*.

Level 5. (Level of Work on Myth and Philosophy) No doubt a thorough description of Matrix Philosophy would require indefinitely many (or even better, parafinitely many) more levels, but in the brief synopsis of Matrix Philosophy I provide in the Manifest, Level 5 is the final and deepest one we reach. It is, in fact, at this level that the *Matrix* movies prove most helpful but also most perilously tempting. It is here where philosophy intersects with myth. The *Matrix* trilogy gives us a highly syncretic mythology, in which the Matrix is juxtaposed with a "real" world, at the center of which lies the proverbial Zion. This picture is, indeed, a "deep" presentation of the "Matrix problem," and this is what gives the movies their lasting power. But the picture is a *mythical* picture, not a philosophical one, so it cannot give us a direct answer to the philosophical question "What is the Matrix?"—although it does provide a mythological one.

In fact, for reasons too complicated to spell out yet, it is precisely the competition between mythology and philosophy that drives this level of Matrix Philosophy. The question here is: "Is the Matrix susceptible to philosophical characterization or can it only be characterized mythologically?" To put the question more economically, we may ask: "Is there a Philosophical Matrix?" This question gets to the heart of Matrix Philosophy. The viability of Matrix Philosophy thus constitutes an overhaul of the philosophical condition in terms of the progressive diagnosis of contemporary philosophy and its intersections with cognitive science, logic, the foundations of mathematics, the finite/infinite distinction, praxis, and mythology. It is at this endpoint of the Manifest that diagnosis and Matrix Philosophy converge. Building on the work of Hans Blumenberg, we call this final stage the Level of Work on Myth and Philosophy, and it is related to the tradition of *Philosophia Perennis*.

Here, I address this final (for us) level of Matrix Philosophy in three stages. First, I compare the mythology of the *Matrix* trilogy to Gnostic mythology; I was first motivated to do this by a paper by my then-student Tom Ball. Gnostic mythology poses a particular challenge to philosophy (standing as it does in close proximity to the Neoplatonist tradition), and there are considerable alliances between *Matrix* mythology (of the movies, that is) and Gnosticism. The *Matrix* mythology is not Gnostic in any consistent way, but contra Will Stevenson, the *Matrix* movies *are* about Gnosticism (and other mythologies, too). The elements of Gnostic mythology one finds in the *Matrix* movies are important, but the departures are equally interesting.

From there, I proceed to consider what this comparison can tell us about Blumenberg's project in his book *Work on Myth*, and how this comparison opens up perspectives on the tradition of *Philosophia Perennis* in its historically mutating (and evolving) forms (Bassler 2012; Schmidt-Biggemann 2004). Ultimately, another important locus of consideration would be the notion of seeing conditionality in the Buddhist tradition and, in particular, what Nagarjuna may point to regarding this issue—but I'll leave that problem to Sonam Kachru. At this deepest level, a (provisional) final answer suggests itself: *The Matrix is the locus for the perennial confrontation of philosophy and myth*.

My goal in the Manifest, then, is to chart a compact for future philosophical investigation, and nothing would please me more than to see readers of this book extend it in their own, individual ways. But before we begin the Theatrical Version, which is to say the Manifest itself, here's a "teaser trailer" for the curious reader. This feature offers a preview of the thematic connections between various pieces of the story. If you tend not to like trailers, skip straight ahead to the "movie."

3 Teaser Trailer: A Thematic Guide

A movie trailer is supposed to offer an enticing hint of what's to come and make you want to see the movie—or in this case, read the Manifest. I've already told you, in a way, what's to come, but here's an additional guide that is more thematically oriented. How can I entice you? With one

particular way to frame the thematic content of this book. The book is thematically multidimensional, so this frame captures just one view, and in that sense, it's an advertisement. By the way, have you ever noticed that trailers sometimes have footage that didn't make it into the movie? That happens a bit here, too.

Consider the question "What can you do with an example?" The Level of Cognitive Science (Level 0) starts with a comparison of two particular arguments: How does Lee Smolin's argument in evolutionary cosmology compare to Nick Bostrom's argument about living in a simulation? These two arguments drive us in the direction of the question "What does it mean to say that an argument involves speculation?" Comparing the two arguments helps us uncover the places (often hiding) where speculation "hangs out." The chapter moves toward a crux of speculation in each case, and allows us to see the similarities and differences in the arguments. The general structure of my argument at Level 0 is that Smolin's and Bostrom's arguments are of the same form, but that Bostrom's is way more speculative than Smolin's because of how the premises are factored into his argument. In particular, what I call the "wetware equals dryware" premise (discussed in Chap. 2) is way, way speculative. If you use a way, way speculative premise, it's not surprising that you can derive a way speculative conclusion, and that's the sum of what Bostrom does.

At the Level of Logic (Level 1), Hilary Putnam considers a way speculative analog of Bostrom's scenario that we are living in a computer simulation—that is, we are brains-in-a-vat. Unlike Bostrom, Putnam argues against his analog, saying that we can't in any reasonable sense be brains-in-a-vat. He holds that the *novelty* of his argument lies in claiming it does not appeal to the *physical* impossibility of the scenario but, rather, to an antecedent *logical* impossibility. This is important for Putnam, because it suggests that philosophy has something to accomplish that can't be accomplished by, say, physics. But the problem with Putnam's argument is that to rule out the way speculative scenario of brains-in-vats, Putnam has to parse lots of really wild examples of what he claims is *logically* possible without their necessarily being *physically* possible. And the basic problem is that using these examples to draw this distinction itself seems way speculative. Putnam has ruled out something that is way speculative on the basis of way speculative examples used to support his distinction

between logical and physical impossibility. That makes the distinction between logical and physical possibility way speculative as well, and so Putnam has disproved something that is way speculative on the basis of a way speculative distinction.

So, what's logically possible? A traditional answer would be what isn't logically contradictory. But what is the relation between being logically contradictory and being physically incoherent? At the Level of the Parafinite (Level Two), we sample a conversation between Wittgenstein and Turing that circles around the question "If mathematics contains contradictions, will bridges fall down?" Turing basically says yes, Wittgenstein says no. I suggest an example of my own: that our omnipresent experience with computer malfunction, historically postdating Wittgenstein and Turing's conversation, intimates that the Wittgenstein-Turing debate may be inadequate, or perhaps even mistaken. But if this is so, then we have to ask: "What was it about the way Wittgenstein and Turing were thinking about mathematics that was correct and what was mistaken? And what implications does this have for how we should think about mathematics?" What this leads us to—perhaps surprisingly—is that mathematics is a kind of Matrix we are "living in." But the sense of living in a Mathematical Matrix—what I call a *Web of Proof*—is very different from the way speculative senses of Matrix that Bostrom and Putnam argued for and against, respectively.

This Web of Proof is something that involves all of us any time we're involved with mathematics. And so, this gives us our first intimation of what I call Matrix Philosophy. It is framed by the example of our use of computers, and this example is way different from the examples of living in a computer simulation, or brains-in-vats, or the wild examples of things that are logically but not necessarily physically possible. This computer example is drawn from contemporary experience—not just any contemporary experience, but also from the canonical experience that generates the sort of speculation we've encountered in Bostrom (obviously) and Putnam (I suggest). The general thesis here is this: *to diagnose contemporary philosophy, our best source of examples is contemporary experience.* This may sound obvious enough, but doing this is intensely philosophically challenging. It is a central problem of what I call *philosophical praxis*.

Focusing on the Web of Proof, we track the question "What sort of matrix is the Web of Proof?" Proofs are curious for two features in particular, both of which exemplify essential aspects of counting. First, it is vague when we have two *different* proofs of the same thing, versus different *versions* of the same proof; and if we are to count proofs at all, we have to resort to extreme methods of formalizing those proofs in order to disambiguate them. Doing so, however, leaves us with many distinct formal proofs that we then identify in any practical, informal context; and so counting proofs means that we inevitably overcount them. Second, if there is any reasonable sense in which we can count proofs (or, to hedge our bets, "candidate-proofs"), it's pretty clear that the number of actually existing proofs is growing all the time: new proofs for new claims are constantly being offered. These two features—the need for disambiguation and the growth process of quantification—are not mere accidental features of proofs as "messy." (Proofs are supposed to clean up the mess, after all!) They are intrinsic features of quantification—a thesis I've argued for at length elsewhere (Bassler 2015). Indeed, these features are part of what I refer to as the *parafinite* nature of quantification. And in the particular case at issue here, this quantification can be expressed by saying that the Web of Proof is neither finite nor infinite but, rather, parafinite. At this level, the answer to the question "What is the Matrix?" is this: the Matrix is parafinite.

Thinking about the general philosophical implications of the parafinite leads to a new approach to philosophy in general, which I call *paraphysics* and which I sketch at the Level of Paraphysics (Level 3). This is the general philosophical orientation that results from the diagnosis of contemporary philosophy conducted at Levels 0, 1, and 2.

So far, all of this is like Neo's getting ready to take the red pill, but it is backwards in the sense that what I claim we need to descend into *is* the matrix we've been inhabiting all along. That is, I'm suggesting that our life so far has been a dream we haven't recognized as such; but waking up doesn't have to do with exiting the dream, in the way that Neo leaves a simulation. My earlier example of Bostrom's and Putnam's arguments suggests that this can't be *literally true* in any coherent sense. But what does the *fact* of Bostrom's and Putnam's arguments suggest? What philosophical examples do Bostrom and Putnam give us? They illustrate

how speculation is a kind of *philosophical dreaming*, and the diagnosis of this condition of speculation is a kind of waking up. So, what I call *paraphysics* at Level 3 is a philosophical program for disentangling ourselves from the speculation of philosophical dreaming. Paraphysics, then, is a wake-up call for transforming this speculation into something I call *visionary design*.

It's not as if the speculation simply needs to be "dispensed with"—Kant recognized that, philosophically speaking, we are inherently speculative creatures. Rather, we need to ask what the speculation is *doing*. That's the question that jolts us into philosophical waking and, ultimately, into action. Viewed theoretically, it's an impossible transition, reflected in a discontinuous, and intentionally troubling, break in the Manifest at this point. The "thesis" of the Level of Paraphysics (Level 3) is this: *The distinction between dreaming and waking is the bedrock, the concrete fact upon which our most fundamental philosophical distinctions are built.* At this level, the answer to the question "What is the Matrix?" is: the Matrix is paraphysical.

The transition from thinking as theory to thinking as doing brings us to the Level of Philosophical Praxis (Level 4). If philosophy has traditionally been contaminated by the theoretical impulse to speculate, where are we to look for examples to restore the balance? In his attempt to overcome the pejorative sense of philosophy as theory, the Russian philosopher and literary critic Mikhail Bakhtin turned to the literary genre of the novel. Here, he found just the sort of specificity that the language of philosophy abrogated, and so he turned to the history of the novel as a way of completing a praxical philosophy of action and value. What is most valuable for the purposes of this book is that, in his investigation of Dostoevsky specifically, Bakhtin diagnosed the role played by what he called Dostoevsky's "crisis-dream" in the genesis of his fictional composition, his *polyphonic* novel (discussed in Chap. 6). The crisis-dream is the linchpin in Dostoevsky's strategy for converting theoretical speculation into orienting vision, and so it links the *program* of paraphysics with the *genre* of the novel.

Now, all of a sudden, we have a whole history of examples to allow us to investigate the questions: "What does it mean to turn from dreaming to orientation in waking life?" "What does it mean to wake up and act?"

This transition from the general philosophical program to the advocacy of particular philosophical actions is that of paraphysics to philosophical praxis, and it is this transition that I treat at the Level of Philosophical Praxis (Level 4). Here, the answer to the question "What is the Matrix?" is: the Matrix is the Locus of Real Action.

With a philosophy of action in place, we are ready to move to the next level and to the book's ultimate example: the action movie in general, and the *Matrix* movies in particular. This is what I call the Level of Work on Myth and Philosophy (Level 5). What will be most important in thinking about these movies is the role played by what I call Matrix *Mythology*. Myths often give us answers to the same questions that philosophy attempts to answer—questions like those of origin, for example. So, what's the difference between myth and philosophy? Here is where the *Matrix* movies, which are our companion as we descend down the rabbit hole, assume their full force as the central example in this book. For the *Matrix* movies, I claim, present a *Matrix* Mythology that challenges us precisely, and especially, because it aggressively *incorporates philosophy in manifold ways*. By examining the *Matrix* Mythology, we can arrive at the deepest expression of the Matrix Philosophy I can achieve in this book. At this level, the answer to the question "What is the Matrix?" is: the Matrix is Work on Myth and Philosophy.

The ultimate tease of this trailer is that you have to follow it down the rabbit hole to find out just what this Work on Myth and Philosophy amounts to. Along the way, there are excursions ranging through the domains of cognitive science, logic and mathematics to poetry, the visual arts, and the movies. This range of experience is meant to accomplish a number of things, including showing that while the arts and poetry can broaden and concretize our appreciation of the matter, this alone cannot resolve the concerns raised by philosophical work oriented by cognitive science or logic and the foundations of mathematics. In some regard, the issues raised by the arts and poetry are mirror images of those raised by cognitive science and mathematics. Indeed, this "standoff" is reflected in the Wittgenstein-Turing debate: Turing speaks for the scientific establishment, while Wittgenstein's view is akin to the culture of the arts. On the one hand, by expanding our horizons we neither resolve our problems nor transform them. On the other hand, examples from the arts can

be more useful than examples from the sciences because, with the arts, we are no longer concerned with technical questions but, rather, with human ones.

Although the distinction between the technical and the human is somewhat narrow and artificial—after all, it is humans who create "technology"—it nonetheless shows how the sciences only address the human condition narrowly and the arts more broadly. One of the ultimate diagnostic aims of this volume (and diagnostic dimensions have largely been suppressed in this thematic trailer) is to restore the balance in contemporary philosophy that has been destroyed by our preoccupation with the sciences. We need to do philosophical justice to the sciences, and in fact, we will not be able to do so until we set them within the larger philosophical context of the human condition. Looking back to Bostrom, we see that the question "Are we living in a computer simulation?" ultimately requires diagnosis, rather than a direct answer ("No, you are not"), because the question itself is an allegorical picture of our contemporary philosophical condition, in which the modern preoccupation with the place of science within the human condition is paramount. It is this philosophical condition that the Manifest attempts progressively to diagnose. I'll leave this last claim as the ultimate tease in this trailer.

Note

1. Bostrom is a member of the Future of Humanity Institute and the philosophy faculty at Oxford University; when his recent book *Superintelligence* (2014) came out, it made the *New York Times* Bestseller List.

2

Are You Living in a Computer Simulation? (The Level of Cognitive Science: Level 0)

1 Are You Living in a Computer Simulation?

No, you are not.

About a fifteen years ago, the well-known physicist Freeman Dyson was invited to deliver a lecture at the University of Georgia. Along with Richard Feynman and a small number of others, Dyson was one of the primary architects of the theory called Quantum Electrodynamics (QED), paving the way for the Standard Model, which is still in some sense our fundamental theory of particle physics (Schweber 1994). Dyson started his career as a mathematician, and his visit to the university was in fact prompted by a nomination from a number theorist then resident in the mathematics department. I learned that, in addition to his general lecture to a large audience, Dyson would be meeting with members of the mathematics department in a more informal setting. Having just recently finished a PhD in mathematics, I was even more eager to attend this smaller event, and my enthusiasm was not disappointed.

In the small lounge at the university's mathematics department, faculty and students packed together to listen to what Dyson had to say. At what

O.B. Bassler, *Diagnosing Contemporary Philosophy with the Matrix Movies*, DOI 10.1057/978-1-137-57889-1_2

for me was the high point of the discussion, he remarked in an impishly casual tone, "You know what you have to do to establish the Riemann Conjecture? It's very simple." The room, dominated by number theorists, eagerly awaited Dyson's wisdom, in search of a route to this summit of mathematical problems. What they heard Dyson announce was that you "only" had to solve some other, equally appallingly challenging problem![1] But Dyson's "joke" led to an animated and productive discussion, though sadly for me it was almost all beyond my mathematical ken.

In his general audience lecture, Dyson was apparently even more bold. The main topic of discussion was the feasibility of mass transport from the earth to the moon. As I recall, Dyson argued that the energy cost per person transported should be about the same as what it takes to send someone by jet from Los Angeles to Tokyo. The audience was generally excited by this futurological speculation, which, in contrast, was rather uniformly bemoaned by my philosophical colleagues. The more substantive part of Dyson's presentation involved some passing remarks about gravity and star formation, alluding to topics he'd written about in more detail elsewhere.

Before Dyson's visit, I had learned from a generally circulated email that Dyson would have a small amount of extra time while on campus for individual conversations. While I expected that this opportunity would be eagerly snatched up by others much higher up on the food chain, I followed the dubious wisdom that you can't win if you don't play, and I responded that I'd be very pleased indeed to have a chance to speak with Professor Dyson. To my surprise, I was told that he had time at breakfast the day after his lecture and would be happy to have someone drive him back to the airport afterward. Dyson and I had a quick breakfast together at the Georgia Center, where he was staying, and then we set out for Atlanta. There were many aspects of our conversation I shall not forget, but one in particular is of concern here.

As we were nearing the city, Dyson told me of several speculative physical conjectures and drew by example a distinction between the sort that he liked and the sort he didn't. One he thought full of merit that had not been pursued was the conjecture by Cornell physicist Thomas Gold that life was in fact being produced regularly beneath the earth's surface as temperatures rose on the way down to the earth's core. At a certain

point along this inward path, Gold predicted, the conditions would be right for the production of low-level life forms. The prospect intrigued Dyson and he did not find it, at least, entirely farfetched.[2] In distinction, Dyson told me of a book I had not heard of at the time, in which a physicist proposed an evolutionary cosmology whereby our cosmos was "birthed" out of a black hole in a previous cosmology, following a process of evolutionary selection. According to this speculation, millions of such cosmoi were constantly being "birthed" deep inside the event-horizons of black holes, but most were spontaneously rejected owing to the instability produced by the specific value of their basic physical constants, which varied randomly from birth to birth. Dyson objected to this sort of proposal, it seems, because the mechanism for the birth of cosmoi out of black holes and the generation of cosmoi with varying physical constants were both simply posited as bare "possibilities." This sort of speculation was the kind Dyson objected to, or at least that was how I understood what he was saying.

At the time, I was driving through Atlanta traffic and certainly wasn't taking any written notes. Unfortunately I forgot the name of the physicist, forgot the name of the book, and the matter lapsed for some time. As it happened, it was a friend of mine working in the area of ancient Chinese history and philosophy who at least solved the mystery of the evolutionary cosmologist. He had read something about it, and he chased it down for me. The book was by someone less known then than now, named Lee Smolin; in the years to follow, I would come across his name and his work again and again. I even taught Smolin's book, *The Life of the Cosmos*, to undergraduates in an introductory Philosophy of Science course. It's great fun, but I came to sympathize with Dyson's distrust of the sort of speculation in which Smolin is engaged. At the same time, I've come to balance this distrust with another maxim: extreme problems generally require extreme solutions.

In 2001, shortly after the appearance of the first *Matrix* movie, the first version of a paper by Nick Bostrom was circulated and was titled "Are You Living in a Computer Simulation?" (Bostrom 2003).[3] What I argue here is that Bostrom is engaged in the same sort of speculation as Smolin, but that the particular instance of speculation Bostrom proposes is *much, much worse*. Herein, we have three tiers of speculative proposals:

1. Freeman Dyson argues that mass transport to the moon is reasonable.
2. Smolin argues for a speculative view of evolutionary cosmology, which Dyson finds objectionably speculative.
3. Bostrom argues for a conditional version of the view that we are living in a computer simulation.[4]

The transition from (1) to (2), as I employ it here, is based on an "argument by authority"—namely, an appeal to the wisdom of Freeman Dyson—and the reader is welcome to disregard this appeal to authority at his or her own peril.[5] I focus here on the connection between the arguments in (2) and (3) in an effort to catch the flavor of philosophical work that someone like Bostrom is doing. And I will argue that the status of Bostrom's *conditional* argument is much, much worse (i.e., much, much more speculative) than Smolin's proposal for an evolutionary cosmology. Before I get started, though, I want to make it absolutely clear that I am *not* interested in engaging in the conceptual debate that Bostrom's paper has generated. I will try to present Bostrom's paper as accurately as I can, but my interest is in the *structural form* of his argument.[6]

In its broadest outline, the strategy of Bostrom's project is rather predictable. This is not really surprising, because the sort of argument that Bostrom is making almost inevitably relies on ignoring the distinction between *dryware* and *wetware*. This distinction can easily be illustrated by example. If I look at the genetic code as a sequence of letters, each representing one of four options in the "spelling out" of a genetic word, then I am looking at the genetic code as dryware. If I look at a piece of DNA as a chemical substance in which a chain of four different components is repeated in a pattern, then I am looking at the genetic code as wetware. To explain the pattern of genetic reproduction and mutation, I can appeal to the dryware and talk about the spelling or misspelling of a previous word. To talk about replication as a process in which, for example, proteins fold and unfold in the process of replication, I need to talk about wetware.

I predicted that Bostrom would begin by assuming that consciousness is not something that depends in any intrinsic way on wetware, and this prediction proved correct. For short, I call this the *dryware equals wetware assumption*. Bostrom's shorthand is to call this "the

substrate-independence assumption" (Bostrom 2003, 2). He needs to enlist substrate-independence for the purposes of his argument, but if you look at the motivational description surrounding the argument, it's clear that he's committed to some version of *relative* substrate-independence. First, he is committed to the notion that there is some underlying material substrate. Second, and perhaps more important, his portrayal of what a material brain is doing is part of the larger picture he's painting. In particular, brain function is understood in terms of synaptic firing, and this synaptic firing is understood on analogy with the running of a computer program.

The version of substrate-independence Bostrom enlists is then exemplified by the remark that, on this view, whatever is relevant about the material composition of the brain will bear on consciousness only in terms of the contribution it makes to synaptic firing. Analogously, the material bearing of any substrate supporting the running of a computer program will make its contribution in terms of the program's execution. As Bostrom puts it, "if there can be no difference in subjective experience without there also being a difference in synaptic discharges, then the requisite detail of simulation is at the synaptic level (or higher)" (Bostrom 2003, 3). This means that the substrate-independence assumption allows us to assume the material substrate supports the execution of a computer program and is relevant *only* insofar as it contributes to the "firing" of this program. In this sense, wetware *does* equal dryware, because the only sense in which we are calling on the wetware is as support for the dryware.

On the one hand, if consciousness does depend on the specifics of wetware in some further sense, then it's hard even to imagine what it would mean to be conscious *in* a simulation (although in some sense we may consider it an abstract possibility). On the other hand, consciousness *of* a simulation, at least as I use the phrase, is so utterly unproblematic that we experience it all the time; it happens every time I watch a movie, for example.

If consciousness *doesn't* depend (intrinsically) on wetware, then the possibility is opened for consciousness to be reduplicated in some simulation medium where the same "words" of conscious experience are spelled using different but analogously functioning "letters." Here, the main—and perhaps even only—thing to say is that we have no way of assessing

the status of the dryware equals wetware assumption in any *realistic* sense, because we have no point of comparison. Whether consciousness, or a simulation thereof, could even be supported in a "silicon-based" life form is not something we have any real method for knowing how to assess (short of discovering a silicon-based life form). Indeed, given the wetware equals dryware thesis, neither life nor reproduction seems to be a necessary condition for consciousness, so the demand for such realistic assessment is effectively overruled.[7]

No doubt I can be accused of begging the question. How, for example, do I know that our current computers, or at least some quantitative improvement of them, are not conscious? Of course, I don't definitively know this. But I can equally well accuse someone who insists that I consider this possibility that this, then, is the case of begging the question. My point is that the fact that each of these accusations is possible demonstrates that we have no *method* for deciding. This breaks the symmetry between the two positions without definitively ruling out either of them. What it does show is that the *natures* of the two competing positions are different. If I make the assumption that Bostrom makes to get his project off the ground, I am doing something quite different, philosophically speaking, from not making the assumption.[8] I'm doing something *very*, *very speculative*. That's okay in some abstractly logical sense, but that it is *happening* tends to get overlooked.

Next, note that Smolin's evolutionary cosmology does not require any analogous assumption to get off the ground. While Smolin must posit that cosmoi are being birthed inside black holes, and he must posit that in this process there is random variation in the basic physical constants associated with these baby cosmoi, he does not need to make a speculative assumption that these cosmoi are identical under two radically different descriptions, one of which is a dryware description and one of which is a wetware description. This is so because the dryware "features" are read directly off the wetware description: these features just are the physical constants associated with the wetware properties of mattter. Nor does Smolin need to assume that the basic nature of the baby cosmoi will be independent of their wetware; in fact, the appeal to variations in the basic physical constants will serve precisely to distinguish different baby cosmoi *in terms of their wetware*. Smolin makes specific arguments

in his book (and in more detail in associated publications) about the way this variation in basic physical constants will affect the stability of the birthed cosmoi.[9] Because Bostrom must assume from the beginning that the distinction between dryware and wetware is irrelevant, he will be in no position to make analogous arguments. This is *much*, *much worse* (i.e., more speculative). We will see a little bit later what the postulates of birthing and variations in basic physical constants best correspond to in Bostrom's scenario.

As I have said, Bostrom's argument is stated in conditional form, and since my interest is in the general structure of the argument, let's not get bogged down in details. From a rhetorical standpoint, this conditionalization (or the original "trichotomy form") gives the argument a kind of high-tech, sophisticated ring. From the futurological standpoint, it's kind of disappointing, because Bostrom is arguing for much less than a direct yes answer to the question posed in his paper. But then, once again from the rhetorical standpoint, this may help make Bostrom seem cautious and responsible. Along the lines I've laid out earlier in this chapter, any such caution is misplaced and Bostrom's sense of responsibility is misguided.

Here's the bare-bones version of Bostrom's conclusion:

> This paper argues that *at least one* of the following propositions is true: (1) the human species is very likely to go extinct before reaching a "posthuman" stage; (2) any posthuman civilization is extremely unlikely to run a significant number of simulations of their evolutionary history (or variations thereof); (3) we are almost certainly living in a computer simulation. It follows that the belief that there is a significant chance that we will one day become posthumans who run ancestor-simulations is false, unless we are currently living in a simulation. (Bostrom 2003, 1)

The last line of this passage yields the conditonal implication of Bostrom's argument: if the antecedent is true rather than false, then we are currently living in a computer simulation. Note that from a conceptual viewpoint, the argument could lead to the paper's having any one of three titles: "Are Humans Likely to Reach a Posthuman Stage Before Going Extinct?", "Are Posthumans Likely to Run a Significant Number of Simulations of Their

Evolutionary History?" or "Are You Living in a Computer Simulation?" Rhetorically speaking, the fact that Bostrom chooses the third of these titles is unlikely to be an accident.

Understanding the conclusion of the argument requires that we understand what Bostrom means by "posthuman." Bostrom's most explicit characterization of the posthuman is in terms of his characterization of a "posthuman" civilization being one "where humankind has acquired most of the technological capabilities that one can currently show to be consistent with physical laws and with material and energy constraints" (Bostrom 2003, 4). As such, and as Bostrom acknowledges, our assessment of the notion of the posthuman is dependent on our knowledge of physical laws and material and energy constraints: it is *physical theory–dependent*.

Let's reformulate the argument in a way that, looking less rhetorically high-tech and cutting it down somewhat in scope, brings out its appeal to the futurologically minded. I omit some logical "niceties," and the reader may choose to believe I am aware of them or not, but I insist that the procedure of "calling out" on these grounds is the number one inhibitor of philosophical work. That is, after you learn to wash your hands, you don't go on to become compulsive; you practice reasonable hygiene and good judgment. Overcoming conceptual compulsiveness and forming good judgment are two of the diagnostic goals of this book.

So, if the human species is likely to attain posthumanity and is likely, in its posthuman guise, to run ancestor-simulations, then we are currently living in a simulation. Putting the argument in these terms brings out what is likely to strike us as odd about it. Let's phrase this in a way that resolves some of this discomfort so as to see where the essential discomfort might lie: if we are the sort of beings who are likely to attain to posthumanity, then we are likely ourselves already to be living in a (posthuman) simulation. In fact, Bostrom states this reduced version of his conclusion in almost the same form at the end of his paper: "Unless we are now living in a simulation, our descendants will almost certainly never run an ancestor-simulation" (Bostrom 2003, 14).

Now the relevant parallel to Smolin's cosmological argument is more readily articulated: if we are living in a *stable* cosmos, then we are likely to be living in a cosmos that has already been through a process of evolutionary

cosmological selection, and so we are likely to be living inside a black hole inside a black inside a black hole, and so on. The parallel statement is: if we are the sort of thing that has the stable possibility to develop into something that runs simulations, then that is most likely because we are *already in such a simulation*. Just as black holes in cosmoi give birth to new baby cosmoi that, if they are stable and therefore selected for survival, will themselves contain black holes that will themselves give birth to baby cosmoi, and so on, the suggestion is that (under the conditional assumption) we are the sort of thing that births simulations, and if those simulations are viable, will themselves birth more simulations, and so on. Thus, the likelihood that our cosmos, or our reality, will be "first generation" is vanishingly small. That is the basic structure of the argument.

Bostrom's paper goes on to introduce some equations that are not as intimidating as they might first seem (although it does turn out that nobody noticed some problems with them for quite some time) (Bostrom and Kulczycki 2011).[10] These equations capture some aspects of probability theory that allow Bostrom to nail down the "vanishingly small probability" that we are living in an original, as opposed to a simulated, reality (under the relevant assumptions). And that's basically the general structure of the argument.

So, what are we to make of it all? The main thing we should perhaps make of it is to determine what speculative lengths we can be driven to once we have the dryware equals wetware assumption under our belt.[11] But setting aside that major bone of contention, what we are faced with is the status of a particular sort of evolutionary argument, on the one hand, and the particular context in which that argument form is implemented here, on the other hand. Again, I will use Smolin's argument as a point of comparison.

First, we are demonstrably living in a cosmos with some rather long-term stability. Nobody knows, of course, about the long, long term, but that's beyond the structure of the argument in both cases anyway (the argument is not about infinite runs; it's about very, very long–i.e., parafinite–runs). Are we living—either in a simulation or a reality—in which we are likely to reach the posthuman stage at which we could run ancestor-simulations? Who knows? That's one of the reasons Bostrom has to present his conclusion in conditional form. Once again, this is much, much worse in the sense of being much, much more speculative than Smolin's (already very speculative) argument.

This may be as good a place as any to present a possible rejoinder. So what if the argument is much, much more speculative? Speculation is *fun*.

I'm not against anybody's getting his or her "swooshes" this way per se, but there are a couple of things to point out. The first is that Bostrom does everything in his power to make his argument sound as plausible as possible. The fine grain of the paper is devoted to two things: sharpening the argument and rendering subpremises plausible (though tellingly, there is no attempt to argue for the dryware equals wetware assumption). It is particularly his attempt to render subpremises plausible that makes Bostrom's paper the least speculative he can manage, given what he's arguing for. So, that sort of caution can be subsumed under the rubric of not being wantonly speculative. But rhetorically, the force of such detail is to draw attention away from the fact that the paper *is* wantonly speculative anyway.

This is a familiar rhetorical ploy of pseudo-science. If what Bostrom is doing can accurately be described as metaphysics, then the question of whether he is doing *good* metaphysics requires judgment. And patently, given the texture of his paper, what he thinks counts as good metaphysics is metaphysics supported by as much (pseudo-)scientific argument as possible. That then means that if what he's doing is metaphysics, it's (pseudo-)scientific metaphysics, and the charge of pseudo-science stands. Note that there may even be perfectly good science employed in the course of pseudo-science, and it can often be hard to separate what's decent from what's junk.

Even if it's all decent, the project is pseudo-science because of its *structure*. The fundamental assumptions, like the dryware equals wetware assumption, have no demonstrable *scientific* merit, and are adopted on independent *philosophical* grounds. Moreover, this fundamental assumption is presented in the paper only as "a common assumption," of which Bostrom says, "although it is not entirely uncontroversial, we shall here take it as a given" (Bostrom 2003, 2). That is, he's not offering any argument for it whatsoever. And the qualification "not entirely uncontroversial" is presumably a sociological rather than a conceptual one. I fail to see how, in a conceptual sense, the thesis that consciousness is substrate-independent—or indeed, even the opposite assumption—is anything other than entirely controversial. After all, we're talking about metaphysics here, not the consensus of some scientific community.

A more interesting form of this rejoinder would be this: we shouldn't restrict this sort of speculation no matter how outrageous it is, because something positive may come from such unbridling of our imagination. (Call this libertarian response the "Feyerabend" objection, if you like.) Indeed, if metaphysics is inherently controversial, doesn't such a restriction of speculation simply amount to a condemnation of metaphysics?

Certainly, nothing can rule out this objection, but how plausible is it? In fact, it's not all that implausible: there are even historical precedents to support this sort of objection. Jules Vuillemin wrote an interesting book in which he suggested that the medieval (attempted) proofs for the existence of God paved the way for thinking about the nature of the infinite in directions that contributed not just positively but even essentially to the later development of infinitary mathematics (Vuillemin 1971). Let's grant cases like the one Vuillemin proposes. Still, the objection only supports the point that there should not be ultimate strictures on metaphysical speculation. It does not do anything to tell us how to draw the distinction, which Freeman Dyson drew informally, between *good* speculation and *bad* speculation. Such matters of judgment are, of course, inherently controversial. The rebuttal, then, needs to come from another angle—and it comes by dividing the question. If the point is to promote radical thinking that may generate heretofore unconsidered concepts, no part of Bostrom's paper strikes me as radical in this sense. If the point is to contribute to thinking about whether we are or may be living in a computer simulation, Bostrom's argument has the status of being way speculative and pseudo-scientific.[12] That is, it is *speculative* without being *provocative*.

I think the point is clear, so I'm going to get off my hobbyhorse. Instead, let's look a bit more at the analogy of Smolin's and Bostrom's arguments. What are the analogs of Smolin's birthing and random-variation hypotheses? The comparison of the two arguments is not perfect—no analogy is—so there's no definitive answer to such a question. However, I suggest that the birthing hypothesis is analogous to the hypothesis that posthumans could (and to make the stronger form of the argument work, would) run ancestor-simulations. Each functions to get the evolutionary selection process going, although each works in a different way. In Smolin's case, the evolutionary selection will be according to random

variation in parameter, since he hypothesizes that black holes *just do* birth baby cosmoi. In Bostrom's case, the evolutionary selection works in terms of whether a given reality/simulation will attain the capacity to produce internal simulations (i.e., simulations in the case of a reality, simulations of simulations in the case of simulations, etc.).

Is there, then, anything analogous to Smolin's random-variation hypothesis in Bostrom's argument? There is. There needs to be a mechanism to ensure that some, but not all, of the possible ancestor-simulations will attain the capacity to produce internal simulations. That is, there needs to be a link between being a "good" simulation and being a simulation that attains the capacity to produce simulations internal to the simulation. One could establish this link by stipulating that a "good" simulation *just is* a simulation that will attain the capacity to produce internal simulations. But that's cheap in all possible senses, and Bostrom doesn't do that. He recognizes an independent sense of being a "good" simulation: a good simulation is one that "passes"—that is, that is not taken for a simulation by the participants. Why should a good simulation, in this sense, be one with the likelihood of attaining the capacity to produce internal simulations?

This question naturally breaks into two parts: First, what is needed to make a simulation that passes? Second, what sort of simulation is likely to attain the capacity to produce internal simulations? Only once we have answers to these questions will we be in a position to attempt to make a link between the two sets of concerns.

In addressing the first question, Bostrom focuses on the issue of how fine a grain is necessary to make a simulation passable. I won't go into the details, but Bostrom makes estimates about the computing power that would be required for such simulation, and in particular argues that the major computing cost of such simulations would lie in "simulating organic brains down to the neuronal or sub-neuronal level" (Bostrom 2003, 5).[13] According to Bostrom's estimates, this cost would be on the order of between 10^{14} and 10^{17} operations per human brain. (Note that these estimates are based on brain science, and so appeal to wetware. Bostrom then needs the wetware equals dryware assumption to pass over to his abstract, numerical estimations.) Then, to simulate human history, we would need to simulate 100 billion human brains times an average of

50 years per human life times 30 million seconds per year; the "scenery" surrounding these humans would contribute, according to Bostrom, vastly less to the overall number of operations and so can be neglected in our estimate. This gives a rough estimate of somewhere between 10^{33} and 10^{36} operations, and Bostrom asserts that even using current nanotechnology, a computer of planetary mass (of course, we don't yet have one of those) would be capable of 10^{42} operations, which well exceeds the needs for such a simulation (Bostrom 2003, 6). This prepares the needed link between the two senses of "good" simulation: the resources needed to make a simulation "pass" are roughly *just* those required to run ancestor-simulations. We see this from the only position that we can see it from, namely our own position. We estimate what resources are required to run an ancestor-simulation, and then we compare that to what resources we currently have, and then we extrapolate to what our resources might be like in the future. In partcular, Bostrom argues that both capacities—to make a simulation that passes, and to run ancestor-simulations—are reasonably within our future power.

I will return to the comparison between Smolin's and Bostrom's arguments in the next chapter, when we have the resources of Hilary Putnam's argument about brains-in-vats under our belt; but for now, that completes my discussion of the parallels. Now, I want to turn to a drastically different orientation of the question: what is required to make a simulation pass? I need to address the strategy of Bostrom's argument from a different angle to show what I find is drastically wrongheaded in this part of his project.

On pain of begging the question, we cannot *assume* that we are living in a simulation; even at the end of Bostrom's paper, we don't get this much as a conclusion. So, the obvious fact to recognize is that *we have absolutely no idea what is required to make a simulation pass*, *since we don't necessarily have any experience of that sort of simulation.* This seems obvious in a way, but the implications of taking it seriously are not obvious at all. They point us frontally to the question: How on earth are we to evaluate what would make a simulation pass?

It seems the only way to evaluate it is to proceed by analogy; this is what Bostrom does, in any case. So, we know what it's like to play a video game, and we may know what it's like to strap on video headgear for a "surround"

experience. Both of these experiences, of course, are completely different from living in a simulation because in either scenario we are able to tell that the former is just a simulation. Indeed, playing a video game is not consciousness *in* a simulation, even in a rudimentary sense, but consciousness *of* a simulation. We might "forget" for a moment that it's a simulation, but it's not even clear whether the term "forget" is being used metaphorically. At minimum, until we have the experience of people living 24/7 inside a simulation for considerable periods of time, eating in them, sleeping in them, and doing all sorts of other things, we don't have anything like a *phenomenological* account of what it would be like for people like us to live in a simulation (unless, of course, we are already living in one, which we are not allowed to assume).

To be clear, I'm not saying that when people are eating, they would have to be eating-in-the-simulation, and when they are sleeping, they would have to be sleeping-in-the-simulation. What I'm saying is that they would have to be in the simulation even while they're eating or sleeping (or whatever it is they're really doing all the time). And what it means to say that they are "in the simulation *while* they are eating" is not all that clear, at least to me. One imagines some scenario (we are heading toward the movies) in which feeding is intravenous. But now we are imposing conditions on the simulation, although putting demands on the coordination of the simulation has nothing to do with the general fineness of the grain. Effectively, the body in a simulation has been turned into part of the *wetware* of the simulation, but in any case, it seems clear that the simulation can in no way be independent of this substrate.

But this is not the sort of simulation that Bostrom is talking about anyway. In what Bostrom means as a simulation, there is only some generic substrate that is experiencing the simulation of which *we know not what*. That is, the simulation *just is* the running of a computer program of sufficient grain.[14] But this brings us back once again to the question: How would we know how to assess what fineness of grain would be required for a simulation in *that* sense? One begins to feel just how powerful, and potentially misleading, the wetware equals dryware assumption is, and specifically the sort of work it's being expected to do in the context of Bostrom's paper.

Returning to the issue of assessing our own experience as simulation, we may ask: Do we have analogous experiences of confusion about whether we are experiencing reality or not? Sure we do. The most famous one is,

no doubt, the experience of dreaming. While I believe this precedent is strongly behind the sort of speculation in which Bostrom and company engage, it is not ever mentioned directly in his paper, nor is it clear how it could be relevant given the structure of his argument. Indeed, we still have little neuroscientific idea of how dreaming works or what it's for. Unlike, say, our neuropsychological treatment of vision, the treatment of dreaming remains quite rudimentary, relatively speaking. This, I suggest, is because it is not (yet) even clear how it should be investigated from a neurocognitive perspective. So, the experience of dreaming is not only irrelevant from the perspective of the conceptual status of the argument, it is also effectively methodologically divorced from the underlying orientation of Bostrom's argument. This *is* relevant for the purposes of this book, because I will make the controversial claim that the dreaming/waking distinction is philosophically more fundamental than the appearance/reality distinction. If my claim is right, Bostrom's argument cuts the philosophical issues off at the knees.

In part because it prepares us for some later issues, I take another example, which is at once more amenable to discussion and less considered in the philosophical tradition—namely, art forgery and verification. Here is an example, at least, where we are abundantly clear what the issue is and we also have a variety of techniques for addressing it. I will leave aside all the physical forms of verification in favor of considering the perceptual verification of authenticity, since the analogous question in the simulation case is, first, whether we could perceptually (because consciously) identify a simulation as a simulation. (Note that scientific authentication of the simulation will be much easier to simulate, since the perceptual experiences involved in such scientific verification are typically much simpler. A large part of scientific testing involves work to bring investigation down to very simple perceptual measurement events: the number on a meter or the setting of a knob, and so on. This should, comparatively speaking, be easy to code up in a simulation, since it involves correlating the perception of the right outcomes with the perception of the application of the right probes. If there is a problem here, it would be with the speed of computer processing, but this is a consideration that extends beyond simple estimates of magnitude of operations.)

The practice of art verification was developed into a practice of "connoisseurship" in the nineteenth century, and a beautiful account of it is given in Edgar Wind's series of lectures *Art and Anarchy* (Wind

1963, 32–51). Turning to Wind's lectures also allows us to connect the consideration of Bostrom's argument above to the domain of the arts, and so establish some tentative links between two major sectors of our intellectual culture. For paintings, the verification of authenticity has proceeded in terms of identifying certain salient patterns of local detail that are unique to an individual artist and not readily forged. Although it is more complicated, the general principle is similar to verifying signatures. The process is not that complicated, and one can imagine computers being trained to execute this verification with extreme accuracy. I would hazard that by far the most difficult aspect of doing this would be to get the computers to isolate those identifying marks, by using "learning" algorithms (neural nets, genetic algorithms, etc.). I have little doubt that this could also be accomplished, and for all I know, this process may already have been implemented.

My point in citing this example, however, is that under the description I have given so far, the connoisseur goes into a situation in which he or she already knows that the painting either is an important original or is a forgery of this important original—otherwise, the process of verification is practically of no value. But everything the computer has been programmed to accomplish, or that the connoisseur has been trained to accomplish in the specific process of verification, has no bearing on the *importance* of the original painting and the process of its aesthetic canonization. Now, my point is not to ask whether this anterior process of canonization could be simulated by computer. Rather, the point is that I have no idea what would be involved in such a process of canonization simulation.

If I really study and read, I might be able to engage in the scholarly debate about whether the passage in *Sir Thomas More* is in Shakespeare's hand or someone else's, and if it is in Shakespeare's hand, whether it's really Shakespeare's writing or just Shakespeare's taking transcription. With study, I might also be able to engage in the debate about whether Hölderlin or Schelling or Hegel was responsible for the composition of the "Oldest Systematic Program of German Idealism." But the canonization of Shakespeare or Hegel is something that comes down to me. That means it needs to be part of the simulation. And the point is not that this can't be done, but that I simply have no idea how to estimate the

computational complexity of the processes that would be involved. Nor, I think, do Bostrom and his friends. The assumption is that such *intersubjective* concerns are reducible to the resources of the individual component (human brain) parts.[15] But this is question-begging, because it assumes that this sort of intersubjective capacity has already been accounted for in the estimation of the individual component's complexity.

Why is this important? At least one reason is that there is an important value-judgment inherent in ignoring these sorts of questions as Bostrom does—and as we find in the promotion generally of pseudo-science in the philosophical community. The value-judgment—if it can even be called that, because it's hardly ever explicitly formulated—is that such issues are negligible, either because they are simple or because they are unimportant. But to think that they are obviously (i.e., without consideration) either simple or unimportant is to support an exercise in cleansing the philosophical community of the chief traditional concerns of the humanities. And, in fact, this is the value program that is tacitly promoted by contemporary philosophical tradition, so there are no surprises here.[16]

That, however, is only the beginning of the matter. To see why, we need to head down the rabbit hole. But before we do that, I'll take a short excursion into art and anarchy, which may be skipped by the reader impatient to take the red pill.

2 Excursus: On Art and Anarchy

Edgar Wind's series of lectures, which I've mentioned earlier, supplies an alternative perspective that may be profitably juxtaposed with the orientation Bostrom tacitly establishes. At first, it may seem that what Wind promotes in his lectures in fact combats Bostrom's project and the orientation it represents, but I suggest that matters are less simple than that. Wind's project is not directed toward the problem of simulation per se, and even in his discussion of art forgery and authentication, Wind has other, larger targets in mind. Specifically, Wind is concerned with the changing attitude toward imagination in recent developments of the arts and how this changing attitude reflects a flip side of trends associated with mechanization of the human condition. A comparison of

these points can provide another cultural take on the preoccupation with the "confrontation" between human and machine—one that suggests a broader cultural context for the preoccupation with simulation in science, philosophy, and art.

The main contributor to the modern approach to art connoisseurship was a nineteenth-century Italian named Giorgio Morelli, who wrote a work (in German) under a (Russian) anagrammatized pseudonym. Beneath this playful self-retreating, Morelli presented a practice of connoisseurship that was threatening in both its simple practicality and its implications. Essentially, Morelli reasoned as follows: if we are to detect the presence of an original work, then we must look for details that are so trivial, yet personal to the painter that someone forging the work would not think to copy them. Tautologically, these details will not explain what makes the work of a great artist canonical; nevertheless, they reveal the presence of a great artist through the aesthetic analog of a personal tic. Once again, we see the practical context for the application of this project *presupposes* the canonization of the artwork—to be sure, I could perform this on some amateur artwork, but why would I?

Morelli's insight may profitably be compared to results I have informally heard reported about identical twins separated at birth. The twins may pursue radically different life paths, but often with strange "coincidences" in small, stylistic personality details. They will, for example, both give their cats the same name or both smoke the same brand of cigarettes. (This may turn out to be social folklore, but it's the nature of the suggestion, and not its truth, that ultimately interests me.) In the case of fine art, some art historians have even attempted to invest these small-scale features (the sort of cross-hatching employed, the shape of an ear or hand, as Morelli catalogues in his work) with psychoanalytic significance (Ehrenzweig 1967).

Whether we attach such psychonanalytic significance to art or not is secondary both to my argument and to Wind's, but what is primary for Wind is that, in isolating these details, we cognitively "dissociate" features that are constitutive of the very personality of the practice involved. That is, Wind sees this view as part and parcel of the way in which mechanization, both in the thought process and in the world at large, has contributed to a marginalization of the art experience.[17] In the age of mechanical

reproduction, not only does fine art become identified with its means of reproduction but also the book illustration displaces or even replaces the experience of the canvas. Even more radically, the reproduction process dictates the nature of the artwork involved—the recorded musical performance that is designed for repeated listening influences the tradition of live performances, making them more conservative, more attentive to avoiding mistakes, more risk averse.

It is this dissociation that is Wind's primary target in his series of lectures, and it is that which Wind takes to be the primary engine behind the marginalization of the art experience. This marginalization, paradoxically, proceeds by way of a glut in our exposure to art, which numbs us to an appreciation of the art as a perceptually and cognitively privileged act. With this experience of aesthetic saturation comes the need for an increasingly radical intensification of the imaginative experience. On the one hand, this contributes further to the overall aesthetic saturation, while on the other hand, this intensified appeal to the imagination cannot combat the exponential proliferation of the aesthetic. The outcome Wind identifies is a progressive detachment of the imaginative from rational constraint—hence, what he calls anarchy. As he points out, until roughly two hundred years ago, one found no expression of concern that rational constraint would hinder imaginative production; however, in a contemporary context, one encounters (as a limiting case) anxiety that any rational constraint could be lethal to the imaginative experience. Certainly, this development has only become more extreme since 1960, when Wind gave his lectures.

What is perhaps most disturbing in the art history that Wind presents is his insistence that the progressive mechanization and dissociation of experience, on the one hand, and the progressive radicalization of the appeal to the imagination, on the other, are two sides of the same coin. Yet, if we think about the *Matrix* movies, we see that these two trends fit there as well: the experience of the movie is a blend of overwhelming technical effect concomitant with the promotion of a mythology of (if you will forgive the pun) the imaginative capacities of neological vision. We are not living *in* the *Matrix* movies, but somehow these movies are representative of our experience of the massive dissociation induced by the mechanization of our perceptual and cognitive experiences. It seems

that Wind's brief history of the aesthetic experience is a much better explanation of our interest in the *Matrix* movies than the bare possibility that we might be living in a computer simulation.

I think this is manifestly true, but there is danger that can come along with getting caught up in Wind's style of explanation. Although I can't recall his ever saying so expressly, the entirety of Wind's presentation tends toward the suggestion that we need to do something to recover the special value of the aesthetic experience. There may or may not be a hint of this in Wind's work, but often this sort of work carries a suggestion that we need to go back, that eventually our situation will become so perceptually and cognitively distressing that we will have no choice but to retreat. Everything that has happened since 1960 suggests that this process has only greatly intensified.

The famous Austrian philosopher Ludwig Wittgenstein kept notebooks in which, from time to time, he would make entries in a special "cipher-script." These entries would often have to do with matters that were intimate for him, sometimes of a personal nature and other times more philosophical. One of these entries reads:

> I believe mathematics went through a period of loss of instinct during the last century from which it will continue to suffer for a long time. I believe this loss of instinct is connected to the decline of the arts, it arises from the same root. (Wittgenstein 1994–96, I 157)

Wittgenstein seemed to feel that his philosophical work could perhaps eventually help in treating the cancerous mathematical growth of the recent past and restoring health through a rehabilitation of "instinct." His prediction that this loss of instinct would last "for a long time" has certainly been vindicated so far, but there is no reason to believe that the future of mathematics will regain it in anything resembling the sense Wittgenstein hoped. Wind's lament for the loss of the special privilege of the aesthetic has a similar feel, at least to me.

While this casts doubt on the normative conclusions we might draw from the works of Wind and Wittgenstein, it does little, however, to invalidate their descriptive presentations. There is much to recommend the idea that Wind, rather than Bostrom, provides us with a means for

mapping out our contemporary preoccupation with the Matrix, but we should not take the additional step of thinking, hoping, or aspiring to remove ourselves from its[18] vice-like grip. This, however, is the primary fantasy in which the *Matrix* movies indulge us, and a great deal of work will be required to set up a confrontation with this preoccupation as one of our primary fantasmal desires.

Notes

1. For report of a related lecture by Dyson, see Lapidus 2008, 301. Lapidus expresses doubt that the proposal by Dyson he discusses will lead to a proof of the Riemann Hypothesis.
2. Conway Morris 2003 surveys work on the problem of the origins of life, but there is no mention of Gold's conjecture.
3. As is the case with other articles where web versions are more readily available, I cite page numbers from the version available online.
4. The view is formulated as a trichotomy, but can be recast as a conditional; for details, see the discussion below.
5. For an interesting critique of the sort of speculation in which Dyson engages, see Hardin 1996, 120–28, where Hardin addresses the cost-benefit of the SST.
6. The conceptual analysis of the argument has been pursued, among others, by Anthony Brueckner and Brian Weatherson, with responses from Bostrom in both cases. See Weatherson 2003, Bostrom 2005, Brueckner 2008, and Bostrom 2009. More discussion and references are also available at www.simulation-argument.com. I discuss this debate briefly below and in Chap. 3.
7. For a discussion of the wetware conditions for life on this planet, see Conway Morris 2003, 23–27; for the wetware conditions for DNA, see 27–31. The overall argument of Conway Morris's book supports the view that these wetware conditions are very specific; of course, the matter is a subject of intense debate. For an account of consciousness that takes wetware conditions to be critical, see Llinás 2001; for opposing methodologies in the Artificial Life project, see Dennett 1995, and Harnad 1995.
8. By not making the assumption, I mean remaining agnostic—that is, swearing off. Making the assumption that consciousness does depend on wetware in some more specific sense would also be speculative, and I'm not trying to assess this alternative here.

9. Smolin attributes the idea of the production of cosmoi with variations in basic physical constants to John Archibald Wheeler; see Smolin 1997, 330; the reference is to Wheeler 1974.
10. The glitch in the argument arises "from the possibility that the average number of people living in the pre-posthuman phase might be different in civilizations that produce ancestor simulations than in civilizations that do not" (Bostrom and Kulczycki 2011, 5).
11. In his response to Brueckner, in particular, Bostrom refers to "the substrate-independence assumption, which underpins the entire simulation argument and which Brueckner does not challenge." So far as I can see, this assumption is challenged nowhere in Bostrom's exchanges with either Brueckner or Weatherson, although the consideration of the proper construal and implications of this assumption are crucial in the debate with Brueckner. I will return to this debate in the next chapter, insofar as it bears on Putnam's brains-in-vats argument. Weatherson's article is focally concerned with the evaluation of the indifference principle Bostrom appeals to in the larger argument involving the formulation of his probabilistic estimates.
12. In his response to Anthony Brueckner, Bostrom insists, rightly enough, that he is not making an argument (directly) about whether we are living in a simulation, and he states in fact that he (personally) does not believe we are living in a simulation. But certainly his argument is intended to have bearing on the question, "Are you living in a simulation?," which he takes as the title of his paper. See Brueckner 2008 and Bostrom 2009.
13. This claim depends on the relative version of substrate-independence I've described above. If the simulation is entirely substrate neutral, then why should we expect the demands on the simulation to be comparable to what we find in another substrate? In the strictest sense, it really seems we have no idea what these constraints should be on the basis of brain science insofar as it involves the wetware construction of brains. This is one example among others of the way principles of design are being used to make the estimates relevant to Bostrom's argument. In at least one case, Bostrom refers to these design principles as such (Bostrom 2003, 4).
14. Given Bostrom's commitment to there being some substrate, we must presumably understand this to be the "concrete" rather than merely "abstract" running of the program.
15. Perhaps, more generously, we could say that the arguments are simply "coarse-grained" and so do not take such "fine-tuning" into account. But this still constitutes a de facto reduction of these concerns.

16. A couple of decades ago I might have felt the need to say "the contemporary Anglo-American mainstream," but this tendency is now in the process of becoming pervasive in philosophical culture, at least in the First and Second Worlds, and I am inclined to think globally.
17. Taking a different tack, Theodor Adorno suggests that it (also?) leads to a massive reconfiguration of artistic creation in an attempt to dissociate itself from the dominance of mechanical reproduction: "Modern painting's aversion to figurative representation, which in art marks the same breach as does atonality in music, was an act of defense against mechanized art merchandise, primarily photography" (Adorno 2006, 9).
18. I leave the referent deliberately ambiguous.

3

Are You a Brain-In-a-Vat? (Level of Logic: Level 1)

1 Brains-In-a-Vat

I begin my presentation of Hilary Putnam's paper "Brains in a Vat" by sketching Putnam's strategy in large, cartoonlike strokes. Like Bostrom's paper, Putnam's engages in a form of radical speculation, albeit in a different and more sophisticated way. The best way to bring out what is radical in Putnam's program is to caricature it in a way that puts its radicalism into relief. Fundamentally, the radicalism of Putnam's program is associated with his claim that philosophy can do a sort of work of that physics, or more generally any natural science, is incapable of doing. To insist on this point, Putnam has to uphold a distinction between logical and practical possibility—or something like that. I begin by emphasizing the way he does this.

Putnam begins his paper by imagining that an ant traces a path in the sand that ends up looking like "a recognizable caricature of Winston Churchill" (Putnam 1981, 1). But, says Putnam, we wouldn't say that the ant has *depicted* Winston Churchill unless it first saw Winston Churchill, had the intelligence and skill to draw him, and then intended to do so (Putnam 1981, 2). This seems more than a little strong, since plenty

O.B. Bassler, *Diagnosing Contemporary Philosophy with the Matrix Movies*, DOI 10.1057/978-1-137-57889-1_3

of people could draw Winston Churchill, and have drawn Winston Churchill, without ever having seen him, only other representations of him.[1] It would be fussy to say that they are depicting representations of Winston Churchill. But never mind that.

Recently, I've taken to painting. I'm still a novice, and I'm focusing in two areas: abstract paintings and portraits. When I brought in one portrait to show a group of students recently, one of them said that it looked like Winston Churchill. Others have said the portrait looked like Harold Bloom. In fact, the portrait was made from a photograph of Hannah Arendt. The natural conclusion is that I have not drawn a very good portrait. But, although it is no excuse for my deficiencies as a painter, a portrait is not exactly a *likeness*, although of course they're not unconnected, either. Currently my favorite portrait—the one I think about the most—is Pablo Picasso's *Portrait of Nusch Eluard,* from 1941 (Rubin 1996, 85). When I look at this portrait, I don't first think, "Oh, that's obviously Nusch Eluard." I look at the title and then I think back to the photographs I've seen of Nusch Eluard (three are available in the same volume I've cited, at pages 80, 82, and 83) and also at other portraits, and I think, "Oh, that's interesting." You can simply write off my portrait as a bad likeness (and you probably should), but Picasso, no, that you cannot do. Once again, we see the canonization procedure stepping in to mark the options.

But Putnam speaks of a *caricature* (just as I'm caricaturing Putnam's argument), and perhaps that's significant. Churchill's features are often characterized in terms of their strongly neotenous flavor, their "baby-like" nature. Although I don't plan to do a Derridean number on Putnam's paper, it is interesting that he chooses Churchill here, since stretching the point just a bit one might argue that it is possible to see Churchill in just about any newborn. But once again, never mind; we aren't concerned with these sorts of questions here, are we?

Let's imagine another context. We look up at the clouds and we "see" a wolf, or an elephant, or maybe you see a wolf and I see an elephant. We wouldn't generally say there's somebody up there depicting. But is it *improbable* that I see a wolf or you see an elephant? No, it happens all the time (according, at least, to a certain way of speaking about things). Maybe we might even see Winston Churchill. But here's what we won't see up

in the clouds or anywhere else: we won't see a bunch of monkeys typing on typewriters and producing Shakespeare's *Hamlet* (Putnam 1981, 5). Indeed, I've never even seen a bunch of monkeys typing on typewriters, but we're not interested in that either, are we? We're not interested in the *wetware* problem of monkeys typing out *Hamlet*; we're just interested in the *dryware* problem. This is a dramatic way of saying that the probability of producing *Hamlet* by selecting letters, spaces, and punctuation marks at random—or anywhere near at random—is just about zero, or practically impossible. But, as Putnam insists, it's not *logically* impossible; indeed, according to Putnam, it's not even *physically* impossible (Putnam 1981, 5). Putnam uses this distinction to do a lot of philosophical work, so if this distinction falls out, *Putnam's whole point is sunk*.

How committed are we to the idea that these things that Putnam cooks up—about ants drawing Churchill, or (radically) monkeys typing out *Hamlet*, or (more radically) an evil scientist tending a world full of brains bathed in vats, or (even more radically) a world of brains bathed in vats *not* tended by an evil scientist (which even Putnam says is "absurd," as if the others are not—or not yet quite?)—are logically possible even if they are not practically possible? In particular, are we more committed to this idea than to the conclusions Putnam draws from it—that the brains-in-vats scenario can be ruled out on *philosophical grounds alone*? Or are we concerned that "*meanings just aren't in the head*" (Putnam 1981, 19), a crucial premise of the argument, or that there is no necessary connection between the representations and their referents, another crucial premise of the argument?

More pressing even is whether we should agree with Putnam when he tells us that he isn't claiming all the wild scenarios he details are "very likely"; he simply means "that there is nothing at all unimaginable" about them, or at least some of them (Putnam 1981, 17–18). But how do we know whether we can imagine even the one to which Putnam is referring specifically when he makes this claim—a scenario in which someone under hypnosis is thinking certain words and having the feeling of understanding them without having any understanding of them? I am inclined to think that most people, if presented such a scenario, would simply think that Putnam has been spending too much time around philosophers, the latter who develop a peculiar capacity to think they can

imagine all sorts of things—not that I want to appeal to the proverbial ordinary (wo)man in the street as a referee, either.

The main point is this: we have to be more secure in our beliefs about these premises than we are in the conclusion of the argument, so that the argument gives us valuable information. Or, at least, that's how arguments are usually (which is to say, outside of philosophical contexts) conducted. But the conclusion of Putnam's argument is that we can't be brains-in-vats. Well, I knew that already, so what the hell is going on here? This concludes my caricature of the strategy of Putnam's paper.

Now, leaning heavily on this caricature, we might try to reformulate our sense of Putnam's aim. This reformulation goes as follows: Of course, Putnam isn't doing *that*, even if it may at first seem to be the case. What he's doing is drawing some argumentative support for his main view, which is that meanings aren't in the head. This view has traditionally met with a lot of resistance, but with it in hand, not only can we rule out some desperate science-fiction scenarios that preoccupy a sizable community of recidivist metaphysicians but we can also rule out these scenarios on *strictly philosophical grounds*. That is, we can do this without any appeal to an evaluation of the reasonability of (pseudo-)scientific estimates of the physical limitations on our capacity for computation, or any question about whether wetware equals dryware, without the metaphysical debates that infect the community of cognitive-science-influenced philosophers when they engage in futurology; that is, we can wipe these issues off the table.

This general construal of Putnam's project has several virtues. First, it makes Putnam's target clear. He is not concerned about convincing the layperson that he or she is not a brain-in-a-vat; he's concerned with arguing a community of philosophers out of their collective wrongheadedness. Contextually, this is a much more reasonable way to construe Putnam's paper. As he says, the paper was itself an outgrowth of thoughts on more general and more esoteric issues about the foundations of mathematics that we'll come to later, an explanation that provides some genetic support for this construal. That is, Putnam's fundamental target is not at all brains-in-vats claims.

The other virtue of this general construal is that it exposes what is fundamentally problematic in Putnam's paper, even if we shape it as

generously as possible. The problem this construal identifies is rooted in Putnam's promotion of the philosophical claim that meanings just aren't in the head. The problem is not with the claim itself but, rather, with what Putnam goes on to say positively about where those meanings are. In his words, the understanding of the expression of a thought "is not an *occurrence* but an *ability*" (Putnam 1981, 20). Effectively, although Putnam does not use quite this language, he's saying that meanings are not in the head; they are in our actions.

The problem is that this position continues to construe the question "Where are meanings?" and does so in the *wrong way*. Even if Putnam were to say that my gloss is inaccurate—that in saying that understanding is an ability is not glossing the meaning by locating meanings in our actions—he nonetheless implies that understanding lies in our abilities in a manifestly literal way. And, contrary to what Putnam seems to think, this desire to identify the locus of our understanding in such a literal way is just the *wrong sense* of locus. Or so, at least, I will argue.

2 Excursus: On Understanding Meaning

Last night, after writing the sentences above, I happened upon this passage from the preface to David Jones's long poem "The Anathemata":

> Or, to leave analogy and to speak plain: I believe that there is, in the principle that informs the poetic art, a something which cannot be disengaged from the mythus, deposits, *matière*, ethos, whole *res* of which the poet is himself a product.
>
> My guess is that we cannot answer the question "What is poetry" (meaning, What is the nature of poetry?) without some involvement in this mythus, deposit, etc. (Jones 1965, 20)

Although this is most probably not plain speaking by Putnam's standards, it is precisely those standards that are at issue. What Jones is saying about poetry, that the "is" question about it cannot be answered in isolation from the "mythus, deposits, *matière*, ethos, whole *res* of which the poet is himself a product," I am claiming also holds for meaning and the meaner.

I am also claiming that if it holds for meaning and the meaner, it likewise holds for the understander. But the understander stands to meaning in a condition of potential dissipation, which only makes the issues to be addressed all the more challenging. That Putnam should focus on meaning and the understander is a function of his functional conception of meaning, which is analogous to the relation in which the connoisseur stands to the painting, dissociating its "identifying" features, as described in Chap. 2. Putnam's reasons for focusing on meaning and the understander rather than meaning and the meaner are largely if not entirely rhetorical: he is committed to the negative condition that a creature (like an ant) lacking the right sort of intentions for language production cannot mean, but he is also committed to rejecting any account of meaning that would rely on a phenomenologically introspective appeal to intentions. Why is this? The answer depends on "the mythus, deposits, *matière*, ethos, whole *res*" of which Putnam is himself a product. What holds for the poet holds for the philosopher as well.

This is no doubt troubling, even painful, for the community of aspirant philosophers to acknowledge. The failure to acknowledge the problem does not make it go away, and there is lingering malaise associated with repression of this condition by way of an appeal to canons of scientific rigor and other standards more appropriate to the scientific enterprise. One imagines similar malaise among medieval philosophers in their appeal to the canons of theological rigor. This form of philosophical repression is not entirely unproductive under the right conditions and in the right hands, but in the wrong conditions and in the wrong hands it is a disaster. By now, historically speaking, we are powerfully in those wrong conditions.

Why doesn't what I'm asserting simply amount to another lament for the lost world of philosophical "humanism"? For one thing, we have not, in our accelerated condition, stopped being human through a process of technological advancement. Attempts to reject our humanity are not analogous, in the relevant respect, to attempts to reject our technology. The anxiety that, in becoming technologically advanced, we will cease to be human is generally misplaced, although this anxiety is powerful fuel for *Matrix* Mythology.[2] As mentioned in Chap. 2, Edgar Wind, as a Renaissance art historian, lamented the lost paradise of a time when the

imagination remained creative under its counter-anarchic subjection to rational constraint; and Wittgenstein lamented the paradise of an earlier mathematics still firmly rooted in healthy "instinct." (Now what does that sound like? Wittgenstein was, after all, a reader of Otto Weininger in his formative years. Remember, too, Wittgenstein's rejoinder to Hilbert that Cantor's "paradise" was no paradise at all [I'm just paraphrasing].) These laments are, to be sure, reactionary, and it is precisely here that I part company with these two. In fact, Putnam does also.[3]

The analogy I'm drawing between what Jones (another reactionary) says about poetry and what I'm saying about the philosophical problem of meaning is bound to raise questions about the relation between poetry and philosophy. Because of current intellectual conditions, poetry and philosophy have largely fallen out of mutual conversation. For all his faults, Heidegger's recognition of the ongoing need for this conversation is greatly to his merit, and what he says about its conditions is not entirely without merit. But I think it, too, is sadly misdirected, fueled by an inadequate residual romanticism that Hans Blumenberg has thoroughly unmasked. Blumenberg, on the other hand, largely avoids the confrontation between poetry and philosophy and heads directly for the confrontation between philosophy and myth. But we need both. So, I'm not offering an answer to the question about what distinguishes the poetic and the philosophical conditions here. Erich Auerbach has noted that at the apogee of scholastic theology, the proximity of poetry and philosophy was powerful and productive (Auerbach 1961, 72–73). Dante's proximity to, and arguably even his philosophical outstripping of, Thomas's scholasticism remains a crucial example to confront.

3 Diagnosing Putnam

If we are not interested in Putnam's answer to the question "What is meaning?," what remains of interest in Putnam's paper? That is to say, Putnam's question is a more decent philosophical one than Bostrom's "Are you living in a computer simulation?," but if we don't buy Putnam's answer—if, as I've suggested, his answer is fundamentally misguided—then what value is there in continuing to think about Putnam's paper?[4]

The answer is that we are now at the same point with Putnam's paper as we stood with respect to Bostrom's from the beginning: Putnam's paper ceases to be of interest (for current purposes) in a direct philosophical way—as the answer to a question—and becomes instead of interest *diagnostically*. As in the Chap. 2 consideration of Bostrom's paper, the best way to get at this diagnostic significance is to look at the general structure of Putnam's argument.

Putnam's paper has the structure of what has come to be referred to in the philosophical community as a *transcendental argument* (Baird 2003). I suggest that this transcendental argument replaces—or better even, *dislocates*—the evolutionary argument that served as the centerpiece of Bostrom's paper. This is the strength of Putnam's argument; its weakness is that it is still insufficiently internal as an argument. That is, all arguments rely on premises (contra some Hegelians, nothing is got for or from nothing), but Putnam's argument relies on premises that are insufficiently *integrated*. And furthermore, the near-obsessive focus on arguments in papers like Bostrom's and Putnam's is also an indication that something is awry, stressed out in the contemporary philosophical mind. Such obsession is a mark of scholasticism, and we find it in the period of high theological scholasticism, just as we do here in the context of its scientific counterpart. Recognizing this condition is fundamental to the diagnosis I am conducting, but having recognized it, I now proceed with a diagnosis of the arguments involved. (As a generality, proof theory turns out to be the most intractable part of metamathematics, and its difficulty has historically been a large part of the motivation for the turn to just that model theory that Putnam tries to show us has caused so many philosophical problems. That is to say, the particular sort of scholasticism Putnam engages in has its roots in a very real problem with the structure of proof. This problem will return with a vengeance, and I face it head-on in the Appendix.)

The particular form of transcendental argument that Putnam pursues we might call a de(con)structive transcendental argument. Putnam says that in arguing for the possibility of the brains-in-vats scenario, the arguer presupposes (and must presuppose) something that implies the impossibility of the scenario. This form of transcendental argument is related to the accusation of begging the question. In this case, however, it is not that

the proponents of brains-in-vats assume what they attempt to prove but, rather, what they assume (that we are brains-in-vats) implies the *incoherence* of what they attempt to assume. When suitably inspected, the argument even for the *possibility* of brains-in-vats is, therefore, self-refuting.

Like mine, Putnam's point is diagnostic: what he really wants to show is how certain assumptions lead people ("philosophers") to endorse scenarios that are incoherent—that is, that are internally inconsistent. Once they recognize this inconsistency, the strategy goes, they will be forced to evaluate their perhaps tacit assumptions and give them up in favor of the program Putnam is advocating—namely, that meanings just aren't in the head, and that understanding is an ability, not a mental state we can introspect. Once we get our "position" right, the transcendental argument has done its work and can be left behind. The transcendental argument has served its purpose of helping us to "get our house in order." Philosophers should clean up after themselves. But is this a cleanup job that can be done once and for all, *semel in vita,* as Descartes said of his enterprise in the *Meditations*? Kant asserted that, on the contrary, reason's teleological drive will always tempt us in the direction of such metaphysical viciousness; and in a moment of disarming candor, Wittgenstein made a related admission to one of his students. But if this is true, then how can the transition from recognizing understanding as a mental state to recognizing it as an ability fundamentally help? I don't think it can—at least, not at this level.

In this area, Putnam's philosophical commitments stem from his reading of Wittgenstein. In fact, he says that his justification for the claim about the nature of understanding just is "a very abbreviated version of Wittgenstein's argument in *Philosophical Investigations*" (Putnam 1981, 20). For the record, I think Putnam misses Wittgenstein's fundamental concern in the portions of the *Investigations* at issue, but since I have my own problems with Wittgenstein's fundamental concern, I won't get into this supremely difficult matter now. In any case, Putnam tells us that the motivation for his argument in the paper comes from thoughts originating in the context of his consideration of the Löwenheim-Skolem (L-S) Theorem, and I think not only Putnam's commitments but also the fundamental problems with his attachment to these commitments become clearer in that context. To be sure, the issues there are in some

sense more arcane, but they are by virtue of that fact also purified, in a way, of extraneous matters. We follow this sort of purification all the way down the rabbit hole.

But before we go there, let's look at the specific form the accusation of self-inconsistency takes here. The claim at issue is simply that "we are brains-in-vats"; we are considering whether this claim expresses a *possibility*. Putnam claims it does not, because the assertion of its truth implies that it is false. But why?

First, we must get at least a bit clearer on what Putnam takes "brain-in-a-vat" to mean. Such a "thing," according to Putnam's construal, does not have sense organs, but has "*provision* for sense organs; that is, there are afferent nerve endings, there are inputs from these afferent nerve endings, and these inputs figure in the 'program' of the brains in the vat just as they do in the program of our brains" (Putnam 1981, 12). Putnam goes on to say that the brains-in-vats are brains, and moreover that they are *functioning* brains; this is important for the progress of his argument. First, this state of affairs, rather than any commitment to the dryware equals wetware hypothesis, is what allows us to speak of consciousness here, and so Putnam avoids the premise upon which Bostrom's entire approach is founded. We are manifestly dealing with fully embodied wetware, and then the argument gets going. Despite the fact that "they" are conscious, according to Putnam this doesn't mean that "their" words correspond to the same thing as "our" words; note that this claim presupposes a distinction between "us" and "them." We are considering whether it is possible that there could be brains-in-vats, not whether we are them. For Putnam, the key question is: "Can they refer to *external* objects at all? (As opposed to, for example, objects in the image produced by the automatic machinery)" (Putnam 1981, 12).

No, Putnam says, they cannot. This is because the fact that the program connects language about trees to sensory inputs (which produce what we are informally, and ultimately misleadingly, calling "images of trees"), which do not represent trees or indeed anything external, cannot in any way make it true that I am perceiving trees or indeed anything external. In particular, saying "I am a brain in a vat" cannot correspond to anything external either, but that is just what we are assuming a brain-in-a-vat is. So, we see that what we are referring to is not something that the

brain-in-a-vat *could* successfully refer to. But, one might say, this doesn't seem to mean that there couldn't be a brain-in-a-vat—only that such a thing could never successfully refer to itself as such.

In some parts of his paper, Putnam only seems to be claiming that such a thing could never successfully refer to itself as such (or in any other way, such as "we are brains in vats in the image," which is equally self-refuting). But at least once in his paper, Putnam seems to be making the stronger claims that brains-in-vats are impossible:

> I want now to ask a question which will seem very silly and obvious (at least to some people, including some very sophisticated philosophers), but which will take us to real philosophical depths rather quickly. Suppose this whole story were actually true. Could we, if we were brains in a vat in this way, *say* or *think* that we were?
>
> I am going to argue that the answer is "No, we couldn't." In fact, I am going to argue that the supposition that we are actually brains in a vat, although it violates no physical law, and is perfectly consistent with everything we have experienced, cannot possibly be true. *It cannot possibly be true*, because it is, in a certain way, self-refuting. (Putnam 1981, 7)

It is at just this point that Putnam goes on to mention that this idea first came to him in the context of thinking about the L-S Theorem.

It seems, at least, that the route to the stronger conclusion is evident and direct. A self-refuting proposition cannot be true; the proposition "I am a brain in a vat" is self-refuting; therefore, it cannot be true that I am (actually) a brain-in-a-vat. Indeed, I think that this is Putnam's argument, and that the reason he does not come back to it later is that it is so evident and direct a conclusion from what he goes on to argue for. But is it? No, it is not, because as we have seen, the entire argument is first set up by distinguishing between me and a brain-in-a-vat. The argument is predicated on a version of *semantic ascent*: I am at one level—call it the meta-level—and the brains-in-vats, which I am considering, are at another. But the "I" in the proposition "I am a brain in a vat" is at the brains-in-vats level, and "I" in "it cannot be true that I am (actually) a brain in a vat" is at the meta-level. There is no pre-established identity between these two "I"s, and if Putnam is right, there in fact cannot be

any such identity. So, the stronger conclusion does not follow, which takes a lot of the intuitive wind out of Putnam's sails. Now, Putnam is just arguing that brains-in-vats couldn't successfully refer to themselves. This is much closer to a more traditional skeptical dilemma about language. When Putnam makes his stronger claim, he presents himself as being able to do philosophical—indeed, I would say metaphysical—work, which rules out scenarios that cannot be ruled out physically or phenomenologically ("experientially"); this is really what would give his paper its "swoosh." Once this stronger claim is discounted, though, we are left with the unsurprising assertion that on Putnam's particular account of the referential function of language, there are certain apparently referential linguistic claims that in fact fail to refer. Big deal. No swoosh.

Putnam's "swoosh" points us in the direction of the most fundamental issue in his paper, and one on which he and Wittgenstein have no disagreement (either on Putnam's interpretation of Wittgenstein or on my own). This is the insistence on drawing a hard-and-fast distinction between physical possibility and experiential possibility, on the one hand, and logical possibility, on the other. This takes us back to the issues I raised at the beginning of this chapter, when I caricatured Putnam's paper. Now that we've looked in more detail at the structure of his arguments, it's time to revisit them. Putnam asks us to consider all sorts of scenarios that he admits aren't "likely" but are "perfectly imaginable." But how do we circumscribe what's imaginable and what isn't? This seems about as indefinite as Putnam's commitment to logic is definite. Furthermore, what business do we have considering all these wild possibilities in the service of some philosophical conclusion? That is, what sort of work are the wild examples doing?

It will be helpful, but not yet sufficient, to draw a couple of distinctions. First, we should pay attention to those cases when Putnam is endorsing these examples to argue for his own philosophical ends, versus those cases in which he is considering these examples to argue against someone else's philosophical ends. At the beginning of the paper, the example of the wild ant (not) drawing Churchill is one Putnam is using to motivate a position he endorses about depiction. When he considers the wild brains-in-vats scenario, he is arguing *against* the consistency of (at least some claims about) this scenario. So, it is manifest that Putnam

is not fundamentally averse to using wild examples to support his own position, but he thinks some wild scenarios are self-inconsistent. The problem, then, isn't with countenancing wild scenarios per se; indeed, even internal to the refutation of the brains-in-vats scenario he considers a wild Twin Earth scenario under which the referents of the terms "elm" and "beech" are switched. (Putnam "blush[es] to confess" (Putnam 1981, 18) that his concepts of elm and beech are identical; this seems to make, for Putnam, "all beeches are elms" an analytic, though false, proposition. I am unclear whether Putnam would endorse this implication.)

I do not pretend to know how to judge what is perfectly imaginable or not, but perhaps I am simply "challenged" in this regard. I can only say what I think is going on, and then I will say what function I think it is serving. The claim that something is imaginable implies, at minimum, that it is logically possible, but not necessarily that it is practically possible. The claim is clearly defeasible: we can seem to imagine something—the brains-in-vats scenario, say—that then turns out to be logically impossible. And this means we cannot really imagine it. Presumably, the other claims—about ants in the sand, about hypnotized "speakers" of Japanese, about Twin Earth—are defeasible as well *if* we are able to identify a logical inconsistency in their supposition. Our guide in imagining wild possibilities seems to be some unspecified appeal to what is permitted by physical laws and what we judge possible on the basis of our past experience, but this can only serve as a guide: the aim of Putnam's paper is precisely to point out an instance in which such guides fail us and philosophy must come to the rescue. (All of this is fitted to the stronger claim I've discussed above, which Putnam states on page 7 of his paper, and which would give the paper its "swoosh.")

But this is different from what I would claim the wild examples are doing, which is being a literal proxy for metaphorical language. Philosophers are the most fundamentalist constituents of contemporary intellectual culture; apparently they cannot recognize the meaning of a proposition unless it is absolutely and unequivocally literal. The first problem with the claim that meaning is in the head is not that it is false but that it is not *literally* true. But neither is the claim that understanding is an ability. An ability like what? Like riding a bicycle? Like, how? Are we presuming that all abilities can be brought under one umbrella? In fact, I

would claim, the literal construal of such propositions requires such univocity. If this is right, then Putnam has either miscontrued Wittgenstein, for whom terms like "ability" cover some area of "familial resemblance," or Wittgenstein is internally at odds with himself, or both—as I'm ultimately inclined to think. (We're all at odds with ourselves; we're human. But some of us are more interesting than others. Wittgenstein is one of the really interesting ones, and so I am quite likely mistaken about where he's at odds with himself.)

There are some radical proposals, coming especially out of probability theory but bolstered by developments in quantum physics, that would deny the absoluteness of the distinction between physical and logical possibility. In *Understanding Quantum Mechanics*, Roland Omnès enlists the support of the early twentieth-century probabilist Emile Borel:

> Emile Borel . . . was particularly interested in the meaning of very small probabilities and he proposed a unique axiom for interpretation: one should consider that an event with too small a probability will never occur. His arguments are detailed and careful, but the main one is that this kind of event cannot be reproducible and should be left out of science. Mathematical theory cannot avoid the consideration of these events, but what it says about them cannot be held reliable *nor even sensible*. (Omnès 1999, 84; my emphasis)[5]

Here, Omnès at least motivates the suggestion that if an event has too low a (theoretical) physical probability, it is not to be considered logically possible (i.e., "sensible"),[6] and this makes the distinction between physical and logical possibility an indefinite one (unless, of course, we can fix an absolutely precise finite, positive value for the probability below which events have probability zero, in which case at least on this particular count physical and logical possibility would coincide). Borel's proposal remains exotic, but it presages issues that will become more directly relevant as we descend.

I'd like to end this chapter with a brief discussion of some work that connects the arguments of Bostrom and Putnam. In particular, Anthony Brueckner has written on both Putnam's brains-in-vats argument and Bostrom's simulation argument, and he asserts (suitably qualified)

that Bostrom's simulation scenario is incoherent on grounds related to Putnam's argument about brains-in-vats. Following Brian Weatherson, Brueckner labels Bostrom's "non-human, computer-generated minds" as "Sims," and he makes the claim that "The Sims cannot *really* create other Sims. The idea that 'stacked virtual machines' can give rise to more and more conscious Sims seems to be just a confusion." It is in a footnote at just this point that Brueckner draws the analogy with Putnam's paper: "Just as a brain in a vat is incapable of *really* building another brain in a vat, a Sim is incapable of *really* building another computer which instantiates another humanlike conscious Sim mind" (Brueckner 2008, 224).

Bostrom's response to Brueckner's claim comes in several parts. To begin, he denies that the analogy between Sims and brains-in-vats on the particular point at issue. Although brains-in-vats do not (and cannot) create other brains-in-vats, Bostrom insists that computer programs can "create" other computer programs in the sense that executing a computer program can emulate the execution of another computer program. Since the sense of consciousness at issue in Bostrom's paper is such emulation/simulation, consciousness in Bostrom's sense can produce other consciousness.

I think Bostrom does have the better end of the argument in this case, but it only goes to show once again what powerful consequences can be drawn from the wetware equals dryware assumption, which as Bostrom explicitly recognizes, Brueckner nowhere challenges. On this basis, Bostrom goes on to insist that "stacked virtual machines, implementing suitable programs, could produce more conscious Sims does not seem to present any special difficulty" (Bostrom 2009, 3). But although Brueckner's objection is effectively blocked, the example Bostrom uses to illustrate his point leads to another set of questions about the coherence of Bostrom's position. Bostrom says: "It is perfectly possible (albeit somewhat pointless) to create a Java applet that implements a virtual machine that runs another Java applet" (Bostrom 2009, 3). But the sort of point Bostrom wishes to make presupposes that computer programs are sufficiently individuated that we can pick out such nested instances unequivocally. In particular, if the individuation of programs is either intrinsically underdetermined or requires independent specification in terms of wetware conditions, then Bostrom's argument would be compromised.

This leaves the question whether Bostrom's notion of simulation, while not analogous to brains-in-vats on the point Brueckner enlists, nonetheless falls prey to the same sort of argument Putnam makes against brains-in-vats in his paper. David Chalmers has argued that it need not in his paper, "The Matrix as Metaphysics." There, he asserts that the claim "I am in a matrix" does not fall prey to the Putnam style of argument to which "I am a brain in a vat" does, because the former does not fall prey to the sort of Twin Earth counterexamples that the latter does.[7] For our purposes, it is not important exactly what these Twin Earth counterexamples are, only that they are wild—way speculative—in the sense I've spelled out in Chap. 1. Also, I'm not interested in evaluating whether Chalmers's response to Putnam is convincing. My point is, rather, that the line of defense for the coherency of simulation Chalmers promotes enlists the same way-speculative machinery we've been tracking all along. Absent this sort of speculation, Chalmers provides us with no defense of the coherence of the simulation hypothesis.

4 Interlude: The Quick and Dirty Argument (Transition to Level 2)

Here, for purposes of a first-run Theatrical Version of the Manifest (as presented in Chap. 1) we substitute a quick and dirty axiomatic argument for the rather excruciating technicalities involved in considering the Löwenheim-Skolem (L-S) Theorem. Even if you prefer to proceed straight to the Director's Cut, be alerted that the quick and dirty argument is itself quite ingenious and of merit on its own. You might say it has all the advantages of theft over honest toil, but then that is part and parcel of its charm and interest.

What I'm calling "the quick and dirty argument" is due to Edward Nelson, and it appears in an essay called, "Warning Signs of a Possible Collapse of Contemporary Mathematics" (Nelson 2011). Nelson is one of the very few people who think that contemporary mathematics may be in peril of imminent collapse. I don't share his doomsday pessimism, but I find the argument he presents to suggest that there might be trouble that is very interesting on its own terms. To frame the argument, we begin

with an axiomatic theory: a presentation of basic arithmetic in terms of axioms, much in the way that Euclid originally presented geometry. Since we won't be particularly concerned with the formalism, I list the core axioms here in a more natural way—that is, using the English language as much as possible. We introduce two primitive symbols, "0," which stands for the number zero, and "S," which is a one-place function—that is, taking one input, and which stands for "the successor of . . ." where the ellipsis indicates the input. There are two other basic operations, addition and multiplication, that will be represented by "+" and "·," respectively. I use italicized letters from the end of the alphabet—that is, x, y, z, to indicate variables, and ϕ to indicate some property that a number may or may not have (as, for example, being even, or being a prime, or anything else you might be able to express). Here are the axioms:

AXIOM 1. No number has zero as its successor (i.e., there is no number that comes before zero).

AXIOM 2. If the successor of x equals the successor of y, then $x = y$.

AXIOM 3. $x + 0 = x$ (i.e., if you add zero to something, you get that same thing back again).

AXIOM 4. x plus the successor of y is equal to the successor of $(x + y)$.[8] More formally, $x + Sy = S(x + y)$. For example: $3 + (4 + 1) = (3 + 4) + 1$.

AXIOM 5. x times zero is equal to zero (i.e., $x \cdot 0 = 0$).

AXIOM 6. x times the successor of y is equal to (x times y) plus x, or more formally: $x \cdot Sy = (x \cdot y) + x$. For example, $3 \bullet (4 + 1) = (3 \bullet 4) + 3$.

AXIOM 7. Suppose ϕ is a numerical property and 0 has this property: we will express this by writing: $\phi(0)$. Now, if for all x, $\phi(x)$ implies that $\phi(Sx)$, then for all x, $\phi(x)$. In other words, if 0 has the property ϕ, and whenever x has the property ϕ, then so does Sx, and then *all* x have the property ϕ.Those are all the axioms for Peano arithmetic. It may not look like much, but you can do a huge amount of arithmetic using these axioms. Axiom 7 is the real kicker, because it's not really one axiom but, rather, a whole scheme of axioms, one for each property ϕ. This is a standard axiom system whose properties have been extensively studied.

Now, following Nelson, I want to add a few less standard axioms to this list. We're going to be concerned, roughly speaking, with what numbers we can "count to." Clearly I can count to 0, since it's the first natural number.

AXIOM 8. 0 is a counting number.

Next, I want an axiom that says that if I can count up to a number, then I can count up to the number following that one:

AXIOM 9. If x is a counting number, then Sx is a counting number.

You might say, "Well, what if I got tired and couldn't go on? Wouldn't that violate this axiom?" The answer is, yes it would, but remember that the axioms are intended to model a kind of ideal system. What that ultimately means, philosophically speaking, may not be so easy to say, but it's part of the axiomatic approach—that is, part of what makes it possible to make the quick and dirty argument. In any case, suffice it to say, that almost no mathematician would be worried about such an idealization. In fact, almost no mathematician other than Nelson would even be interested in an axiom like Axiom 9. Of course, if I can count to x, then I can count to $x + 1$. What's the problem? It's not like we're going to run out of numbers, right? (Right?!)

If a number can be counted to, then call it a *countable number*. In the same way, we can define what it means to be an *additionable number*. This is a number such that if you add any counting number to it, you still get a counting number. And we can also define what it means to be a *multiplicationable number*—that is, a number such that if you multiply it by a counting number, you still get a counting number. I'll spare you the details because they don't really concern us.

On the basis of these definitions, and the axiom system consisting of Axioms 1 through 9, Nelson is able to prove a couple of theorems.

Theorem 1 If x is an additionable number, then x is a counting number.

This theorem tells us that if, whenever you start with a counting number y you get to a counting number by adding x to it, then x itself is a counting number. This leads in a straight line to a second theorem, which says that the sum of two additionable numbers is also additionable:

Theorem 2 If x and y are additionable numbers, then $x + y$ is additionable.

Maybe that seems obvious, and if so, you won't be surprised to hear that the same sorts of thing holds true for multiplication:

Theorem 3 If x is a multiplicationable number, then x is a counting number.

Theorem 4 If x and y are multiplicationable numbers, then $x \cdot y$ is multiplicationable.

These results are a little bit harder to prove, but only a little bit, because the proofs go from multiplicationable back by way of additionable to countable. This makes them a little bit longer, a little more complex, but everything works just fine.

Does this seem obvious, too? Well, what if we try to go to the next operation? First, note that adding is multiple succession. For example. 4 + 3 is the same thing as "the successor of the successor of the successor of 4"— that is, SSS4, which is ((4 + 1) + 1) + 1. Adding three is the same as taking the successor three times. Similarly, multiplication is successive addition. 4 · 3, or "four times three," is equal to "four plus four plus four," which is to say "add four three times." The process of successive multiplication is called *exponentiation*. Four multiplied by itself three times, that is, "four times four times four," is notated 4^3, which is expressed as "four to the third," or "four raised to the third power." Now, what is surprising is that if we define *exponentiable number* on analogy with countable number, additionable number, and multiplicationable number, we cannot prove in the same way that an exponentiable number raised to the power of an exponentiable number is a counting number!

The problem arises because exponentiation is not associative. To take a simple example, consider:

$$4^{(3^2)}.$$

Here, 3^2 is 3 · 3, which is 9, and 4^9 is 4 multiplied by itself nine times, which is 262,144. (Notice that we've generated a decent-size number out of two uses of the exponentiation operation and some small numbers.) On the other hand, consider:

$$\left(4^3\right)^2$$

This is equal to 64^2, which is equal to 4096, a considerably smaller number! So it matters in which order you exponentiate. We express this by saying that exponentiation is not *associative*. But we used associativity to prove Theorems 2, 3 and 4, so we can't prove an analogous theorem about exponentiation in the same way.

Roughly speaking, this means that writing down expressions using exponential notation does not guarantee that we will stay "within" the counting numbers. Using the term "numerals" to refer to such numerical notations, Nelson expresses the point this way: "the belief that exponentiation . . . applied to numerals yield numerals is just that—a belief. Here we have the . . . most serious . . . warning sign of trouble in contemporary mathematics" (Nelson 2011, 85).[9]

This problem motivates our transition to the foundations of mathematics, which is the proper concern of Level 2.

Notes

1. Putnam 1983, 18, acknowledges an analogous point.
2. The one important exception to this claim I have just made, and it is a very important exception indeed, is in the consideration of what might be called "prosthetic intervention" into the human bodily condition, the history of which goes back to the Neolithic period, at least. But note that it is precisely here that we face an ineliminable question of wetware. And just as in Bostrom, such wetware considerations are methodologically ruled out by Putnam's manner of proceeding.
3. Consider, for example, his remark in "Models and Reality" that "set theory may not be the 'paradise' Cantor thought it was, but it isn't such a bad neighborhood that I want to leave of my own accord, either," Putnam 1983, 21.
4. For my purposes in this book, the most interesting response to Putnam's line of argument is that given by Hartry Field in the context of Putnam's work on the L-S Theorem, which, as he says, motivates and underlies his argument against brains-in-vats. I discuss this response in detail in the Appendix. The secondary literature surrounding Putnam's brains-in-vats argument more specifically is indicated in the bibliography provided by Kickey (n.d.). I briefly discuss Brueckner's responses to Putnam and Bostrom later in this chapter.

5. Omnès refers to Borel 1937. An even more fundamental (and even more exotic) development can be found in Borel 1952, a fundamental text in the (pre)history of the mathematical parafinite. Omnès uses Borel's position to argue for the completeness of quantum theory. See also Omnès 1999, 236, 239.
6. My interpretation of "sensible" here as logically possible—that is, logically meaningful— is not the only possible one, but I think the context of the passage I cite supports it.
7. Chalmers's point is more subtle than my description in the text would make it appear: it turns out that "I am in a matrix" and "I am envatted" are both coherent on his account, whereas "I am in the Matrix" and "I am a brain in a vat" are both incoherent. The problem with the latter claims, roughly, is that "the Matrix" (i.e., the one I've seen in the movie *The Matrix*) and brains-in-vats are both given to me through my perceptual experience in my world. As such, although they are examples of "matrix" and "envattedness," respectively, their particular possibility is ruled out. The general possibility of matrices and envattedness, however, is not. Chalmers thinks this makes Putnam's argument a red herring. On my view, it points back, rather, once again to the role wild examples are playing in purported abstraction from our perceptual experiences.
8. The parentheses tell us which operation to do first, so here we first add together x and y and then take the successor.
9. I've excerpted from the passage only to bring it in line with the restricted context of our discussion here; the full passage reads: "The belief that exponentiation, superexponentiation, and so forth, applied to numerals yield numerals is just that—a belief. Here we have the third, and most serious, warning sign of trouble in contemporary mathematics." The point Nelson is making is that we have a belief that such notations have *in general* a numerical sense, but that belief is just that: a belief.

4

Introduction to the Parafinite (Level of the Parafinite: Level 2)

1 Wittgenstein versus Turing

In 1939, Alan Turing attended a seminar in Cambridge, England, given by Ludwig Wittgenstein, on the foundations of mathematics. Alan Turing contributed, in significant ways, to our modern notion of an algorithm, the idea that launched untold thousands of computer programs. He also did pioneering work with early computers; devised the idea of an in-principle computer now known as the "Turing machine"; is responsible for the "Turing test," a thought experiment used to think about the potential for machine intelligence; and is co-named in the "Church-Turing Thesis," which is a specific proposal for how the notion of algorithm can be related to basic mathematical formalism. Ludwig Wittgenstein was in his day, and continues to be held to be, one of the greatest philosophers of the twentieth century.

The confrontation between Turing and Wittgenstein during the seminar could not unreasonably be described as a battle of giants. In this seminar, Turing and Wittgenstein carried on an extended conversation in which they talked about many things, but probably the one for which this conversation is best remembered has to do with whether bridges will fall down, whether we expect them to fall down, if there are contradictions in mathematics.

O.B. Bassler, *Diagnosing Contemporary Philosophy with the Matrix Movies*, DOI 10.1057/978-1-137-57889-1_4

The thought that there could be contradictions in mathematics was a more focal intellectual concern at the time than it is now. Debates on the foundations of mathematics reached a kind of historical peak in the 1920s, when David Hilbert first tried to prove, using a special form of mathematics called "metamathematics," that mathematics was both consistent (contained no contradictions) and complete (all derivable from basic first principles); and then slightly later, when Kurt Gödel proved results that dramatically called into question the possibility of doing any such thing. In the wake of Gödel's Incompleteness Theorem, the confidence in securing an ultimate, consistent, and all-encompassing foundation for all of mathematics waned. Wittgenstein's 1939 seminar transpired in this shadow, but in an intellectual atmosphere where the specter of mathematical contradictions was still extremely potent. Wittgenstein suggested that we shouldn't expect contradictions in mathematics to lead to doubts in the solidity of our real, physical bridges. Turing, meanwhile, stood for the more "standard" (at the time, anyway) view that should we discover basic contradictions in mathematics, this would radically call into question the security of our faith in building construction—and lots of other things, too.

Wittgenstein was making a fundamental point about the nature of logic and its relation to "reality." As with the debates concerning simulation and brains-in-vats (Chaps. 2 and 3), I'm less interested in evaluating the debate between Wittgenstein and Turing on its own terms, and more interested in using it to diagnose a sea change in our response to what we might call quite informally "living with contradiction." Even more so than in the consideration of Bostrom's and Putnam's arguments, I'll be running roughshod over the philosophical finesse on display in the Wittgenstein-Turing debate. As we descend down the rabbit hole, the diagnostic ambition must keep pace with the depth of the philosophical problem at issue, and so an acceleration is required. If you think, for example, about medical diagnosis, the idea is not so alien: medical diagnosis must be conducted in real time, with a view toward the seriousness of the problem.

We now live in a different day and age than did Wittgenstein and Turing, and I think it's considerably easier to accept Wittgenstein's basic point, at least on this issue. (When, in his last manuscript, circa 1951,

Wittgenstein wanted to assert something wildly implausible, he picked as his example that he had been to the moon! Two decades later such an assertion would no longer be so wildly implausible.) Ironically, one of the main reasons it's easier now has to do with computers—the very novelty Turing did so much to help introduce. We live, in fact, with computer errors on a daily basis. For instance, my laptop computer has a program for "synchronizing" my laptop files with the "mainframe" (to be honest, I'm not even sure on which server this information is stored), yet eight times out of ten I get a message that reads "synchronization error." Whether I should or not, I just ignore it. What am I to do? Getting in touch with the computer technician takes time, and if I chased down this problem I'd lose time that I could spend on something else, like writing this book. Sometimes, of course, it's worse: my computer doesn't boot up correctly or, God forbid, there's a crash. But even in these "disaster" scenarios, life does go on.

In any case, we know that there are plenty of computer programs that "contain contradictions" in some informal sense, and that many but far from all of them cause the programs to halt, so there are no doubt plenty of internal contradictions in our codes that, under unfortunate circumstances, could cause our systems to go down at some time. I'm intentionally not trying to express anything technical or formal here, just to give an "average" sense of how reliable computers are. They are generally reliable but break down often, and we all live in (some, usually mild) fear that something either personally or collectively catastrophic could hit us at any time.[1] It's questionable whether we want or ought to describe such situations in terms of contradictions *causing* these breakdowns; nevertheless with the proliferation of electronic technologies available to us, we all experience these breakdowns on a regular basis.

The main point is this: we don't respond to these situations by putting all programming on hold until we've verified that there are no contradictions. We don't even devote a majority of our computing resources to error checking. As with plumbing or wiring, or more generally, with buildings, we fix them as the problems manifest themselves. We try to educate young programmers in good programming habits, in hygienic principles of programming architecture, but sooner rather than later they go out into the world and write programs—fallible pieces of human software.

If they carried these principles of hygiene to their "logical" conclusion, they would become the programmer equivalents of compulsive handwashers and never get any programs written. I also don't really care to dispute how much of our computer woes should be chalked up to "incompatibilities," as opposed to "contradictions." What I'm offering isn't a fine-grained portrait but just a sketch, and for that purpose it's enough that when there's a bug in the program, the program can, and often does, halt.

On the one hand, this picture might support Turing's position to a considerable extent, since as a matter of fact, if there are software problems we can easily imagine them leading to problems with power grids or worse. In such cases, we are right to worry that incompatibilities lead to a loss of integrity in our "architecture," even literally speaking. But Wittgenstein isn't denying incompatibilities, and at a more fundamental level, I think the basic picture supports Wittgenstein's point. We've always lived with incompatibilities, and because of our contemporary experience of incompatibilities in "programming," as we call them, we are all the more explicitly aware of this term. Incidentally, I'm not sure this reason for taking Wittgenstein's point is compatible (!) with his view of why it's so—but that's another, difficult matter I don't want to get into, at least right now.

In fact, we can see the worldview of "living with contradictions" in the *Matrix* Mythology itself. The competition between the Oracle and the Architect is described as antagonism between programs, and the evolving story has to do with the way these incompatibilities, or "contradictions," get worked out. Mythologically, in the *Matrix* movies we're already in the land of evolving computer ontologies, and this reflects, in a rough-and-ready way, the point David Isles makes when he introduces evolving proof-structures. (Isles's work is discussed at length in the Director's Cut of the Manifest; see the Appendix.) At the end of the first movie, as I've mentioned, Neo is identified as an anomaly. An *anomaly* is a conflict under pressure to be resolved. Will Neo-as-anomaly be resolved through his assimilation into Smith? Or, through his sacrifice to the Matrix? In fact, it will be the latter by way of the former, and at this point the cumulative experience of the movies is to leave us anew with the question "What is the Matrix?"

Turing proposed that we think of computer programs as procedures, and when he asked what sort of procedures computers could (in principle) accomplish, he came up with a model for computation known as the "Turing machine." But whether what can be accomplished by our intuitive idea of a procedure—or "algorithm," as Turing came to call it—is equivalent to this formal model of computation is not itself a mathematical problem but, rather, a philosophical one, though one lying, metaphorically speaking, right at the interface of philosophy and computation. Part, but only part, of the problem is that there are various models of algorithm besides the Turing machine, and their equivalence is a delicate matter of some philosophical interest. What is perhaps surprising, and generally viewed as reassuring, is that if we adopt the Church-Turing Thesis (which I'm intentionally not giving explicitly), all the major formalizations of the notion of algorithm turn out to coincide. The Church-Turing Thesis is not itself a formal model of algorithms, but instead is a thesis concerning what we mean by the term "algorithm." It is widely accepted but not entirely (or perhaps not at all) uncontroversial, though in practice it usually serves as bedrock for considering the various formal models of algorithmic computation.

In the nineteenth and early twentieth centuries, different initiatives were taken to provide secure foundations for mathematics in terms of fundamental concepts that could be accepted by all. As is to be expected, there were various approaches to this project, but the one that has become most canonical through this development is the notion of a mathematical set. A set is, roughly, a collection—a grouping of objects. These objects may themselves be sets, or they may be primitive objects that are not themselves further analyzable as sets. Even in the wake of the challenge to the foundational approach posed by Gödel's Incompleteness Theorem, it is still the case that mathematicians typically take the notion of a mathematical set as fundamental and build "all" of mathematics (in some sense, which needs much more specification) from this notion. Despite the problems, there are good reasons to think that the notion of set basically does a good job serving as the fundamental mathematical notion.

What is interesting to me in this context is that the notion of algorithm does not do a similarly good foundational job. This is a difficult and somewhat technical point, but it is worth our while to pay some

attention to it. The basic reason is simple and deep, and amounts to this: no matter what version of the notion of algorithm we choose, we have to have the notion of replacing a variable by a value, and we can't (or don't want to)[2] specify in advance the *range* over which the values (i.e., inputs) are being drawn without reducing the *power* of the notion of algorithm down to the power of the notion of set. In other words, either we stick to the notion of algorithm and leave vague what all can be "plugged into" the algorithm, or we fix what can be plugged in, and this effectively reduces the notion of algorithm. To put it yet another way, either we stick with the foundational notion of set and "code up" the notion of algorithm in terms of the notion of set (but then we need something like the Church-Turing Thesis to argue that we've "caught" our intuitive notion of algorithm in terms of this formal set-theoretic model); or we go with some approach in which the notion of algorithm is fundamental (and then we face the ambiguity of what the values are that the variables in the algorithm range over).

There are various attitudes to view this dilemma. One begins by noting that the issue of the range over which quantification is taking place in set theory already poses all sorts of problems, from the technical to the philosophical. But that point can cut two ways: we might decide we've got enough problems on our hands in set theory already and that this should dissuade us from taking any approach where these issues only become more telling. Call this *drawing the conservative moral*. Or, we might *draw the radical moral* such that if we can't even solve these problems when we're being good set-theoretic conservatives, then there is reason to embrace the dilemma of quantification-over-range with radical, perhaps even reckless, abandon. That is effectively what people do when they replace set-theoretic foundations by some algorithmic "alternative"—say, a combinatory calculus like λ-calculus, for example (don't worry if you don't know what that is). A strong argument for such radicalism is that even in ostensibly set-theoretic contexts, mathematicians are regularly involved with "large domains" that beg the question of quantification-over-range; many category-theoretic arguments in mathematics (never mind if you don't know what that is, either!) fit this bill.

From a strict set-theoretic perspective, these arguments are fallacious, but there is generally a moderately permissive attitude toward

this sort of thing, which is expressed in the faith that "in principle," the arguments could be carried out coherently. Although I am doing my best to thread a path between the Scylla of "there's no problem here" and the Charybdis of "mathematics is a house on fire," mathematical emotions run high in this area, and I doubt I will have satisfied most everybody who has a strong, prior opinion. What is really needed if the issue is to be appreciated is the sort of careful attention that Colin McLarty has been giving it by focusing on real-live mathematical examples in context (McLarty 2010).

There is a second problem here, however. Even if we make the leap from a set-theoretic to an algorithmic perspective, we've not yet addressed the status of *proof*. Typically, within the algorithmic worldview, algorithms are identified with proofs: a procedure for computing 12 from 7 + 5 is taken to be a proof that 7 + 5 = 12. But on closer inspection, it turns out that there are all sorts of reasons for not being satisfied with this identification. For example, and although it may seem only to be a psychological issue, computations not only can be performed by rote but in fact their practical value is essentially bound up in their routinization. But at least prima facie a proof does not seem to be something that can be made by rote: the whole point of a proof is to establish conviction. Even if a particular proof can be routinized—"learned and recited by heart"—the recognition of it *as* a proof is not something done by rote. The same could be argued for computation—that in addition to performing a computation, we must recognize it *as* a computation. But still, the roles played by proof and computation seem different with respect to the issue of routinization.

I don't mean to resolve this issue here, I only bring it up, for it turns out that there is a close but subtle relation between proof and computation at the heart of the most interesting models of the logic of computation. This leads to something called the Curry-Howard Correspondence, which says roughly that each proof can be tracked by a computation and that each computation can serve as the "decoration" for an associated proof (Simmons 2000). But precisely this model of the relation shows that if we identify proof and computation, the delicate interplay between the two is reduced to nothing.

This leads to my suggestion for the answer to the question "What is the Matrix?" at Level 2. The answer is: *The Matrix is a Web of Proof*. Instead of

the traditional idea of foundations of mathematics, in which the security of mathematics is predicated on our surety of the notion of set, and set-theory is fashioned in a such way as to guarantee consistency and as much completeness (breadth) of mathematics as possible, I recommend we discard the notion of an independent mathematical foundation and replace it with a Web of Proof that constitutes the justification-condition for mathematics. Note that this is *not* to identify mathematics with proof: mathematics consists of many things, of which proof is only one. For example, mathematics involves stating theorems (which must be proved), framing conjectures (for which we may seek proofs), applying mathematics to real-world problems, teaching, organizing, presentations at conferences, writing grant proposals, and the like. Mathematics thus construed is an activity, but it has a special relation to its justification-condition, and we don't propose to understand this justification-condition in terms of any *antecedent*, *unified foundation*. Proofs can be given relative to axiom systems, as systems of definition, and may even be sought for ill-posed problems as a way of converting heuristic mathematics (which is mathematics) into proof-justified mathematics (which is mathematics).

As Imre Lakatos has illustrated in his series of dialogues, *Proofs and Refutations*, what is taken to be a proof has varied over time (Lakatos 1976). This is not to say that we shouldn't try to clarify the notion of proof, only that we shouldn't expect that this clarification will be ultimately satisfying or that we should wait around for an ultimately satisfying clarification before we can move along with the business of mathematics. To take only one example, but a crucial one, the axiom that asserts the existence of a set with infinitely many elements is routinely assumed by mathematicians, but the investigation of the L-S Theorem indicates that there are reasons to question the firmness of the distinction between the finite and the infinite. Nonetheless, mathematicians "get along" with this axiom, and (so far) the consensus is that working with it is turning out just fine. (There are some people on the fringes who continue to attempt to clarify the status of the mathematical infinite, but they're either called philosophically minded or philosophers outright.)

If we view "the" Web of Proof as "the" Mathematical Matrix, then we are at least prepared to see one aspect of the nature of the Matrix at large, for the Web of Proof is neither finite nor infinite—it is what I call

parafinite (Bassler 2015). Here, the natural way to see this parafinitude (but not the strongest way) is in terms of the idea of the "finite, but unbounded." At any given time, there will only have been finitely many proofs actually produced by the mathematical community at large. If we are permissive in our notion of mathematical proof, these proofs may extend well back into the Neolithic age, in which hominids "demonstrated" to each other the nature of complicated string figures, weaving patterns, and patterns of knots and pseudo-knots generally.[3] We might even argue that there is no fact of the matter exactly how many proofs there have been; if you think about differences in presentation, it's even a good question when to count two proofs of a geometric fact as either the same proof or relevantly different proofs.

This exemplifies a deeper sense in which the Web of Proof is parafinite, but that's not my focus right now. The main point is that we're thinking of proofs, on the one hand, as *activities*, so that they don't "exist" in any relevant sense prior to being supplied; but at the same time, we're thinking of them as *structures of activity*, so that they can be repeated, compared and contrasted, joined together to produce something new, and so on. Obviously, the repository of proofs is growing, since mathematical activity is ongoing, and it is in this sense at a very minimum that the Web of Proof while "finite" is "unbounded." So, what is the problem with this description?

The problem, simply put, is that this description relies tacitly on a conception of the potential infinite as something unfolding indefinitely, and arguably this way of thinking about the indefinite unfolding as potentially infinite conceptually presupposes the idea of the actual infinite, so the distinction between the finite and the infinite is once again presupposed as a definite conceptual distinction. In the seventeenth century, Pascal made the argument that the potential infinite conceptually presupposes the actual infinite, and his argument is still taken seriously, even if it is far from uncontroversial (Bassler 2015, 86). What I want to suggest, however, is that Pascal's argument opens the door to seeing that we're thinking about the indefinite in the wrong way here. In fact, the indefiniteness of the Web of Proof has to do with the very relation between logic and mathematics when they intersect in the domain of mathematical justification-conditions. To explore this point in all its dimensions would require tremendous work, and it is initiated in the Appendix.

A proof, understood as a syntactic, truth-preserving structure for the transformation of premises into conclusions, involves a tight connection of the logical and mathematical registers. The fact that the proof is truth-preserving is *logical*, and it is the syntactic structure that carries this truth-preservation; on the other hand, the premises and conclusion of the proof are *mathematical*. Together, they provide the justification-conditions for the conclusion. As David Isles has argued, over the course of the proof the range of *mathematical* expression is constrained by the *logical* demands of the syntax. Each proof tacitly specifies a range of meaning-reference over which it is valid. *This range is delimited but not uniquely specified; it is intrinsically indefinite.* This indefiniteness reflects that mathematics is always under way, in the sense that it has both already started and not yet finished. In this, it is like activity generally, and indeed can serve as a kind of constrained model for it.

That mathematics has this modeling nature (and we could make this point equally about language) means it is possible for mathematics to be properly geared, in fact, to the description of activity, but this is not an assertion I will take up here. So, the Web of Proof gives us our first exemplification, in this Manifest, of the Matrix, and also a model of it. It is, however, not a formal model because we are considering proofs not as formal entities but as activities. This model is better than the mythological model of streaming numbers we see at the beginning of the first *Matrix* movie, because numbers are more nearly objects—bits of information—and we need to move through the progression from set ontology to information ontology to algorithm to proof, and ultimately to activity more generally, which is what I call *praxis*.

While there is something wrong with the idea of identifying algorithm and proof, so that Turing's algorithmic approach doesn't give us the most adequate picture of the Mathematical Matrix, still there's also something wrong about Wittgenstein's attitude in his lectures on the foundations of mathematics. What Wittgenstein holds onto that is residually foundational is the idea of the *definiteness* of logical form. It might seem that while the reference-range (i.e., the range of inputs) associated with a proof is a dynamically evolving configuration, constrained by the ongoing structure of the proof and the substitution and identification conditions it imposes, that nonetheless the proof "itself," as a syntactic-structure (i.e., a structure of rules or procedures), is something static, non-evolving, *definite*. But in

fact the proof-structure is not something we can understand independent of the content-laden mathematical premises and conclusions involved. For this would require us to make the premise- and conclusion-places that the syntactic structure supplies themselves subject to ranging over arbitrary collections of premises and arbitrary conclusions—*but we don't even know what such arbitrary ranging would mean*.

Wittgenstein once said that what it means to say that something is red is that, for all that, if I say instead that it's blue, nonetheless it's still red. This is not, of course, to say that I couldn't mean by the term "blue" what I currently mean by "red"; rather, it's to say that what "red" *means* is specified (defined) in terms of its truth conditions.[4] But to say this is to hold on to a residual element of logical foundationalism. For all we really know is that the meaning of "red" is constrained by its truth conditions, and we have every reason to believe that the extent of this constraint is something we're not in a position to specify in advance. That, too, is "on the way." Wittgenstein was a notoriously intransigent thinker, and his power comes from this intransigence. When he made the famous "turn" from his earlier *Tractatus* view to the later view represented by the *Philosophical Investigations*, he "liberalized" his views in a profound way (in fact this liberalization had to do, on the one hand, with the finite/infinite distinction and on the other hand, with the status of color terms), but he never liberalized on the issue of logical definiteness (Bassler 2015, 189–97). This further step leads to what I call *paraphysics*.

We can see this transition in terms of the rejection of a phenomenological language that Wittgenstein argued for.[5] The argument for the rejection of a phenomenological language that Wittgenstein gave is well presented in David G. Stern's *Wittgenstein on Mind and Language*, and I follow his presentation here. The argument was developed in the crucial period around 1929, when Wittgenstein was shifting away from the philosophical orientation indicated in the *Tractatus Logico-philosophicus* and moving to the later position of the *Philosophical Investigations*. A phenomenological language would be a complete and accurate description of contemporaneous experience, or as Stern calls it "a canonical analysis." The sense in which such a phenomenological language constitutes an analysis is well expressed in a passage from Wittgenstein's essay from early 1929, "Some Remarks on Logical Form":

> [W]e can only substitute a clear symbolism for the unprecise one by inspecting the phenomena which we want to describe, thus trying to understand their logical multiplicity. That is to say, we can only arrive at a correct analysis by, what might be called, the logical investigation of the phenomena themselves, i.e., in a certain sense *a posteriori*, and not by conjecturing about *a priori* possibilities. (quoted in Stern 1995, 135)

The language of phenomenology would be the language in which this analysis was conducted, and it would lay bare the "logical multiplicity," which is to say the logical structure, of the phenomena at hand. To say that this language is "in a certain sense a posteriori" is to say that the language is a descriptive one, but one that purifies the presentation of experience of the logical opacity we find in ordinary descriptions of experience.

Wittgenstein was to reject this notion of a phenomenological language before the year 1929 was out. In a passage from the *Philosophical Remarks* drafted in October 1929, he wrote that he no longer had "phenomenological language, or 'primary language,' as I used to call it, in mind as my goal. I no longer hold it to be necessary. All that is possible and necessary is to separate what is essential from what is inessential in *our* language" (quoted in Stern 1995, 136). Somehow, the removal of opacity from "our" language is to be performed without the provision of a logically transparent language; in a closely associated passage from the *Philosophical Remarks,* we find the crux of Wittgenstein's argument against a phenomenological language: "How strange if logic were concerned with an 'ideal' language and not with *ours*. For what would this ideal language express? Presumably, what we now express in our ordinary language; in that case, this is the language logic must investigate" (quoted in Stern 1995, 136). In other words, if we attempt to replace "our" language by a logically purified one and analyze the logical structure of this now logically transparent new language, then in fact we are not analyzing the structure of "our" language! For Wittgenstein, we cannot resolve the problem of the logical structure of our language by substituting another language.

A closely associated problem is revealed in Wittgenstein's analysis of the attempted claim that "all is in flux." Wittgenstein finds this traditional expression of the Heraclitean doctrine unsuccesful because, effectively,

the fixity of the expression violates the flux it attempts to express. For something to be fixed in language, according to Wittgenstein, we have to be able to imagine it otherwise. But what "all is in flux" attempts to express is so fundamental that we cannot imagine its opposite (Stern 1995, 161). This means that the assertion is inadequate to what it seeks to express, and so what it seeks to express cannot be expressed, but only indicated—"shown."

There is, in a sense, an inadequacy of language to the expression of such a fundamental "essence," and yet there is no sense in which we could hope to eliminate this inadequacy by moving from one language (ours) to another (purely phenomenological, or primary). That is to say, there is something *indefinite* in our language, something which reflects the indeterminacy (flux) of life, and this Wittgenstein indeed explicitly asserts: "If a pattern of life is the basis for the use of a word then the word must contain some amount of indefiniteness. The pattern of life, after all, is not one of exact regularity" (quoted in Stern 1995, 190). Furthermore, Wittgenstein insists that this does not imply any inadequacy or defectiveness in our language. Wittgenstein has permanently and thoroughly rejected the notion of a purely definite, fully analyzed phenomenological language.

But rejection of a phenomenological language, and the linguistic definiteness thereby implied, does not amount, in Wittgenstein, to a full rejection of *phenomenological definiteness*. To begin with, the notion of a phenomenological language is replaced in the *Philosophical Remarks* by the notion of a phenomenological investigation, and this notion of investigation will continue to serve as the model underwriting the general philosophical orientation of Wittgenstein's late philosophy as expressed in the *Philosophical Investigations*. What has changed does not in any sense involve a rejection of the ideal of logical definiteness, but only a change in view about where and how this logical definiteness is to be located: "Am I not getting closer and closer to saying that in the end logic cannot be described? You must look at the praxis of language then you will *see* it" (quoted in Stern 1995, 190; emphasis mine). The question seems to be "rhetorical," and yet it is difficult to see how the claim at issue—that logic cannot be described—escapes the very problem faced by such a claim as "all is in flux." I will not attempt to resolve this issue, but it points to an

impasse in Wittgenstein's own philosophical praxis. What is needed is a rejection not just of the phenomenological definiteness of language but also of the phenomenological definiteness of logic. This latter, as I understand it, is tantamount to a rejection of phenomenological philosophy per se, and is a step that I do not read Wittgenstein as taking. It *is* a step motivated by, and perhaps even implied by, the program for an "evolving proof theory" along the lines of Yessenin-Volpin and Isles, which I discuss in the Appendix.

The work of Yessenin-Volpin and Isles has repeatedly been criticized for being difficult or impossible to follow, and it is true that the level of technical difficulty is quite severe. Isles himself admits that it is unlikely his alternative proof theory will be readily adopted, since it requires us to give up the idea of a fixed reference ontology–but then so does the very notion of algorithm! The technical challenge here masks a more fundamental source of discomfort, which has to do with what is truly radical in Isles's program. Isles enlists the support of Wittgenstein's philosophy, in fact, to promote this philosophical radicalism. But if I am right, the radical implications of Isles's work, and the work of Yessenin-Volpin which stands behind it, go beyond the philosophical reorientation that Wittgenstein's work demands. If, in particular, we can see the way in which Isles's proof-theory requires the rejection not just of a language of ideal definition but also of a logical ideal of definition, then we will be in no position to replace the function of phenomenological language by phenomenological (philosophical) investigation in the way Wittgenstein hoped.

But why should Isles's proof theory motivate the rejection of logical definiteness? So to speak, Isles has insisted that the reference ontology "plugged into" a syntactic proof-structure yields, in general, a different output ontology. On analogy with the discussion of algorithm given earlier in this chapter, we may say that the domain of the input ontology has been unfixed, rendered indefinite. In an evolving proof theory, the logical dimension is specified in terms of the structure that converts premises into conclusion in a truth-preserving way. But crucially, this structure is no longer tied to the notion that reference remains fixed across this structure in the way it is assumed to be in "traditional" proof theory.

When Wittgenstein concluded that language must admit of indeterminacy, it was because the language was geared to the expression of "forms of life" that were themselves not fully definite. The point is that now we find ourselves in exactly the analogous situation with respect to the relation between *logical structure* and reference ontology. If the reference ontology is logically evolving, that just means it is not fully *definite* in the traditional sense, and the logical structure is geared to reflect the *indefiniteness* in this underlying reference ontology. *This, then, is precisely what it means to reject logical definiteness*. And such logical definiteness is in fact presupposed by Wittgenstein's conception of philosophical investigation—logic is the fixed, underlying essential structure that cannot be imagined otherwise, and therefore can only be shown.

In fact, it is just this philosophically intransigent distinction between logical structure and "reality"—what I have been calling reference ontology—that lies at the heart of the standoff in the Wittgenstein-Turing debate. This is even clearer if we look at a second-generation version of this debate, appearing in the writings of Charles Chihara and Stuart Shanker (Chihara 1977; Shanker 1987). Chihara and Shanker both rigidify the positions of Turing and of Wittgenstein, respectively, in a rather academic way, but this does have the virtue that the "rub" becomes even clearer. Chihara essentially represents Turing's orientation and Shanker presents Wittgenstein's—what becomes clear in Shanker's response to Chihara is precisely that it is discomfort with Wittgenstein's insistence on a lack of causal, ontological connection between logic and reality that drives the responses of Turing and Chihara to Wittgenstein's position.

Shanker insists, rightly enough, that for Wittgenstein, logic cannot cause anything at all (Shanker 1987, 252). But if we reject the ideal of logical definiteness Wittgenstein hangs onto, there is no need to insist on the categorical distinction between logic and reference ontology. In fact, Isles's program shows precisely how reference ontology *is* transformed by the underlying logical structure of proof. But when we do so, the merits are redistributed to both sides of the debate, which in some sense is effectively "brooked"—although in a way that is not likely to satisfy either of the competing parties! On the one hand, logic and reference ontology are connected, so that the distinction between "logic" and "reality" is relativized in a way that makes manifest the underlying commitment to

logical indefiniteness. But this connection is itself still *logical* rather than *causal*. What is needed, then, is a more radical reevaluation of the status of logic than anything we find in either Wittgenstein or Turing.[6] In particular, Wittgenstein's conception of philosophical investigation must be rejected. A similar problem saddles Husserl's conception of phenomenological investigation, which we consider in some detail at Level 3.

2 Excursus: A Poetic Analog

All this no doubt sounds very esoteric, and indeed it is, but we can find analogs in the poetic domain that are at least geared to the concreteness of our daily experiences, and their concreteness leads us more powerfully back to the filmy proximity and confusion of reality and dream. Here, I consider Mark Strand's "The Story of Our Lives" (Strand 1980, 97–103). In the process, too, we'll see a connection to what I've previously called *pace* (Bassler 2012), and a way to translate ourselves out of the phenomenological impasse in terms of pacing. As a transition, let me begin, first, with a nod to issues of pacing in the mathematical domain.

In the "quick and dirty argument" (Chap. 3), Nelson showed that the attempted proof that the exponentiation of an exponentiable number is a counting number fails. To introduce a shorthand notation, we can say that Nelson showed the attempted proof of the *totality* of exponentiation fails, since the exponentiation operation cannot be proved always to generate counting numbers. In Chap. 3, this "quick and dirty argument" is a proxy for the more detailed consideration of exponentiation that David Isles gives, and which is treated in the Director's Cut in the Appendix. In the more complicated, but in some ways analogous argument David Isles makes that the proof for the totality of the exponentiation operation fails, he identifies a circularity that has to do with the way non-numeral exponential notations[7] are getting reduced to numerals in the process of attempting to prove the totality of the exponential operation for the numbers these notations name. That means that in order to prove the totality of the exponential operation for *numbers*, we already have to rely on the totality of the exponential operation for *names*, since the reduction of a non-numeral exponential notation like 2^3 to a numeral notation like

|||||||| really just *is* the result $2^3 = 8$ at the level of numerals rather than of numbers. It "just is" this exponential result because the reduction of the one numerical notation (2^3) to the other (||||||||) is an *exponential* reduction. Isles's analysis has the virtue that it shows much more explicitly than Nelson's "quick and dirty argument" the way in which the attempted proof of the totality of the exponeniation operation is question-begging: it is tacitly assuming what it seeks to prove.

One way to state the question-begging at issue here is to say that the implied notation reduction *keeps pace* with the operation at issue. If the notational reduction is slower than the operation at issue, then the proof is not question-begging. The notational reduction for addition goes at the rate of the successor operation; the reduction for multiplication goes at the rate of addition, so these are both okay. However, if the notational reduction is equal or faster in pace, then the proof *is* question-begging, and this is the case for exponentiation, where the notational reduction goes at an exponential pace. This frames the issue in terms of the *pace of functions*: the notational-reduction function versus the operation at hand.

From another angle, it turns out that if we ask questions about the completeness of mathematical systems—say, arithmetic—we can show that functions growing at a sufficiently rapid pace cannot be proved total in the standard presentation of arithmetic. This, in turn, is linked to a way to look at the Gödel Incompleteness Theorem, which frames it in terms of rapidly growing functions. In fact, there is a general framework in which we can use rapidly growing functions to test the features of particular mathematical systems. And all of these functions are "finite" in the sense that for each numerical input, an output can be produced (in principle) by a finitary algorithm. But in fact there are good reasons to think that, while the apparatus involves nothing but finite numbers and input/output operations, the system as a whole is not finitary in the most fundamental sense. The typical conclusion is that it is *infinitary*, but this seems odd, too, given that everything we are dealing with falls under a rather strict definition of algorithm. I recommend that we think of these concerns about the growth rates of (recursive) functions as *parafinitary*.[8]

These issues of pacing become much more concrete when we look at ordinary or poetic language. In particular, how can engaging in a

description of our experience help but slow down this experience? In the limit, the attempt to give a complete description of our experience would reduce this experience to a kind of surreal paralysis of melting time, engendering a phenomenological analog of Daliesque paranoia. This atmosphere is part of what is caught in Mark Strand's poem, "The Story of Our Lives," which begins: "We are reading the story of our lives / which takes place in a room" (Strand 1980, 97). In the room we (you and I, that is) sit on a couch next to each other, and the story tells what we say about the couch, that it is "ideal." Our saying that it is ideal matches the couch, which is indeed ideal, and the story says this, too. The room in which we read the story of our lives is "almost as if" the world is generating a sense of the claustrophobia that such a closed description of our experience induces. But this telling, which is a telling of our experiencing and yet is part of the experiencing itself, and which threatens to cannibalize our experience, comes to seem unreal, the way a film often tries to capture the intensity of experience bordering on irreality by gearing down to slow motion. It is like the destination I can never reach in my most frequently occurring nightmare (and maybe yours, too).

So, the description cannot *keep pace* with the reality it would seek to describe, and we are beset in a way with a problem that seems at once the same and the opposite of the one we faced mathematically. That is, it is the same because there's something futile in our attempt, something about what we aim to do in describing our experience totally that frustrates our attempt to do it. As Wittgenstein said, language is frustrated in its attempt to say "all is in flux" (Stern 1995, 160–67). And yet there's something that's seemingly opposite here because our language seems to be too slow to keep up with our experiencing, inevitably arresting this experiencing, boxing it in. But that just means that if we looked at what language would have to do *not* to be boxed in, it would need to presuppose the very experience it sought to ascribe. In this way, the problem seems the same as (or at least similar to) the one we faced about mathematics. The experience of considering the problem now itself seems to whirl madly, spinning out of control.

But Strand's poem goes beyond just this. The second section begins: "We are reading the story of our lives / as though we were in it, / as though we had written it" (Strand 1980, 97). The point this declaration drives

home is simple but pervasive: when we read (or dream), we fall into our reading (or dreaming) in a way that becomes confused with reality, and so confuses reality. Strand interpolates passages from the stories of our lives into his poems in italics, and often these italicized sections say the same thing as the experience the poem describes; sometimes, the poem claims, the accuracy is astonishing, beyond our capacity to imagine. The story it captures is uncanny, and in being uncanny, it captures the uncanniness of reality, and then the two uncanninesses become uncanny in their interposition. As we descend into the next level of the poem (just as we descend lower and lower into the Matrix), the story takes over our lives and our lives become the story of our lives.

What would life be without story? There are mysterious parts to the story, as mysterious within the story as our dreams are within our lives. And then these stories are stories of dreaming and we descend farther down. All the while, we are spinning, spinning, spinning, experiencing the intensification of the story that is the story of our lives and that paces our lives with its meaning; but then in its mysterious parts it overtakes this meaning with outpacing, and there is no fixed point in the book—only too many dark approaches and narrow escapes. And then as Wittgenstein will insist that logic cannot be said, only shown, there will be no explanations, only revelations. And then follows a waiting worthy of Samuel Beckett, and a silence worthy of Samuel Beckett, and a nothing worthy of Samuel Beckett. The Strand poem ends in a final nightmarish convergence of themselves (ourselves) and the book.

We can read "The Story of Our Lives" as the poetic refutation of a phenomenological language, a language in and of the flux itself, but I think a more powerful way to read it is in terms of the de-realization that attends our attempts to totalize our experience in language. The power of Strand's poem is predicated both on our desire for foundations and on our recognition of their inaccessibility. In our experience, the indefinable border between dreaming and waking is a paramount figure not just for indefiniteness but also more specifically for the indefiniteness of the boundary between the paradigmatically indefinite (dream) and the definite (reality). And any attempt to "define" this indefinite boundary is outpaced by the indefinite itself, so that definition never "catches up." And yet, as we know, somehow we do manage to fall asleep, dream, and awaken.

Strand's poem powerfully attends to the inadequacies, paradoxes, and even nightmarish aspects of description by focusing on the "boundary" between description and reality and its inherent indefiniteness and instability. But this only amounts to the poetic analog of an argument against a phenomenological language, and not against the inadequacy of phenomenology itself. For this, more needs to be said about the nature of experience and justification, just as more needed to be said above about the nature not just of mathematical language but also of proof. What we have learned in Chap. 3 from Nelson, and in the Appendix from Isles, is that although the failure of total justification is involved with language, it can be traced to more fundamental features of mathematical proof itself.

Poetry has the power to shock us into recognition, over and over again. But then, inevitably we take our distance from the poem, and the shock of reading it fades, and we are back in the "old world" where phenomenology seems plausible and desirable, and desirable because plausible, and plausible becomes desirable. When Wittgenstein, in a flash, asks "How could logic ever have been concerned with an ideal language?," he registers this shock; but as later thinkers have acknowledged, among them such seemingly contrary figures as Jacques Derrida and Leszek Kolakowski, the metaphysical urge always slips back upon us: we are revisited by the "search for certitude." Kant insisted that reason contains an internal drive toward ultimate explanations, and only counseled that we rationally guard ourselves against it. Post-Kantian philosophy cultivates this counsel while recognizing the labyrinth into which it draws us, for the ruses of reason are truly labyrinthine. And thus we are drawn back into the arena of dream, truth, and the dream of truth.

3 Matrix Mathematics

How does mathematics get along so well if it is lacking foundations, if even quite possibly it is full of contradictions? The answer is that it gets along much as our computer software gets along: we check for "errors" as vigilantly as is possible within the bounds of constraints that are generally practical and tacit, and only exceptionally explicitly theorized, and we fix the errors as we go along. My mathematics dissertation adviser, for

example, is reputed as a particularly vigilant reviewer of mathematical manuscripts. He told me that on reviewing manuscripts submitted for publication, he often finds significant errors—meaning errors that were more than just mathematical "slips of the tongue"—but that they are generally far from fatal. With some further work, these errors could be "repaired," but work is required. In addition, it is important to stress that the justification-conditions for mathematics are strict and mathematicians tend to be justifiably (!) conservative in applying them.

In the nineteenth century, there was a real threat of the mathematical community's dividing into subcommunities along roughly national lines, and the Italian geometers of the Cremona school, in particular, were brilliant mathematicians but inordinately difficult for those outside this school to understand. People in the mathematical community often point to the rigor that foundational ventures supply, but historians of mathematics will tell you that the establishment of a mathematical lingua franca and an international standard of communicability were also important parts of the overall venture. In the twentieth century, and now into the twenty-first, a different problem has faced the mathematical community. In one sense, the standard of universal mathematical communicability has been all too successful; as a consequence, there has been such an explosion of mathematical publication that so many mathematical results are being generated—sometimes in nearly the same language but more often in a language that is related but relevantly distinct—that different mathematicians are routinely proving new results multiple times in slightly different forms, without recognizing the work of their colleagues as nearly identical.

Peter Johnstone has noted how "theorems and techniques which are commonplaces in one field are laboriously and imperfectly rediscovered in adjacent ones" (Johnstone 1982, xx),[9] and this phenomenon highlights the existence of different, relatively independent subcommunities. Although this sort of reduplication is less than ultimately efficient, on the other hand, it does have its reassuring side. If these same results are being repeatedly but (relatively) independently generated, it does serve as some indication that the mathematical domain continues to remain internally coherent. Perhaps in this reduplication, ironically, we have the best "praxical" proof for the coherence of contemporary mathematics. And in the absence of internal incoherence, we may ask, Why worry about foundations?

In the long run, the nineteenth-century emphasis on a common language of mathematics may prove much more significant than the idea that mathematics be grounded in "solid" foundations. What is important in mathematics is not "rock-solid" foundations in an unquestionable domain of mathematical objects (sets) but, rather, a tightly interconnected web of justification-conditions of a special sort: a web that is unprecedented among other domains of investigation. Understanding the tightness of these connections will emerge more naturally and powerfully from a consideration of the nature of proof than from any "grounding" of mathematics in a privileged collection of objects. In a way, this points both to what was both right and wrong in Putnam's proposal. Generally speaking, Putnam's insistence on the role of proof is prescient,[10] but rather than a proof-based semantics, which is not really quite the right direction, what we need is a shift of view from semantics to proof—that is, to a focus on the way proof *binds* semantic concerns to syntactic ones.

While the notion of grounding mathematics in notions of mathematical objects like sets or numbers may not be the right way to look at mathematics, mythologically the picture of the Matrix as a stream of numbers and symbols is far from bad—indeed, it is highly suggestive. If the Matrix is depicted at the beginning, and throughout the *Matrix* movies, as a stream of numbers and other symbols, we would do well to ask about the mythological significance of these symbols. The question of proof as justification-condition is an inherently praxical one—"What will I take to be a proof?"—and this anticipates our concerns at Level 4, the "Level of Philosophical Praxis." Our considerations at Level 3, though, will first be directed toward looking at the mythology of number in the *Matrix* movies, and this anticipates Level 5, the "Level of Work on Philosophy and Myth." Generally speaking, as we move deeper into Matrix Philosophy, the levels become entangled in all sorts of complicated ways. At Level 3, we will ask the question "What is the Matrix?," not as a question in foundations of mathematics, but as a motivation for paraphysics—a philosophical program I outline in the next chapter.

Notes

1. Shortly after writing this, forty pages of this manuscript were corrupted during an attempted "synchronization." Fortunately, I had backed them up elsewhere. At this point, I got in touch with the computer technician.
2. The difference between "can't" and "don't want to" is huge, of course, but beyond my purview in this thumbnail sketch.
3. I am using the term "pseudo-knot" to refer to a configuration in which there is no ineliminable knot (if, say, you are allowed to pull on the string), but in which a pattern is manifest due to limitations imposed on the capacity for "unweaving." The simple "slip-knot" is an example of a pseudo-knot. It can be held in place by interposing a stick, for example.
4. Among many "Wittgensteinians" it's seen as tasteless to saddle Wittgenstein with this, which looks at first blush like a commitment to Tarski's T-schema. That's not quite what I'm saying–the T-schema is a fully explicit logical idealization, and Wittgenstein himself bears no analogous commitment. But the commitment to logical definiteness that commitment to the T-schema tries (unsuccessfully) to reflect *is* shared by Wittgenstein (Bassler 2015, 189–99).
5. The point can also be made, in an even more explicit way, in terms of the rejection of phenomenological philosophy to which David Isles's Buridan-Volpin structures lead (on the latter, see the Appendix).
6. The point I am making here about Wittgenstein's "anti-causalism" runs parallel to the discussion of Frege's and Husserl's anti-psychologism in Bassler 2015, 119–51.
7. A numeral notation is a name for a number.
8. There are technical issues about primitive versus general recursion that I'm suppressing here, and which are controversial. Tait 1981, takes a hardline position within the mainstream and insists that only primitive recursion should be viewed as finitary. If one takes this view, then what is at issue is not finitary; but since it is *recognized as finitary* by much of the community, it at least makes sense to suggest that it is somehow vexed with respect to the traditional distinction between the finite and the infinite, and that it therefore satisfies a third description of parafinite. In fact, at the end of the day I'm not convinced that Tait's position is "hard" enough: a still more conservative candidate would be the sub-exponential hierarchy, and Isles's work reinforces this suggestion, as does, in a differ-

ent way, the work of John Mayberry; his analog of Nelson's and Isles's results is stated in Mayberry 2000, 348. Ultimately I think the issue can only be treated reasonably by abandoning the debate about what is "finite" and what is not, and by asking rather about various candidate "cores" for mathematics. Girard (2011, 448–50) has suggested, and in the end I think reasonably enough, that the hyperfinite fragment of arithmetic serve as a core. This comes from a different vantage, but it's a better place to start than focusing on the nature of recursion per se.

9. Johnstone identifies this as a consequence of the "enormous increase in the number of practicing mathematicians since the 1930s." On the other hand, he notes of a mathematician like Marshall Stone that, "although his interests may lie in one particular area of mathematics, [he] has nonetheless a sufficiently general perspective on the whole subject to recognize the significance of his work for other fields" (Johnstone 1982, xx).
10. This is discussed in detail with respect to his paper on the Löwenheim-Skolem Theorem in the Appendix.

5

An Introduction to Paraphysics (Level of Paraphysics: Level 3)

1 Two Mythologies

On our way down the rabbit hole of Matrix Philosophy I've enlisted the help of a number of companions: Bostrom, Putnam, Wittgenstein, Nelson, and (in the Director's Cut, in the Appendix) David Isles. In this chapter, I begin by invoking the work of Gregory Chaitin, but in a somewhat different spirit. We have passed the level at which it makes sense to look at work in the foundations of mathematics for its value as work in the foundations of mathematics. Chaitin promotes an approach to proof via what he calls Algorithmic Information Theory, and this will effectively (if not indeed intrinsically) involve the identification of algorithm and proof I have argued against in previous chapters.

As Chaitin himself says, his approach is "very 1930's. All I add to Turing is that I measure software complexity, I look at the size of computer programs" (Chaitin 2007, 324). So, if we were to look at Chaitin's work simply as a venture in the foundations of mathematics, it would constitute a step backward. As a matter of fact, my ultimate view about work at Level 2 is that we need to take a more pluralist approach, since there seem to be almost inevitable tradeoffs as a consequence of the

O.B. Bassler, *Diagnosing Contemporary Philosophy with the Matrix Movies*, DOI 10.1057/978-1-137-57889-1_5

particular approach any work in this area endorses: if we focus on proof, we do an injustice to algorithm; if we focus on algorithm, we do an injustice to proof; and if we focus on the interaction of algorithm and proof (as in the Curry-Howard Correspondence) we do better, but we train our eyes, at least temporarily, away from a whole host of other equally important issues. Mathematics is indeed the "science," the *scientia* in the Latin sense, of a life form that is incapable of doing everything at once. As Husserl remarked, if there is a God, he would have no need for mathematics (Husserl 2003, 202).

So, my approach to Chaitin's program is different: I am interested in it not so much as a proposal to clean up the foundations of mathematics as an incipient *mathematical mythology*. Chaitin introduces a "mysterious" number Ω (omega), which he even calls the *oracle*. Names are important: we are heading in the direction of *Matrix* Mythology. Let us look at the way Chaitin describes this number and its magical properties. In a collection *Formulas for the Twenty-First Century*, Gregory Chaitin provided an explanation of the Ω-number according to the specifications for inclusion in this volume—namely, that the description consist of 120 words or less. Here is the description he provided (Chaitin 2007, 333):

The Halting Probability Ω: Concentrated Creativity

> The number Ω is the probability that a self-contained computer program chosen at random, a program whose bits are picked one by one by tossing a coin, will eventually stop, rather than continue calculating forever:
>
> $$\Omega = \sum_{p \text{ halts}} 2^{-|p|}$$
>
> Surprisingly enough, the precise numerical value of Ω is uncomputable, in fact, irreducibly complex.
>
> Ω can be interpreted pessimistically, as indicating there are limits to human knowledge. The optimistic interpretation, which I prefer, is that Ω shows that one cannot do mathematics mechanically and that intuition and creativity are essential. Indeed, in a sense Ω is the crystalized, concentrated essence of mathematical creativity.
>
> —*Gregory Chaitin*

When I use my word processor to count the number of words in this description, it gives me a tally of 106 words. But how exactly is it counting the words? Do we count the included formula as no word, or one word, or several words? Are the various appearance of Ω to be counted as words? In fact, in our everyday use of language, we don't have one fixed method for counting the number of words in a passage, especially if it includes mathematical formulas. Such a method can be fixed with precision, but the work needed to do this is more than you might expect. Practically, what we do instead is rely on word counters inside word-processing programs and the like. And when we get an answer like 106, and the limit is 120 words, we conclude that we are "safely" within the parameters. This may all seem like nit-picking, but careful consideration of the problem shows that this issue lies at the heart of a great, big problem.

To see this, consider the following definition: The least number not definable in 37 symbols. If I've counted correctly (and you'd better check), there are exactly thirty-seven symbols in this definition. But then, this means that I have defined the least number not definable in thirty-seven symbols using just thirty-seven symbols. What's going on here?

The typical way to resolve this problem is to note that we are tacitly relying on some definite, therefore formal, characterization of what it means to be definable in x symbols for a variable quantity x. If this characterization is really definite, which is to say it is formal, then what you find in the definition above most likely won't even count as a formula. Why? Because it's in English, which is not a formal language. Doesn't that feel like cheating?

Okay, enough already. My point is just that we make constructions in some formal, mathematical context and then we interpret them in our ordinary, everyday language—English or whatever your preferred language may be. That transition is *mythological*: there's no more "possible" transition from formal language to English than there is from a syllable to blood.[1] Pessimism, or optimism, is a mythology, or in the generic sense of the term, a philosophy. How did we get from the Ω-formula to optimism? We got there because, as Chaitin straightforwardly says, optimism is his preference. Preference is a function of will: it is a choosing, a doing, an activity. Mind you, this choosing need not be fully conscious, thematic, or even thematizable. In limiting cases it may not be conscious at all, but that, too, seems like a mythological issue.

Go to the oracle: you will find the concentrated essence of mathematical creativity. Possession of the oracle will put you in touch with the answers to all your mathematical questions. But as we know from watching Neo, the oracle can only tell us what we are already in a position to understand. And as Gregory Chaitin tells us, the oracle is uncomputable, "irreducibly complex," in some sense radically inaccessible. Is that really so surprising?

What may be more surprising is that, in fact, there is a sense in which we can write computer programs to provide partial approximations to the Ω-number. But it also turns out that by computational standards these approximations are very difficult to compute in terms of computing resources: we have to use all the resources available to get an approximation for this particular program or computer. What we can't know is the rate at which these approximations are converging on the Ω-number, and so practically speaking, these approximations are of no use at all. Mostly what we know about Ω has to do with how "inaccessible" it is. For example, a formal axiomatic system (FAS) will be able to compute only as many bits of information about Ω as the bits of information that count the complexity of the formal axiomatic system itself (Chaitin 2005, 132). That means, effectively, that there is no "meaningful" way to trick the oracle into revealing her secrets. What we can know is in some very strong sense what we already know. And so an oracle starts sounding like a very paradoxical (or parapsychological) thing, indeed—which, if you've watched the *Matrix* movies, is something you already know.

All this "mystery" has to do with the fact that Chaitin's fundamental approach to information, while conceptually natural, makes it difficult to calculate the information associated with any particular object. To simplify this somewhat, suppose that I want to compute the complexity of the number π. Chaitin's approach tells us that we should define this complexity in terms of the simplest algorithm that can be used to generate this number. Now, suppose we have some algorithm that does the trick; how could we possibly tell if it's the simplest one? We're in a position somewhat like the one taken by Edward Nelson or David Isles, but the routes these two take to confront the problem are very different from the route Chaitin takes. Both Nelson and Isles admit straight up that they are only looking at particular proofs of the totality of arithmetic,

multiplication and exponentiation. These proofs *seem* to be very natural, to be sure, but Nelson and Isles recognize that their claims are only about these particular proofs. Chaitin, on the other hand, unapologetically characterizes information-content by quantifying over all possible algorithms. But what exactly does it mean to "quantify over all possible algorithms"? That involves a lot of presuppositions about the nature of computation involving the Church-Turing Thesis, how computation is modeled set-theoretically, and so on. So far from advancing beyond set-theory, Chaitin appeals to it unabashedly. The mathematics of Chaitin's Ω starts to look like an allegory of all of mathematics "writ small." We can prove theorems about Ω in the way that we can prove theorems in mathematics at large. These theorems will never exhaust Ω, just as the body of mathematical theorems will never exhaust mathematics. Thus, Ω is an *allegory* of mathematics' concentrated creativity, and Ω is an oracle in the sense that every theorem of mathematics finds its representation in this allegory.

It is important to note that Chaitin does none of this naïvely. Chaitin's heroes are, after all, Turing and Gödel. From Turing, Chaitin has inherited the notion of algorithm, as well as, as he notes, his approach to Gödel's Incompleteness Theorem! (Chaitin 2007, 324). From Gödel, Chaitin inherits his conviction that set-theoretic foundations are not only compatible with open-ended mathematical creativity but also in fact are *just the right foundation* for supporting mathematical creativity. This is somewhat paradoxical at best, given that we are appealing to an ontology of fixed mathematical objects to support an optimistic worldview of unlimited creativity.

And while we're at it, let me acknowledge a potential paradox in my own position. It may seem that I've gone to an awful lot of trouble to criticize the set-theoretic view, and in particular the idea of quantifying over a range of fixed mathematical objects, if now I'm simply going to let it all back in as mythology. Indeed! To address this challenge, I need to look at how an example of mythologization works, and the example I want to look at is Darren Aronofsky's 1997 movie π, a lower-budget cousin of the *Matrix* blockbusters soon to follow.

When Gregory Chaitin states his mathematical results about the number Ω, he's not yet engaged in mythology, unless you happen to think

that mathematics itself *is* mythology. But when he calls this number an "oracle" or the "concentrated essence of mathematical creativity," then he's crossed the line into mythology. Since this mythology takes mathematics as its subject matter, I call it *mathematical mythology*. Aronofsky's movie π does mention some mathematical results in passing, but it's really a work of mathematical mythology from the get-go. Whereas Chaitin's oracle number has an unending decimal expansion and, in a sense that can be made rigorous, encapsulates an infinite amount of information, the main number at issue in π is a 216-digit number (let's call it the "216-number") that encodes the ultimate "name of God." (The name has 216 Hebrew characters, but why the 216 digits correspond to 216 characters is never explained. It's a movie, after all. Unlike mathematics in some ways, movies require a "suspension of disbelief." That, too, is a characteristic of myth. I don't know about you, but I've never seen a phoenix rise up out of its own ashes. I've never even seen a phoenix.) What the two scenarios do have in common, however, is the idea of an essence concentrated into a number.

But the parallel doesn't end there. In both cases, the point is not really the number per se, but as our hero Max puts it in π, the connections expressed by the number. Even if we were in full possession of a decimal expansion for Ω, we would still have to know how to read it, how to "extract the concentrated essence of creativity" from it. Even if Max gives the rabbis the 216-number, he says, they wouldn't understand what it means (so why does he make such a fuss about giving it to them, then?). Only Max can do that because Max is the one who discovered it. Chaitin didn't discover the concentrated essence of mathematical creativity per se, because Ω can't be computed, but he did "discover," according to his own mathematical mythology, that mathematics is an open-ended activity whose creativity is "concentrated" in Ω. That may not be the name of God, but mythologically speaking it's still pretty good.

Max discovers the 216-number while looking for patterns in the stock market. The first year of my teaching career, I served as a lecturer in the philosophy department at Yale, before coming the following year to the University of Georgia, where I've been ever since. During the first of the two semesters I taught there, I attended a lecture by Benoit Mandelbrot. Mandelbrot had a position in the Yale Mathematics Department, but as

I later learned, among his earliest papers were some dealing with changes in the price of cotton in the United States in the years following the Civil War. In his lecture, Mandelbrot spoke on the general problem of fluctuations in prices, and he strongly insisted that especially large fluctuations were seriously underrepresented by contemporary mathematical models of the stock market. This underrepresentation, he warned, could have serious implications in particular for the way markets are insured against collapse. (It would be interesting to know if large fluctuations of prices contributed in any way to our recent economic debacle.)

The next semester, Mandelbrot offered a semester-long course on this subject, though to say the course was on this subject is a serious underrepresentation of the breadth of topics he treated. I attended this course along with only about half a dozen others, and by the end of the semester there were but three or four of us left. Mandelbrot's way of lecturing was extremely intuitive and his manner intimidating. I found him a thrilling intellectual presence, and I especially prized his capacity for conceptual cross-connection—which drove a few of my fellow auditors right out the door. I came to know Mandelbrot better over the course of the semester, and we had several talks in his office. As with Dyson, I had the sense of being in the presence of a major intellectual, but my experience with Mandelbrot went in a very different direction. In addition to their differences in intellectual strength, Dyson and Mandelbrot were almost antithetical in personality. Though both were willing to speculate, Dyson was accurate and self-effacing, while Mandelbrot was powerfully suggestive and self-aggrandizing. Mandelbrot was, in short, dramatic. In one of his lectures, Mandelbrot told us that what we were experiencing was like attending the master class of an operatic diva. For my part, I would have divas no other way.

Like Mandelbrot, Max studies the stock market, but whereas Mandelbrot was looking for statistical patterns in price changes, Max uses his computer to predict individual changes from day to day. The first is mathematics, the second is mathematical fiction. When Max gets a computer output that indicates a stock that's never fallen below 40 is going to dive to 6½, followed by his computer's crashing, he tosses the computer printout into the garbage. Big mistake: the next day he sees in the paper that the tip was right on the nose. Worse yet, before

the computer crashes, it spits out a 216-digit number. Max's mentor Sol, it turns out, has found 216-digit "bugs" associated with computer programs, but his own trip down the rabbit hole of number theory has ended with his suffering a stroke, and he's retired to tending pet goldfish and playing Go (the latter a Japanese board game). Sol steers Max off course from the 216-number pursuit for a while, but Max is persistent. When pressed, Sol finally coughs up his conjecture that the number is something the computer spits out as it momentarily attains consciousness, just before crashing. Still, he thinks, the number is just a random string of digits. Max isn't convinced; besides that, he's pulled his mainframe apart and it's sprouting some kind of goo that looks like primitive brain tissue. Maybe even the ants that were walking around in there had something to do with it. There are three registers of blinding light that get superposed here, in terms of mythology: Max's blistering migraines, the origins of consciousness, and his neighbors in sexual congress in the adjacent apartment. It's bound to end badly.

Two groups are after Max: the rabbis mentioned earlier, who float a goofy-looking low-ranking piece of bait named Lenny (who nonetheless knows a lot more than he lets on), and a Wall Street cartel fronted by Marcy, who lures Max with a super-secret, still-classified computer microchip. Max barely escapes these latter buzzards before being swooped off in a station wagon full of rabbis to go meet the grand master, who attempts to extract the number from Max. I won't spoil any more of the movie with plot summary.

I claim that the movie π provides us with a mythology of the parafinite. The 216-number is not small, but it's not infinite, either—it is in between, just in the way good mythology should be. (By the way, 216-digit numbers can be very difficult to factor.) It stretches the bounds of our world picture symbolically. Like God, who is still personal, only a whole lot bigger than we are in every sense, the 216-number seems connected to us in a way the Ω-number never could be. And yet equally, it is mysterious, perched at the boundary of human understanding—as is our hero Max, the migraine-inflicted Jewish Chinatown resident recluse. (Marcy gets his address from Columbia University, where Max has presumably studied or worked.[2]) It's interesting that both the economic and the religious interests are after Max, and though he doesn't identify with

either, he clearly favors the rabbis. The only people he trusts are his mentor Sol, a young Asian girl who poses him elementary math problems, and perhaps the woman next door. Sol lies to Max to try to lead him off track, and when Max defies him and goes ahead with his research, it's bad news for them both.

In the movie π, mathematics is viewed in Pythagorean fashion, not just as fundamental knowledge but also as the fundamental stuff of being: *all is number*. So, the 216-number is not just the blueprint of the universe but also something like its underlying armature. Economics and religion—the two central sources of social power—have sold us out, and we're left with knowledge, community, and possibly love. And all three are built into (or equivalently, out of) number. All those parafinite faces we look out at, and that pass by our gaze during the movie, swirl about us in patterns of arithmetic complexity, just as the parafinite stars of the Milky Way swirl mathematically about a pivot (probably a supermassive black hole). The parafinite is adapted to each of these dimensional vortices; it is the blueprint of our spiraling outward. And the underlying message is that we are spinning out. We are feeling massively outpaced. Well, I do, at least. How about you? From the perspective of π, what is the Matrix? It is a spiraling web, a vortex of numbers.

So, why resurrect set-theory as mythology? That's not quite what Chaitin has done. He's built uncritically on set-theory to fashion an approach to computation that has (for him) certain mythological implications. But the question why we should resurrect set-theory as mythology, nonetheless, basically stands.

The first answer is this: we should approach mythology inclusively. We should look at Gödel and Turing and Chaitin and Mandelbrot and π and the *Matrix* trilogy. Mythology has a history with a much longer scale than either mathematics or philosophy as they are typically construed, and the rate at which mythology changes is concurrently slower (Witzel 2012). As Hans Blumenberg puts it, mythology responds to longstanding assertion-needs, and these change more slowly than more specific mathematical or philosophical assertion-needs.

The second answer is this: the set-theoretic orientation continues to be the dominant foundational orientation in mathematics, and so long as it satisfies this community's foundational desires (which are themselves

a mixture of rational and irrational impulses), the mythology stemming from this orientation will give us valuable information about the currently "representative" picture of mathematical mythology. Strangely enough, the mythologies of π and the *Matrix* movies, which are further removed from the foundational desires of mathematicians and much closer to the mythological impulses of the community at large, are much closer to alternative orientations like the one Nelson (1992, 2011) or Isles and Yessenin-Volpin promote. The social unconscious is closer to the vanguard of mathematical dreaming, and the comparison cannot be made in full without understanding the mythologies of both the mathematical mainstream and the vanguard. The vanguard aspect will be pieced out somewhat in remarks later about the *Matrix* trilogy.

The third answer is that some of the points made in purely mythological fashion in a movie like π can be traced back to conservative anticipations in the mathematical mythology coming out of set-theory. Both the similarities and the differences—for example, a master number in both cases, but of infinite precision (uncomputable) versus 216 digits (computable), respectively—are interesting. To some extent, as we move from the mainstream to the vanguard, the similarities with the "social unconsciousness" reflected in the popular medium of the movies increase and the differences dissipate. Why is this? Or, is this really the case? Even the mythologies of two movie ventures as similar as π and the *Matrix* trilogy diverge in many ways. In any case, neither of the movie mythologies moves beyond a static ontology of numbers, sets, and algorithms to the more sophisticated, vanguard commitment to proof and activity. Whatever dynamism there is in the mythologies has to do with static entities —numbers or number-like symbols—streaming down or spiraling around. And in both cases these "movements" come from the grinding out of computer programs. None of that moves in any way beyond Chaitin's orientation. So, to look at these mythologies, we need some sense both of the mainstream tendencies they reflect at the level of fundamental commitments and how these commitments get moved around in ways that make no literal sense from the mainstream perspective—but are closer in some metaphorical way to making sense of the vanguard ones. There are whole hosts of important issues to be investigated here, but I pass them by in rapid descent. The next step is to look at the philosophical methodology appropriate to paraphysics.

2 The History of Bracketing, from Paraphysics via Husserl's Phenomenological Bracketing and Kant's B-Preface to Descartes's Meditations

The prehistory of this methodological trajectory lies in the transition from Descartes's cultivation of algebraic geometry, in which the issue of geometric exactness is paramount, to the methodological doubt presented in his *Meditations*, where Descartes is concerned with what we might call metaphysical exactness (Bos 2001; Klein 1992). With regret, I omit a discussion of this transition and begin at what is often taken to stand at the fountainhead of modern philosophy, yet sets us squarely *in medias res*. As you probably noticed, the title of this section gives the reverse trajectory; we'll get to this shortly.

Descartes's *Meditations* is maybe the most famous text of the modern philosophical period and the bane of many a university student. A friend of a friend of mine was once asked by his philosophy professor, who seemed to him crazy, "But, Antonio, how do you know I'm standing here in front of you?" His response, with the telling rebuttal, was "You lookin' at me, ain't cha?" I find this pointedly brilliant, and have spent many more hours thinking about it than Antonio would have ever guessed. But for all that, the Cartesian method of doubt is not so pointless (though perhaps it is quite so insane) as Antonio must have felt. Indeed, it is important to feel how radically preposterous Descartes's methodological doubt would have seemed to a Classical Greek philosopher. This was not anything like the Classical skepticism that a philosopher of antiquity might have encountered, which was always aimed one way or another at morals regarding how we should live in this world. Cartesian skepticism, rather, is in the service of a method for how we can change the world, and as such it is radically modern in flavor. But this forces upon us the question of how radically casting doubt on all our knowledge can be put in the service of an orientation that favors a dramatic assertion of our capacity to change not just our this-worldly condition but, indeed, the conditions of the world in which we live?

The answer is that the Cartesian methodology of doubt is "hygienic"; it is a slate-cleaning so we can begin afresh. But to extend the metaphor, it is a bit like scrubbing off one's skin to begin dermatologically anew. Or better yet, scrubbing off our nerves and muscles, too, until we arrive at the bone, which we will then also cast into the furnace of hygiene. Driving home this metaphor and its implications of bodily annihilation may help to also drive home the point that Descartes is working with a powerful conception of mind—of the mind as something which, unlike the body, can emerge from its own ashes like the proverbial phoenix. And what should the mind be like so that it would be capable of that resurrection, and why shouldn't the presupposition that mind has this capacity also be cast into the fiery furnace during our extended exercise in mental hygiene? Such troubling questions constitute the complicated casuistry of Descartes's unprecedented exercise.[3]

My favorite book on Descartes's philosophical project remains Jean-Marie Beyssade's *La philosophie première de Descartes* (Beyssade 1979). Among the lessons I carry away from my reading of this book—whether Beyssade would endorse them or not—are that Descartes was in full possession of a new method for treating the continuum of mental processes, that he understood the proximity but also the difference of this continuum of mental processes to and from the continuum of bodily or physical processes in a powerful and sophisticated way; and that this particular form of "continuum philosophy" established a precedent for philosophical orientation that we can trace through the modern tradition at least as far as Bergson in the early twentieth century. And all of this arises, organically as it were, out of the methodology of doubt we find at the beginning of Descartes's *Meditations*, whose most obvious and esteemed progeny is to be found in the explicitly denominated *Cartesian Meditations* of the twentieth-century philosopher Husserl. In a complicated tale that must necessarily pass by way of Leibniz, the bonding of Cartesian methodology to continuum metaphysics becomes constitutive for some of the deepest currents in modern Western philosophy. But here, for now, I focus on the methodological orientation from which this continuum metaphysics flows. And since I almost always think that history reads better backwards, I begin with the incipient instances of paraphysical doubt intimated above and read backwards from them to this post-Cartesian and then Cartesian modern tradition.

In Chap. 3's "quick and dirty argument," we looked at problems concerning how operations like addition, multiplication, and (most crucially) exponentiation "meshed" with the counting numbers, 1, 2, 3, . . . Usually we take these counting numbers to go on forever (whatever that might mean), but the "quick and dirty argument" already suggests that how the counting numbers "go on" might not be a matter as simple as marching toward the horizon. We can see the beginning of the problems we face if we just consider this image of marching off toward the horizon. What would it mean to march off toward the horizon forever? In any realistic sense, when we march off toward a horizon, the horizon itself moves.

But suppose we attempt to respond to this situation by saying this: The point is we are considering the possibility that we could march along forever. But what is *forever*? Suppose, for example, we are bugs on an island, and we are not close to the shore. Suppose, in particular, we are far enough inland that we could march for our entire lives without ever reaching the shore. Isn't that a version of marching along forever? In particular, how could I tell in such a situation whether there was some ultimate end to the island? And why shouldn't it be the same for the counting numbers?

It seems that any attempt to resolve this dilemma would prejudice the very question we are trying to resolve. It would be question-begging in a sense analogous to the way the arguments Isles and Nelson consider exponentiation to be question-begging. This leaves us in a position of not being able to give a positive or a negative answer, and so the obvious next move is to adopt an agnostic position. This, at least, maintains a neutrality between the two possibilities (insofar as we are capable—more on that in a bit) and pushes the problem back, in the case of Nelson and Isles's arguments, from the level of numbers to the level of our numerical notations—that is, to names for numbers. If our commitments to numerical notation presupposed at the level of naming whatever we were attempting to demonstrate at the level of numbers, then we would assert that we had begged the question and thus *beg off* answering the question. If, on the other hand, we could successfully prosecute our agenda at the level of numbers without assuming equally much at the level of names for these numbers, then we would confidently assert that we had, indeed, accomplished something: we had supplied a proof with *soul*—that is, a proof that doesn't fail to go through for the sort of reason that is exemplified in

Edward Nelson's result about exponentiable numbers. As I would put it, we would promote *liberation mathematics*.[4]

This example supplies an instance of what I want to call *paraphysical bracketing*. As it turns out, the name is well-suited to do double duty, which, although it's no proof that one is heading in the right direction, is always pleasant. First, it identifies the chief methodological move—what I have earlier called the adoption of an agnostic position on a potentially question-begging issue—by which paraphysics progresses as an enterprise. But second, and more important, the bracketing is not just characteristic for paraphysics as an enterprise—in fact, it can be seen to be paraphysical in its nature. For what is being bracketed in the example supplied above is precisely a metaphysical issue about the status of the natural numbers and one which paraphysics, in its agnosticism, insists that we put on indefinite hold.

In this stance toward such metaphysical commitments, it is quite helpful to compare the notion of paraphysical bracketing to the notion of phenomenological bracketing developed by Husserl, the latter which is perhaps the most proximate antecedent for a methodology with ambitions at the same level as the one rooted in paraphysical bracketing. For Husserl, the act of phenomenological bracketing consists of "filtering" our experience of all incidental features, so that essential elements of structure can be disclosed to thinking. These essential elements of structure yielded by phenomenological bracketing are not exactly metaphysical in the traditional sense, for they are disclosed descriptively rather than as the result of any normatively rational procedure. As such, we may in fact call them *constitutive* in a way that makes them *pre*-essential. Nonetheless, Husserl uses the vocabulary of essentiality to describe them.

In his characterization of phenomenological bracketing—or as he also calls it "epochē," from the Greek ἐποχή—Husserl speaks of what is bracketed in terms of a putting-out-of-play what he calls the "natural attitude." The natural attitude is the attitude in which I stand when I simply and naïvely provide a description of facts at hand: I see my table, I see my computer sitting on the table, my hands are typing on the keyboard, I am seeing all this through my glasses, I am listening to a symphony by Haydn, it is 1:40 in the morning, and so on. This would be a descriptive account not in the pure language of phenomenology for

which Wittgenstein was looking in his earlier philosophy, but in our ordinary, everyday language—the language Wittgenstein ultimately decided must already be in logical order itself. Husserl's bracketing constitutes not only a "reduction" of this ordinary language description but also one more focally of the orientation-in-consciousness that such an ordinary language description reflects, and that Husserl calls "standing in the natural attitude." In the phenomenological epochē, "[*i*]*nstead of remaining in this attitude*, *we propose to alter it radically.*" In the natural attitude, I am committed to all sorts of acts of positing, as Husserl calls them: in asserting the existence of the table, I posit in judgment the table as something standing over against me. In the phenomenological attitude that results from bracketing, I do not cancel these posits, but in a radical sense, I put them out of play by *changing their valence*. As Husserl describes it:

> The annulment in question is not a transmutation of positing into counter positing, of position into negation; it is also not a transmutation into uncertain presumption, deeming possible, undecidedness, into a doubt (in any sense whatever of the word): nor indeed is anything like that within the sphere of our free choice. *Rather it is something wholly peculiar. We do not give up the positing we effected*, *we do not in any respect alter our conviction* which remains in itself as it is as long as we do not introduce new judgment-motives: precisely this is what we do not do. Nevertheless the positing undergoes a modification: while it in itself remains what it is, *we*, *so to speak*, "*put it out of action*," *we* "*exclude it*," *we* "*parenthesize it.*" It is still there, like the parenthesized in the parentheses, like the excluded outside the context of inclusion. We can also say: The positing is a mental process, *but we make* "*no use*" *of it*, and this is not understood, naturally, as implying that we are deprived of it (as it would if we said of someone who was not conscious that he made no use of a positing); rather, in the case of this expression and all parallel expressions it is a matter of indicative designations of a definite, *specifically peculiar mode of consciousness* which is added to the original positing simpliciter (whether this is or not an actional and even predicative *positing* of existence) and, likewise in a specifically peculiar manner, changes its value. *This changing of value is a matter in which we are perfectly free*, *and it stands over against all cogitative position-takings* coordinate with the positing and incompatible with the positing in the unity of the "simultaneous," as well as over against all position-takings in the proper sense of the term. (Husserl 1982, I, 54–55, using marginal numbers)

This is one of the most important passages in the history of philosophy for considering the pre-history of paraphysics, and in particular the notion of paraphysical bracketing. It is a passage that reveals its secrets only slowly, and I will not pretend to "gloss" it here. Instead, I only point out some features of Husserl's phenomenological bracketing that are typically underemphasized and of particular importance for the comparison with paraphysical bracketing. First in line is that Husserl insists that when we engage in phenomenological bracketing it is not a matter of taking something away from our conscious experience but, rather, of *adding something on*. This something that is added on effects a change of "valence" in our consciousness, and specifically with respect to position-takings. We do not abandon these position-takings, but in a certain sense we neutralize them of their assertoric force: not by depriving them of this force but, rather, by training our consciousness away from the direct encounter with this force. We could say that we "overcome" this force of direction, or even more pointedly and literally, that we force this force of direction into indirection.

The analogy with putting matter into parentheses is helpful (up to a point!): we do not literally remove the material, but its parenthetization qualifies it in a way that requires us to approach it indirectly. The analogy (like all analogies) is only valid up to a point, since I do not necessarily neutralize the assertoric force of an assertion by putting it into parentheses. (What I do is more subtle than just that.) Another analogy that is helpful (although Husserl does not himself use it) is the activity of mentioning a piece of language rather than using it. I see a table in front of me: here, I am using the word "table" to refer to something I see in front of me. But in the second half of the previous sentence I have mentioned the word "table" to refer to the use of the word "table" in first part of the previous sentence: "I see a table in front of me" (which I am now mentioning rather than using). Mentioning something is a sort of using it, but a *special* sort. It does not deprive the word of its referential function, but it neutralizes this function when we mention a word rather than using it. This analogy, too, has its limits. In the end, what Husserl is attempting to develop in terms of the idea of phenomenological bracketing is *sui generis* and cannot be reduced to linguistic activities of parenthesizing or quoting.

What distinguishes paraphysical bracketing from phenomenological bracketing? First, at least in the example of paraphysical bracketing we've seen, the philosophical context for paraphysical bracketing is not descriptive but, rather, argumentative. We are in the process of trying to conduct an argument, and we cannot settle on the appropriate premise: whether the natural numbers are only internally unending, or whether they are in some sense externally unending as well. To predispose the decision would be to beg the question in the context at issue, which is the sort of relativization engendered by Löwenheim-Skolem (L-S) type results. So, in particular, the context for paraphysical bracketing is a *located* one: we are not in the business of the "slate-cleaning" that Husserl's phenomenological bracketing implies in its attempt to neutralize the entire natural attitude; this sense of phenomenological investigation must be given up along with Wittgenstein's. But what is the nature of this "location"? We are not speaking of a particular physical location, exactly, though historical locatedness in a particular context-of-thinking is relevant. Most specifically, it seems, the locatedness must be specified in terms of a particular rational context of argumentation.

Let me insist immediately that I am precisely not endorsing a reduction of philosophy to the trading of argument-stances. On the contrary, the point is rather that in neutralizing the position-taking stance in phenomenological bracketing, Husserl commits to the idea that the force of position-taking can be put (forced) radically out of play. As Leibniz would no doubt insist, that idea, like the ontological proof for the existence of God, is one whose possibility implies its actuality, but its possibility remains to be proved. And what could possibly serve as such a proof, given that Husserl has put the position-taking stance out of order through the purported act of phenomenological bracketing? We are backed into another version of self-presupposing foundationalism. Paraphysical locatedness commits to our apparent incapacity to escape this sort of circularity in the way that Husserl claims to be endorsing phenomenological bracketing.

Ultimately, this is a question of philosophical intuition, I suppose: if you try to engage in phenomenological bracketing and feel yourself successful, more power to you. Paraphysical bracketing does not require such a feat of entrance or entrance fee, since it is predicated not on a force of

neutralization but, rather, on an explicitly agnostic recognition: I do not find myself in a position either to confirm or disconfirm, for example, the possibility of phenomenological bracketing. This recognition is radically distinct from, and in some sense weaker than, the recognition that would be involved in the capacity either to endorse or to refute the possibility of phenomenological bracketing. What can we make of such a weak, agnostic recognition?

Well, we can make a different sort of argument, which is indeed much weaker but which I find quite interesting (and if the phenomenological reduction doesn't succeed, then we would need to ask: weaker than *what*, anyway?). This sort of argument has the form: it seems to me that if you want to prove the totality of the exponentiation operation the onus is on *you* because I can show that the most obvious way to try to go about it is question-begging. (Isles makes this point explicitly; Nelson's argument suggests a similar response.[5]) Not only that, but then we can go on to ask the larger question: What is implied by its being the case that this argument, which looks not just good but even trivial from a traditional perspective, in fact begs the question? This opens up the philosophical land of paraphysics.

In one way, at least, the strategy of paraphysical bracketing is closer to the way that Kant describes the enterprise of the *Critique of Pure Reason* in the preface to the second edition of this work. This preface, usually referred to as the "B-preface" (since the second edition is referred to as the "B-edition"), has troubled me ever since I first read it over thirty years ago. What most perplexed me about Kant's description of his venture in the B-preface is that he frames it in unapologetically hypothetical terms. Let us, he says, provisionally adopt a distinction between the world of appearances and the world of things in themselves. If we see that this distinction allows us to resolve a whole host of traditional problems with which the metaphysical tradition besets us, then we will conclude it successfully confirmed. This sounds like nothing so much as adopting a scientific hypothesis, then going out and running experiments to confirm or refute the hypothesis. (Rummaging through Chaitin's books, I happened on a passage in which he quoted a de facto proponent of the empirical or quasi-empirical nature of mathematics, who said: "mathematics is just like physics only the experiments are cheaper." According to another

rather nasty joke, experiments are even cheaper in philosophy because philosophers don't need to buy trash cans. In particular, it's not clear that Kant ever threw anything away.)

Is there some better—or at least more interesting—way to understand what Kant is proposing? I think there is. To reframe Kant's strategy, it helps to keep two things especially in mind. The first was that Kant said the goal of his philosophy was to *limit reason to make room for faith*. The other thing to remember is that the enterprise of the First Critique specifically, the *Critique of Pure Reason*, was to find a way to resolve the history of metaphysical dilemmas which Kant found pressing upon himself. These dilemmas generally took the form of two antithetical claims and arguments for both. Kant lists, in particular, four sets of antinomies, of which the first is the pair of assertions: (Thesis) The world has a beginning in time and is also enclosed within bounds as regards space. (Antithesis) The world has no beginning and no bounds in space, but is infinite as regards both time and space (Kant 1996, B 454).

To present the story a bit retrospectively, what Kant discovered when he looked at the arguments for these two assertions was that both arguments were sound once the premises were granted. Could we, then, favor one or the other argument by arguing that the premises supporting one or the other were more defensible? No, Kant thought, we can't. So, what should we do? What Kant proposed was to identify a *common ground* for the adoption of both sets of assumptions. His guiding principle was something like the following: When we find opposing philosophical camps adopting different assumptions leading to antithetical results, look for the problem not in the specific assumptions but in the *rational drive* that has led these philosophers (on both sides) to make these arguments in the first place**.**

It is here that Kant's project of limiting reason to make room for faith comes in. Kant felt that the mistake these philosophers had committed was more fundamental than adopting false premises. The fundamental mistake was to appeal to reason in an *unbounded* way, to seek *final* philosophical explanations where, in fact, the resources of human reason were unable to support definitive answers to these questions. The fact that equally defensible, but irreconcilable philosophical explanations had historically been offered was, for Kant, evidence that reason had overstepped

its own limits of defensibility. What needed to be done, then, was to find an appropriate way to limit reason that would not, on the one hand, artificially constrain our ability to engage reason in those contexts where its use was legitimate, while on the other hand, to delimit that region in which reason has a natural drive toward overstepping the legitimate domain of its application. Kant did not believe that we could ever prevent human beings from overstepping these bounds, since it is part and parcel of the rational drive to desire such overstepping; but he did think we could rationally protect ourselves against it—much in the way we can attempt to moderate any of our desires rationally while recognizing that we will never fully succeed in this venture of self-moderation. This does, and should not, change the fact that the problem here is fundamentally one about what we want, and so a matter of *desire*. Our capacity to say what we *should* do, and our capacity to find a systematic way to support these norms, is what is at issue here, not the extent to which we will be able to put those norms into practice. The latter is important, of course, but philosophically (though not practically) speaking, the former is a precondition of the latter.

Kant's invocation of the conceptual distinction between appearances or experience, on the one hand, and things in themselves, on the other, is the master *conceptual* distinction that allows him to prosecute his agenda. In particular, it allows him to disambiguate the use of terms that are being equivocated by opponents in the battle of the antinomies. Seen in this way, Kant is not himself enlisting a metaphysical distinction between some "world of appearances" and another, distinct "world of things-in-themselves"—though Kant does often use the phrase "two worlds," and though this appears in fact to be essential once he gets down to the business of practical philosophy, in which what matters is a "kingdom of ends" that is of the world of things in themselves. This larger outlook on Kant's philosophy is not my interest here, however.

So far as Kant's strategy as described in the B-preface stands in productive analogy to paraphysical bracketing, the point is that his distinction between appearances and things in themselves is no more (and no less) than the distinction we've made between the "bugs-on-an-island" picture and some "external picture." And Kant's "provisionality" in invoking this distinction in the B-preface is analogous to the agnosticism involved

in the paraphysical bracketing of the distinction between the "internal" and "external" perspectives. In the latter, paraphysical case, we do not attempt to dissolve the distinction; like Husserl, we neutralize it for the purposes of our investigation. Unlike Husserl and Kant, the paraphysical neutralization does not involve either a forceful change of valence or a critical legislating but, rather, an agnostic backing off. Kant is certainly not recommending that we adopt the distinction between appearances and things in themselves in a neutral way: the positive analogy between Kant and paraphysical bracketing comes instead from recognizing that, in both cases, the employment of the conceptual distinction is part and parcel of unraveling a false dichotomy. The major difference, and it is major indeed, is that in the paraphysical case we make the distinction (or better, allow the distinction to be made) while recognizing the possibility that it is incoherent, but we bracket it for purposes of argumentation, and then we show how it was the underlying distinction that was incoherent in the first place. This has, and can have, no *legislative* force in the way Kant intends, and in this regard, the ambitions of paraphysics are weaker than Kant's.

It is this agnostic backing off that leads us to the projection of a concept of the parafinite as a way of undermining the original distinction between the finite and the infinite (or the "internal" and the "external") as a clear-cut dichotomy. Because Kant relies on a definite distinction between appearance and thing in itself, he in fact sets himself up for a metaphysical rebound, a "return of the repressed." This is the unwonted metaphysical consequence of his legislative intent. And yet in more informal moments, Kant himself anticipates the innovation of the parafinite. When he speaks of the theses and antinomies of the respective antinomies, he often phrases them in tellingly paraphysical terms:

> *First*, assume *that the world has no beginning*. In that case the world is *too large* for your concept. For this concept, which consists in a successive regression, can never attain the entire bygone eternity. Suppose *that the world has a beginning*. In that case the world is in turn *too small* for your concept of understanding in its necessary empirical regression. (Kant 1996, B 515)

The world is *too large* if it has no beginning and *too small* if it has one; this is because the concept of successive temporal regression is inadequate both to the idea that this regression could be bounded and that it could be unbounded. In a way, paraphysics only attempts to turn Kant's metaphorical language into something more resilient and to support it in a more methodologically direct fashion, avoiding the metaphysical return of the repressed that riddles Kant's work in the form of the "two worlds" hypothesis— a world of phenomena, on the one hand, and of noumenal things in themselves, on the other. In comparison with Husserl, Kant's program seems more philosophically "realistic" in the sense that it does not require us to commit to the assumption of some mysterious "other" attitude in which "nothing changes but everything changes" to engage in philosophical work.

Paraphysics would seek to emulate this Kantian reality-orientation. Indeed, paraphysical bracketing is the methodological application in difficult philosophical contexts of something we all do (or at least should do) all the time when we engage in controversial conversations: we work to identify what we can't agree on, articulate it as well as we can, and then try to move forward. This, at any rate, seems to me a much more realistic attitude toward dispute than one that focuses on "resolving our conflicts and (only) then moving on."[6] There are always too many conflicts, and the conflicts are too deep, for us to expect to resolve them before we move on. This is because, as Kant already recognized in the limiting context he considers, the conflicts are a matter of desire. The consequence is that when we preach such resolutions, we are consistently pushing some ideological program, generally fueled by our own needs and desires. Generally speaking, what we need to do is find formats for discussion in which we are able to float our conflicts rather than letting them dam up and/or explode. Let's get real: the idea of "conflict resolution" is almost always a euphemism for "conflict management." Conflict resolution is the wonderful exception that proves the rule. And when we do agree, we should really ask whether this agreement is the product of social conformity. But how do we figure that out?

This brings us back to what is most idealized in the methodology of doubt that Descartes enacts in his *Meditations*, and what has made him the proverbial whipping boy for all attacks on philosophical "foundationalism."

Archimedes said, "Give me one fixed point and I will move the world," but the whole point is that there is no such external fulcrum for the world-as-totality. Did Descartes recognize this irony when he spoke of finding an "Archimedean point"? Did he in any way acknowledge that any Archimedean point would be *located*, and therefore less than a total, absolute vantage? While we should be less ready to saddle Descartes with such thoughtlessness, Descartes was insistent on the process of the *Meditations* as radical, going to the root, and establishing certain foundations. Can we extract the radicalism of Descartes's project from the metaphysical foundationalism that has historically caused so much trouble? To a considerable extent, Husserl's *Cartesian Meditations* seeks to do so, but by invoking its own commitment to the absolute transition from the natural to the phenomenological attitude, so that Husserl's advance, while real, is only incremental.

As I have suggested above, what is most to be retained from the Cartesian venture is the radical *modernism* of his enterprise. This is an aspect of the Cartesian heritage that paraphysics seeks to affirm. Descartes's methodology is in the service of an *activation* of knowledge, and for all the emphasis that has been laid on the metaphysically static implications of the Cartesian conception of (dual and non-interacting) substance, once we look at the location of Descartes's metaphysical enterprise within his larger philosophical program, we see that ideas are in the service of the modification of our world-condition and even our world. As Hans Blumenberg might put it, Descartes is, if not without reservation or internal tension, promoting a philosophy of human self-assertion. It is easy—all too easy, in fact—to castigate this philosophy of human self-assertion by pointing to its potentially disastrous consequences, ranging from nuclear annihilation to global warming (welcome to the desert of the real). But though this problem is real and not to be ignored, what must go hand in hand with this recognition is that tautologically, if we are going to *fix* the situation, then *we* are going to fix the situation. The reparation of damages associated with human self-assertion must come through the human capacity for self-assertion, even—and especially—if this human self-assertion takes the form of limiting the domain of permissible self-assertion. For, again tautologically, as soon as we have the capacity to change the world for better or worse, choosing not to do so is as much or more a form of self-assertion as is choosing to do so.

Only were we to abandon the worldview in terms of which humans are *normatively* endowed with the systematic capacity to change their world-condition—and this is what is specifically modern and radical about the post-Cartesian worldview—would the issue of self-assertion recede. This would mean, for starters, giving up the scientific enterprise as we know it.

None of this seems to me either likely or desirable, despite the natural fact that our dissatisfaction with the modern condition is expressed, and will continue to be expressed, in these terms. The dystopic elements of the *Matrix* Mythology are an excellent example. And yet, as Charles Rozier has insisted, there is another side of the *Matrix* Mythology. This is the side that has to do with the progressive interfusing of the domains of human and machine, so that in fact as the *Matrix* trilogy progresses, there comes to be a blurring of the boundary between these two realms. Charles pointed out that the Merovingian's wife asks for a kiss. The Indian family Neo meets in the train station is kind but also nervous, facing its potential annihilation, and most especially, the danger to the daughter, in a distinctively human fashion. And at the end of the trilogy, it is in fact this nonhuman daughter we see alongside the oracle as symbol—an indication or instance of future generations. The distinctions between symbol, indication, and instance blur. The horizon we face, exiting from the desert of the real, is beautiful, but seems unreal (or perhaps hyperreal). Will we see Neo again, the daughter asks, and the oracle says: "I expect so." But this would require Neo's resurrection. For the time being, the Matrix has "worked out" its anomalous contradiction, a task that was so dangerous because it was dangerously uncertain. It involved huge gambles, and the gains, we are told, will not be permanent. None of this guarantees Neo's return, but all of it is surely a precondition for it.

Notes

1. The idea of an impossible transition from syllable to blood is inspired by an unpublished poem by J. Patrick Fadely, "Fantasy Baseball."
2. Compare Preston 1992. Thanks to Isadora Mosch for chasing down this reference (and for a lot of other help besides).

3. It has become somewhat fashionable to trace everything in Descartes back to Augustine. Even if every move in Descartes's exercise can be traced back in some way or another to Augustine or some other antecedent source, the radicalism I identify cannot possibly be traced back. Descartes does not possess it alone, but he does share it with an essentially modern cohort. If you want to look for roots of this radical spirit of thinking, seek them in Montaigne (or Shakespeare), not Augustine.
4. David Isles makes this point in terms of the idea that we have liberated a small piece of mathematics from the hidden recesses of what was packed into our bound variables. As Isles puts it, we would *free the bound variable*. This, and the idea of liberation mathematics, are discussed extensively in the Appendix.
5. The argument form is presented by David Isles and discussed at length in the Appendix.
6. Compare "ruling out the possibility of contradiction before we prove any mathematical theorems" or "getting our philosophical foundations established before we make any philosophical claims."

6

An Introduction to Philosophical Praxis (Level of Philosophical Praxis: Level 4)

1 We Interrupt Our Program to Bring You This Special Report

All of a sudden there was a disastrous wrenching. It was as if something had been sacrificed. I awoke in my reclining chair and looked around. The room seemed the same, but it wasn't clear if it was night or day. It seemed to be one of those in-between times. I walked downstairs, opened the front door, and walked outside. The sky was strange, covered with clouds but dimly glowing. You were standing there looking up, too. "It could be any time of day," I said. You agreed.

Had I walked out the front door of my house into the world, or had I exited a movie theater, or had I crossed the threshold into the movie theater of the world? "Is this the Matrix," I asked you. But you did not reply. I held out my hand toward yours; you held out your hand toward mine. The hands looked almost like mirror reflections of each other. As our hands approached, they grew increasingly bright, and as they came close to touching, the light increased and pushed them back apart once again. We were unable to make contact, you and I. If this is a dream, then what am I dreaming? If I am awake, then what am I living? What is the Matrix?

O.B. Bassler, *Diagnosing Contemporary Philosophy with the Matrix Movies*, DOI 10.1057/978-1-137-57889-1_6

I stumbled when the wrenching occurred. It felt like the Manifest had almost jumped the track, but then, suddenly, everything seemed back on track again. What had happened? Had we passed into a new reality? And yet everything seemed just like the old one, which left no account of how the wrenching could have come in between. When I stumbled, you faltered. But unlike mine, your faltering wasn't a physical stumbling. You weren't anywhere to be seen, not until I came outside and noticed you looking up at the sky. But how did I get off the train?

If life is a dream, then who is dreaming it? I am not dreaming it, I am living it. I am in the dream that is my life. If I were dreaming it, then how could I be living it? But what if I am dreaming I am living it? Then, the strange color of the sky, the fact that it could be any time of day or night, might alert me that I am dreaming, not living. But if I am living it, and the color of the sky is strange and it could be any time of day or night, what then? To live with this, you have to give something up. Until you give something up, space comes apart, and time comes apart, and they can't fuse together into motion. There are only anticipations of motion in the violent wrenching of the Manifest, during which time falls apart and then rights itself again. But a time filled with wrenchings is no time at all, only a series of moments with no real connections to each other. The moment after the wrenching, when the Manifest rights itself and runs again along the tracks, both is and isn't the moment before the wrenching. In fact, there's no clear before and no clear after. And if you looked out the window of the Manifest, you wouldn't see countryside moving by. What would you see? Only a roster, with symbols streaming down.

How do you know the Truth? In this world, if you can call it a world—there are not yet forces of Evil and forces of Good. This world, if you can call it that, is still too primitive. There is only the wrenching and the righting, and this comes before the forces of Good and the forces of Evil. The wrenching, the righting, and the writing of the roster. It is as unreal as Mr. Rogers asking, "Will you be my neighbor?"

So, we cannot even say that some time has passed, that is how impoverished we are in this district. When I say that something is practical, you take it to mean that it is only a different and anomalous kind of theory. And I have to admit that you are right. But when you try to explain what it would really take to be practical, the language is full of strange-sounding

words all mixed up together. There are even words missing, illegible words, right when you get to the most important points in the Manifest. And I think, *that's just how I feel, too*. You give reality names like "the practical event-of-Being of acknowledgment-affirmation." I can tell I need to brush up on my late nineteenth-century German philosophy. But it's funny, and I keep reading, and then the Manifest seems almost like a dream. After all, it comes from another time, in another part of the world, in a language I can't read in the original; it's so remote it's hard to imagine as continuous with anything that's going on just now. If you can even talk about something going on just now. If you can even think of it. This is the hole we've fallen down into. The question is not how will we get out but, rather, what the hell do we do now?

That's just where we find Neo.

2 Action!

But, in fact, for Neo the action has stopped, he has finally hit rock-bottom. "He says, 'I seem to have stopped falling; now I *am fallen*, consequently, I lie here in Hell'" (Bloom 1973, 21). Hell, at least, turns out to be someplace real: it has *solidity*. And in Hell, there are Others, the Helpful Ones. Now, Neo can recompose himself, or have himself recomposed. Now, there are forces of Good and forces of Evil. Now is the time for Action. But first, Neo must be instructed by Morpheus. In order to be really real, Neo must apprentice himself to the power of dreams. Morpheus's power is the power of belief.

The Matrix is an action movie. But what is an action movie—what is this genre? This is not such a simple question to answer as it might at first seem. What is the action in an action movie? There are, to be sure, all sorts of motion—grand fights and chases, swordplay and shootouts. But action? Whose action? Who is active in the action movie? The answer, of course, had better be Neo. But Neo finds the power of resistance, the power of sacrifice, and the power of assimilation, in that order. Neo finds the power to take a bullet, get up, and walk away. Neo finds the power to die and die again. Where is the action in all this? Is this the self-assertion of not engaging in self-assertion? As the motion in an action movie becomes more stylized, it moves away from action and toward

the *quotation of action*. It is as if the action movie is talking the language of action without engaging in action—talking the language of violence without feeling truly violent. Is there violence in an action movie? What about all those dead people? What about all those guns?

Let's take a particular example. In his training under Morpheus, inside the Matrix Training Program, Neo tries to jump from one rooftop to another. No one has ever made it on the first jump—no one *believes* that much—and Neo doesn't make it, either. When he emerges from his training, back in base camp, there's blood around the edges of his mouth. *If you feel violence as real in the* Matrix, *it is real*. And yet if this were true to the full extent of the law, Neo would be dead—he just jumped off a skyscraper. Is this because the Training Program is more forgiving? But clearly we're not supposed to ask such a question. So what is violence? Violence is something like the *impossible transition from dreaming to waking*. Violence is the trace of blood you feel in your mouth on waking. If we could pinpoint this wrenching transition from dreaming to waking, we would have the master key for unlocking the meaning of action. This mystery is solved by the *deus ex machina* of the red pill, the results of which suck Neo down the tube. Now, at Level 4, he has hit the floor. He may be knocked out, but he's cold-dead awake. In fact, this is the first of Neo's resuscitations. He keeps hitting bottom, and he keeps getting up. This is a repetition compulsion of the wrenching transition from dreaming to waking. In the limit, it would happen with every breath, which is the smallest transition from living to dying and back again.[1]

In fact, actually, nothing moves on the movie screen: it's all an illusion. So, if anyone is acting, it isn't Neo, or Keanu Reeves (the actor, who has already *acted*) but, rather, you. Are you acting? Maybe, actually, you are the hero of the action film. Neo is only a quotation. But if you are the hero, then you aren't a hero in your life, but in the dream-of-the-movie. But you aren't *in* the movie, you are *watching* the movie; you are watching a simulation, just as you are reading this Manifest.

In his book *A New Theory for American Poetry*: *Democracy*, *the Environment*, *and the Future of the Imagination*, the literary theorist and critic Angus Fletcher develops the conception of a new genre of poem, a genre he calls the "environment-poem" (Fletcher 2004). I doubt that he

had action movies much on his mind when he was thinking about this, but what he says is that in the environment-poem, the reader becomes the hero of the poem. In the most traditional form of poetry, which is the epic poem, the hero of the poem is, well, the epic hero—Achilles, for example. And as Mikhail Bakhtin stresses, the presentation of the epic as temporally remote is designed precisely to frustrate our direct identification with the epic hero. At the other end of the historical tradition, so to speak, by the time we get to the environment-poem there's no ostensible hero in the poem at all.[2] All that's left is the "position" of the hero, which has been vacated for the reader, who is thus invited to supply the missing presence of the otherwise absent hero.

In the action movie, there certainly seems to be a hero: our hero is Neo. And, short of delusional fantasy, we do not presumably identify ourselves as "the chosen one." (But what about what we might call "movie-solipsism"?) In this sense, the action film seems more like the epic. And *The Matrix* is set in some future disjoint from our current reality. We can imagine this future as a dystopic extension of our current reality, but hardly literally. Giant swooping machines harvesting human warmth as a source of energy? Please. There is clearly a sort of distancing function built into the setup. So, in what sense could you possibly be the hero of *The Matrix*—or, for that matter, any other action movie?

Here's the answer: the action movie allows you to reinvent yourself. That's why the name "Neo"—new—is so perfect! In fact, let's go ahead and say you renew (re-Neo) yourself, at the hands of Morpheus, the creative power of sleep, no less. This is starting to look like allegory, but allegory of a special sort (Fletcher 2006). For what is required is that you, as reader, insert yourself into the allegory. It's like the way Chaitin's Ω is allegorical: it becomes allegorical when you plug yourself into it as the source of mathematical creativity—making new, re-Neω-ing. And with the parallel to Ω in mind, it gets even better: The oracle is the internalization within *The Matrix* of the principle of allegory itself. And what is the oracle? A computer program that balances out the Architect. And what is Neo? Neo is the quotation of anomaly. He represents how you stick out when you jack into *The Matrix*. You are the ultimate flash drive. When you plug into the movie, are you plugging yourself into the computer, or into the program? It's both—the

latter by way of the former. So, what is the action in an action movie? The act of plugging yourself in and getting a jolt. How active is that?

How active is the descent into a dream? And how active is the return from dreaming to waking? Perhaps this liminal stage is "meta-active" (van Gennep 1960). "It is not so much in the dream state, as in the preceding delirious stage, particularly when one has been immersed in music that a relationship is established between colors, sounds and perfumes" (Hofmann, quoted in Lockspeiser 1973, 69).[3]

The Matrix tells a story, and the story makes a mythology. In the *Matrix* Mythology, there is a clear-cut distinction between the Matrix and the Real (until Neo discovers that his powers in the Matrix extend to the Real, which suggests that the Matrix has levels). The distinction between the Matrix and the Real is the distinction between the blue pill and the red pill. And throughout the first movie, there are plot points geared to the question: Should I have taken the red pill? And would I go back if I could, even at the price of betrayal? And how do I cope with a choice I made in the absence of a knowledge of its consequences—as, in more or less dramatic ways, all our choices are? But though this armature supports the mythology, the getting-going of the *Matrix* Mythology machine isn't what is most interesting about this mythology. The mythology only really gets going when the strict dichotomy breaks down, when there are passages back and forth between the Matrix and Reality, when the dichotomy between the Matrix and Reality begins, by implication, to break down, when border zones appear (like the Train Station), when the dichotomy of Reality and the Matrix takes a backseat to the dichotomy of the Human and the Machine, which in turn breaks down when computer programs start infiltrating the domain of the Real, and on and on,– just like natural numbers.

If it feels like action, it is action. And yet, if this were true to the full extent of the law, you would be dead. We have jacked into an action movie training program. How much do you believe? Have you hit bottom? Are you in Hell? Have you gotten up and walked again? Are you ready to move on from the training program to some real action? *The Matrix* (movie) is not the Matrix.

What is the Matrix? *The Matrix is the locus of Real Action.*

3 A Novel Philosophy of the Act

In the period 1919–21 the young Mikhail Bakhtin (b. 1895) worked on a manuscript, part of which was posthumously published under the title *Toward a Philosophy of the Act* (Bakhtin 1993). Bakhtin was living in Nevel and Vitebsk in the wake of the Russian Revolution, among such distinguished artistic figures as Chagall, Malevich, and El Lissitzky (Clark and Holquist 1984). The manuscript as it has come down to us is incomplete and in a degraded state, so some of the passages that remain are impossible to decipher. This situation is at least better than the one we find ourselves in with respect to Bakhtin's later book on the eighteenth-century German novel, which was lost when the publishing house where the manuscript was placed was bombed during the German invasion of the Soviet Union, and Bakhtin used much of the only copy he had retained as a source of paper for rolling cigarettes.[4] In fact, Bakhtin seems never to have held his manuscripts in high regard, and the works published posthumously had to be coaxed by friends from a rat-infested shed in which Bakhtin had stashed them (Bakhtin 1981, xxiii–iv).

What remains of *Toward a Philosophy of the Act* is a precious indication of the origins of Bakhtin's thought and, in particular, the larger ethical context in which he viewed his research in aesthetics. Bakhtin was well versed in both the Russian and European philosophical traditions, and his knowledge extended to such then-vanguard figures as Bergson, Husserl, and Scheler. In particular, Bakhtin takes as a point of departure the distinction between formal ethics and nonformal content ethics of values that one finds in the work of Scheler. Scheler offers his own nonformal content ethics as an alternative to the formal ethics of Kant, which he indicts as guilty of a false "theorizing" of the ethical. Roughly, this is to say that if we attempt to derive ethical value from the mere form of ethical judgment, then our derivation will necessarily restrict our notion of ethical value to the rational nature that we can find embodied in the pure form of ethical judgment per se; but then, according to Scheler, the notion of ethical value must remain as empty as the form from which it is ostensibly derived. Instead, Scheler insists that ethical value must derive from content, not form, and he supplies a theory of this notion

of ethical content. Bakhtin, however, going Scheler one better, convicts Scheler's own content ethics of the very same fallacy as Scheler identifies in Kant's formal ethics, only at a different level. This is because, according to Bakhtin, the replacement of form by content changes nothing about the fact that both Kant and Scheler are attempting to develop theories of ethical value. But this, Bakhtin says, is only to promote another form of theory, which in Scheler's case is content-based and therefore ideological, when in fact what is needed is, rather, an account of the practical sources of value. In other words, and as Bakhtin puts it, a theory of the practical is only one more version of theory. And so a "theory of ethical value" is, for Bakhtin, a *contradiction in terms*. It is only ideology promoted under the banner of theory.

Where, then, are we to look for an account of the practical sources of value? Bakhtin's answer to this question is provisional, but it does shed important light on the future trajectory of his work, and though the answer is not fully satisfying, it is quite prospective:

> In order to give a preliminary idea of the possibility of such a concrete, value-governed architectonic, we shall analyze here the world of aesthetic seeing—the world of art. In its concreteness and its permeatedness with an emotion-volitional tone, this world is closer than any of the abstract cultural worlds (taken in isolation) to the unitary and unique world of the performed act. An analysis of this world should help us to come closer to an understanding of the architectonic structure of the actual world-as-event. (Bakhtin 1993, 61)

Bakhtin acknowledges that his recommendation to look at the "world of aesthetic seeing" is only a first step, but what is more disappointing is that he does not give any direct indication why or specifically how this analysis "should help us." Yet, if we look at the trajectory of Bakhtin's overall career, we find that his focus within the "world of aesthetic seeing" came to be on the still-nascent genre of the novel, and looking at his various works on the poetics of the novel can help us to see what lay behind Bakhtin's original proposal in the passage quoted above.

As a genre, the novel is particularly suited to the analysis of "emotion-volitional tone" because of the way verbal intonation is reflected in

novelistic language. Following his invalidation of theories of value whether formal or content-substantive, Bakhtin emphasizes that in literary language there is no division between form and content, and to this end, he introduces the category of the *chronotope* to reflect the way literature, as a language in which form and content are inextricably fused, can carry the concrete value that is unavailable to any theory that would separate the two.

> In the literary artistic chronotope, spatial and temporal indicators are fused into one carefully thought-out, concrete whole. Time, as it were, thickens, takes on flesh, becomes artistically visible; likewise, space becomes charged and responsive to the movements of time, plot and history. This intersection of axes and fusion of indicators characterizes the artistic chronotope. (Bakhtin 1981, 84)

Philosophically, Bakhtin understands this as a replacement of the Kantian transcendental categories of space and time by the chronotope as a "formally constitutive category" under which space and time are fused together "in the process of concrete artistic cognition (artistic visualization) under conditions obtaining in the genre of the novel" (Bakhtin 1981, 85). Bakhtin compares the notion of chronotope to Einsteinian space-time, and says that while the relation is mostly metaphorical, it is not entirely so (Bakhtin 1981, 84).

In some ways, the position of the author in Bakhtin's conception of the novel anticipates the role the reader will occupy in Fletcher's environment-poem (though in other ways they are rather different). Specifically, the author of the novel enters into a "zone of contact with the world he is depicting" (Bakhtin 1981, 28), and as creating consciousness,[5] lies "at the boundary of language and style" (Bakhtin 1981, 60). The presence of concrete value insinuates itself into the text by way of emotion-volitional tone, which the author as creating consciousness manipulates by presenting a wholly concrete language in which there is a continuous fluctuation of representation and parody, direct expression and indirect quotation, so that the texture of the language itself opens up to the multiplicities of the various worlds that the novel "includes" rather than simply "depicting."

Poetic metaphors, by being pitched in a register other than the author's own, are converted into indirect quotations where intonation is internalized as irony, since the distance between the author and the various worlds he or she indirectly enlists casts the traditional metaphor into the chiaroscuro of the author's only suggested appraisal, and the author sits, metaphorically, at the intersection of all these worlds in collision. "The author (as creator of the novelistic whole) cannot be found at any one of the novel's language levels: he is to be found at the center of organization where all levels intersect" (Bakhtin 1981, 48–49). As distinct from these levels, "the author participates in the novel (he is omnipresent in it) with *almost no direct language of his own*" (Bakhtin 1981, 47). Thus, value is embodied in language interstitially—between registers, between direct description and indirect quotation, in the "queering" of pitch and stress we find when a speaker enters a foreign language. And this interstitial embodiment of value in language reflects the way we find value in the world, distributed among people, groups, and cultures in constant exchange.

Thus, we find value on the ground floor, both in the world and in language. Bakhtin's poetics is *locative*: the creative act is always primary and is doubly located, in the world and in the linguistic chronotope. It is from this doubleness of location that the division into genres emerges. But equally, this location is not to be saddled with a false sense of definition; for each locus is in motion, straddling the competing value-tensions, pitched across registers, irreducibly multicultural and polyglot. Genres emerge as the praxical substitute for definitions in the intersection of loci, and the genre of the novel is of particular power and suppleness in accommodating the sense of locative intersection. Such, at least, is Bakhtin's conception of generic strength in the novel. The generic strength of novelistic language is that it is the *quotation of action*.

Where are we when we are in the Matrix, and what should we do? The mythological answer to this question, the one on which the plot of the *Matrix* movies is predicated, is that when we are in the Matrix, we are in a "world" of illusion and that we should take the red pill and get out, come hell and/or high water. If the action movie is anything like a popular equivalent of the novel or the environment-poem—and suitably qualified I believe it is— then Bakhtin (and Fletcher) suggest(s) that the

power of the *Matrix* movies lies elsewhere. To criticize the *Matrix* movies for perpetuating the simulacrum seems strangely humorless, confusing the *Matrix* movies with their depiction of the Matrix. The real Matrix that the *Matrix* movies create is not the Matrix they depict but, rather, the sense of phantasmagoria they inherit, as does the novel, from the festival tradition. This is as if didactically confirmed in the second *Matrix* movie, when Morpheus stands as Roman orator addressing the crowd at the Bacchanalian festival that precedes the day of battle, girding the troops and civilians alike for the coming onslaught. There are many levels to keep straight here, from the representation of the ritual festival within the movie, to the sense of phantasmagoria that the *Matrix* movies engender as action flicks (playing subtly and not so subtly against the depiction of a dystopic Matrix within the movies themselves), to the larger question of what it means at the level of philosophical praxis to inhabit the Matrix and how this could be exemplified in any sense by our experience of watching an action movie.

All of a sudden, Stephen Mulhall's suggestion that in the tetralogy (teratology?!) of *Alien* movies, "a fundamental part of the philosophical work of these films is best understood as philosophy of film" (Mulhall 2002, 3), may presage an even stronger claim than he may intend: there is something specific not just to the genre of the movie but also to the genre of the action movie and para-action genres (horror, sci-fi, samurai) that is fundamentally relevant to the capacity for film to transact philosophy—that philosophically, "achieving our country" (Rorty 1998) is more deeply rooted in action movies than in any return to literature.[6] From Bakhtin's perspective, this is perhaps even obvious, since there is no aesthetic cultural artifact more central to contemporary American culture. One step further along, and if Abbas Kiarostami is right in his suggestion that the power of Hollywood is even more coercive for world conditions than the American military (Kiarostami 2004), we have a mighty problem on our hands. What the hell do we do now? Perhaps what Cornel West (who, of course, appears in the *Matrix* movies) called "The American Evasion of Philosophy" (West 1989) really is a deep problem of contemporary American culture, and not just a lament to be sung by a handful of academics.

Bakhtin, alas, was writing in code, as anyone who has read about the plight of the creative intellectual in Soviet Russia understands (Mandelstam 1970; Haight 1976; Schweitzer 1992). Ruth Coates relays a conversation between Bocharov and Bakhtin that took place on June 9, 1970, but was only published by Bocharov in 1993:

BAKHTIN:	Everything that was created over the course of this half-century on this graceless soil under this unfree sky, it is all depraved to some degree or other.
BOCHAROV:	Mikhail Mihailovich, leaving [Voloshinov's] book aside for a moment, that's a complicated matter, but what is depraved about your book on Dostoevsky?
BAKHTIN:	Oh come now, could I really have written like that? I tore the form away from the most important thing, you know I couldn't talk directly about the main questions.
BOCHAROV:	Which questions, M. M.?
BAKHTIN:	Philosophical questions, what Dostoevsky tormented himself with all his life: the existence of God. I had to prevaricate all the time, to and fro. I had to take a firm hold of myself. As soon as a thought got going it was necessary to stop it. To and fro (Bakhtin repeated this several times during the conversation). I even qualified what I said about the Church. (quoted in Coates 1998, 9)

Later in the conversation Bakhtin said directly to Bocharov, "You, at least, do not betray. If you don't assert, it's because you're not sure. But I prevaricated—to and fro" (quoted in Coates 1998, 9). This last addition is critical, since it confirms that there were not just sins of omission but of commission as well. What is perhaps most tragic about the case of Bakhtin is that it was to just the sort of theoretical strictures he criticized to which he was subjected while promoting a praxical approach designed as an alternative to the sort of top-down constraint such theoretical programs enforce. Given that Bakhtin speaks of "tearing the form away from the most important thing," his repeated use of the phrase "to and fro" might even be taken to refer to the shuttling back and forth between theory and content that he criticized in *Toward a Philosophy of the Act*. But then, in a way, Bakhtin's work itself becomes a field in which the

praxical ironies of self-betrayal play out soundlessly beneath the parodic linguistic elements in the novelistic texts Bakhtin himself so radically exposes. *The text becomes a praxiological minefield.* The project of philosophical praxis is a labyrinthine one, simultaneously spiraling down a bottomless well: we are beset with a novel form of the dimensional labyrinth for which Piranesi's prisons would only serve as a first approximation. This is reflected, too, in the architecture of the Real World we find in the *Matrix* movies, with Zion deep beneath the surface figuring a mythological core accessible only through a multidimensional system of conduits. It is part of the *Matrix* movies' inverted architectonic that Zion (whose outer wall the machines breach with spiral drills) would occur roughly where we find the deepest pit of Hell in Dante's *Inferno*.

There is no sense in saying that Bakhtin would have been more successful under "better conditions," though his text would surely have been more transparent.[7] In turning from Bakhtin to Albert Camus, we must resist the temptation simply to say that we can find in Camus a report on aspects of the novel Bakhtin "could not talk about." Camus fought his own battles, none of them as overtly dominated by censorship as Bakhtin's, but all of them equally imbricated in webs of ideology. So, when we read the "Rebellion and the Novel" subsection from Camus's *The Rebel,* we cannot compare it directly to Bakhtin's program despite their common (and largely unspoken) reliance on a sense of the sacred (theological in Bakhtin, nontheological in Camus), and more specifically, their common, uncommonly deep indebtedness to Dostoevsky. Perhaps, to the extent that there is a common root, it is to be found in their mutual "existential" emphasis on the concrete over the abstract.

Camus's treatment of the novel is the first of three sections of part 4 of *The Rebel*; the second and third are, respectively, "Rebellion and Style" and "Creation and Revolution." The parallel with Bakhtin—the prominence of the novel, the attention to style, and the emphasis on creation—is suggestive. For Camus, art is the activity that "exalts and denies simultaneously" (Camus 1956, 253), and the artist is the one who, to create beauty, "must simultaneously reject reality and exalt certain of its aspects" (Camus 1956, 258). In the force of this rejection of reality we find the principle of rebellion already contained, and this essentially rebellious radix is only intensified by the literature of rebellion, "which

begins in modern times." Generically, the novel is paradigmatic for this literature, and Camus addresses this first by emphasizing the anomaly involved in "writing or even reading a novel" (Camus 1956, 259). Although Camus's conception of the novel genre is hardly as developed as Bakhtin's, he does penetrate quickly to a point central to Bakhtin as well: that the novel makes primary the simultaneous *inclusion of factual material* and the *falsification of facts*. This means that the novel is porous to the world like no other genre, and yet this porosity itself makes the capacity for dissimulation exponentially more powerful and treacherous. What is potentially pernicious about the novel is figured in terms of our capacity to use it "to escape," and thus, as Camus puts it, "common sense joins hands with revolutionary criticism" (Camus 1956, 259). It is no longer possible even in principle to distinguish the revolutionary from the everyday; the revolutionary potential of the novel coincides with its being a silly genre. A similar point is expressed by Bakhtin:

> And here we encounter the specific danger inherent in the novelistic zone of contact: we ourselves may actually enter the novel (whereas we could never enter an epic or other distanced genre). It follows that we might substitute for our own life an obsessive reading of novels, or dreams based on novelistic models. (Bakhtin 1981, 32)

What is a potential danger for novel reading is only all the more obviously so for watching movies. There is a danger, then, that we may enter *The Matrix*—the movie, that is—and substitute the watching of it for living of our lives. But it is out of this danger that the capacity to bore to the center of our consciousness protends. The experience of watching *The Matrix* is simultaneously a displacement of our creative consciousness and a boring-down into its deepest home. What will happen when it drills through the last protective wall of our outer mind? But this, of course, is a demonizing portrait of the movie-as-obsessive-temptation, and is to be taken no more literally than the representation of the Matrix within *The Matrix* itself.

As Camus insists, we cannot simply interpret the act of reading a novel as an escape from reality, for "happy people read novels, too, and it is an established fact that extreme suffering takes away the taste for reading"

(Camus 1956, 260). The activity of reading a novel embodies the contradiction that "man rejects the world as it is, without accepting the necessity of escaping it" (Camus 1956, 260), and so the activity of reading a novel is precisely exemplary of rebellion, a rebellion that springs from not possessing the world completely enough and that Camus also identifies in the activity of turning other humans into novels-for-our-reading. Only in the novel do we experience this peculiarly thorough sense of satisfaction, of *control*: "What, in fact, is a novel but a universe in which action is endowed with form, where final words are pronounced, where people possess one another completely, and where life assumes the aspect of destiny?" (Camus 1956, 262–3). The novel allows us to outpace the frustrating incompletions of our reality (Camus 1956, 263). The potential for rebellion is a potential for violence, a violent rejection of existence as we find it expressed in Proust's melancholy (recreation of, hence, equivocal) retreat into an earlier happiness. Proust's art is unrepentant in its thoroughgoing "extraction" from the dispersion of the world of "tentative, trembling symbols of human unity," as if the wrenching of existence (subjective and objective genitive) could somehow be transported whole as a unity into the domain of art. Such is art's illusion, its dream, its rebellion. Beyond this lies only the theology of Dostoevsky's dream of truth.

Among Dostoevsky's central themes, Bakhtin identifies the theme of the *crisis-dream* as the one we might identify as "genre-shaping" (Bakhtin 1984, 152): "more precisely, it is the theme of a man's rebirth and renewal through a dream vision, permitting him *to see* 'with his own eyes' the possibility of an entirely different human life on earth" (Bakhtin 1984, 152). It is at this point that Bakhtin cites the crucial passage from Dostoevsky's "The Dream of a Ridiculous Man" that may be taken as a proof-text for the entire project of the Matrix Philosophy, and which I therefore quote here:

> Yes, I dreamed that dream then, my 3rd of November dream! They all tease me now, telling me it was nothing but a dream. But surely it makes no difference whether it was a dream or not since it did reveal the Truth to me. Because if you have come to know it once and see it, you will know it is the Truth and that there neither is nor can be any other, whether you are dreaming or awake. Very well, it was a dream—let it be a dream, but the

> fact remains that this real life which you so extol I was going to snuff out by suicide, whereas my dream, my dream—oh, it revealed to me another life, a great renewed and powerful life! (quoted in Bakhtin 1984, 152)

In the topos of the crisis-dream we have a convergence, at a minimum, of truth, myth, phantasmagoria, and praxis. At the root of this convergence lies visionary design; and what is most novel in Bakhtin's treatment of the novel is the way in which he identifies how this traditional core is transposed into the most contemporary of generic forms. But given the inherently theological nature of Dostoevsky's stance toward visionary design, it is just this topic that we find pushed to the margins of Bakhtin's writing. To reconstruct, we must ask how the convergence of truth, myth, phantasmagoria, and praxis may be understood as a function of visionary design, and how the genre of the novel transposes the impulse of visionary design into a modern context—most specifically the context of Dostoevsky's novels.

Dostoevsky's theology—and paradoxically, it is here above all that the nontheological Camus remains indebted to him—emphasizes that it is here on earth that paradise is to be gained. And yet, more paradoxically still, when the Ridiculous Man preaches paradise on earth, he realizes that this paradise can never be gained (Bakhtin 1984, 150). The second aspect of Dostoevsky's theology to be identified is that each person (or character) can possess only his or her own truth (if in possession of truth at all), and it is the individual possession of this truth that "creates the special sort of loneliness these heroes know" (Bakhtin 1984, 151). This loneliness can in turn engender a sense of irreality—the confrontation with nihilism comes not just from the absence of truth but also more deeply from the only way in which truth can be possessed. This sense of irreality leads to a crisis in which the crisis-dream may occur, and truth is reestablished through the collapse of the distinction between the real and the irreal ("they all tease me now, telling me it was nothing but a dream"). The dream reflects the "possibility of a different life" (Bakhtin 1984, 147). The idea of heaven on earth is developed in this dream context in terms of the spirit of a Golden Age, which itself reflects the tradition of the Saturnalian festival (Bakhtin 1981, 147).

The visionary truth of this Dostoevskian kerygma is entirely concrete—"the subject of a living vision, not of abstract understanding" (Bakhtin 1984, 153)—and this visionary concreteness is the locus underpinning Bakhtin's (and Camus's) insistence on the concreteness of value. The time of this vision is the apocalyptic time of instantaneous transformation stemming from the crisis-dream: here, "the only possible time is *crisis time*, in which a *moment* is equal to years, decades, even to a 'billion years' (as in 'The Dream of a Ridiculous Man')" (Bakhtin 1984, 169–70). At the praxical level, we have carnival and a central act of sacrificial dismemberment into parts (Bakhtin 1984, 161–62); at the level of myth (to which we are rapidly descending), we have the explosion of the linear continuum of time and the replacement of linear time by a Golden Age with which the fullness of the present coincides. Bakhtin quotes:

> And yet it could be done so simply: in a single day, in a **single hour** everything would be settled! One should love others as one loves oneself, that is the main thing, that is all, nothing else, absolutely nothing else is needed, and then one would instantly know how to go about it. (quoted in Bakhtin 1984, 153)

Yet before descending into the confrontation of philosophy and myth, we must confront Bakhtin's commitment to the role played by carnival, which stands, praxically, behind his commitments both to visionary design generally and to the myth of the Golden Age specifically. Carnivalization is predicated on the temporary abolition of hierarchical distinction through dramatic reversal, allowing for "free familiarization, scandals and eccentricities, crownings and decrownings" (Bakhtin 1984, 133). This carnivalization enables the embodiment of ultimate questions, transferring them from the abstract plane of philosophical or religious dogma to a ritual enactment on "the concretely sensuous plane of images and events—which are, in keeping with the spirit of carnival, dynamic, diverse and vivid" (Bakhtin 1984, 134).

Carnivalization thus promotes the articulation of visionary design, transporting it from the apocalyptic crisis of the visionary dream to the extended (multi)linearization that the genre of the novel will promote. Bakhtin describes carnivalization not as an "external and immobile

schema" but, rather, as "an extraordinarily flexible form of artistic visualization, a peculiar sort of heuristic principle making possible the discovery of new and as yet unseen things" (Bakhtin 1984, 166). This mechanism works (and in this regard, paraphysical bracketing picks it up) specifically by the *relativization* of external stabilities, most fundamentally through the reversal of positions within a hierarchy. It is a machine for the promotion of indeterminacy and destabilization that allows us to "penetrate into the deepest layers of man and human relationships" (Bakhtin 1984, 166).

Bakhtin stresses that this mechanism is especially suited to the promotion of developing relation[8] —and, we might add, relations of development. Carnivalization is the master creative principle for concrete embodiment, but to speak of it as a principle is already misleading, since it is not a principle in any theoretical sense. Rather, it is a master praxis for the promotion of *sustained praxical proliferation*. Developing relation is the proliferation of the crisis dream into manifold dimensions of narrative complication, articulated around what Bakhtin calls "tight matrices" of linguistic association.[9] In this way, carnivalization is transferred to the literary domain. This master praxis that Bakhtin calls carnivalization I call *philosophical praxis*. That is, the novel externalizes linguistically the controlled praxical proliferation of the ritual event of carnivalization. In its linguistic, novelistic context, this proliferation may proceed indefinitely. Bakhtin notes that in the seventeenth century, the novel of adventure expanded by a factor of ten to fifteen times its maximum size in antiquity, producing examples ranging from three up to five thousand pages (Bakhtin 1981, 94–96).

What is most striking about these descriptions of carnivalization—in which I have followed Bakhtin quite strictly—is how closely they come to describing the fruitfulness of ritual practice in general. As such, carnivalization is not so much a part of Dostoevsky's, or Bakhtin's, theological commitment as it is what the theology must account for. It is in precisely this sense that carnivalization lies at the praxical root of Bakhtin's philosophical theory of the act. Thus, what I call *philosophical praxis* is the deritualization of carnival ritual through paraphysical bracketing.

My reconstruction of Bakhtin's theology of the novel in Dostoevsky may, of course, be objected to on a whole host of grounds, ranging from the particular charge that it involves the conflation of material

from Dostoevsky's novels and his correspondence (but so does Bakhtin's account of Dostoevsky's themes) to the most general charge—which is undeniable—that it is a reconstruction. Bakhtin's works invite us to engage in such reconstruction, just as much as the novel, and Dostoevsky's novels in particular, invites the challenge of Bakhtin's generic reconstruction of its status. I would only like to point out how concentrated the material from which I have drawn is in Bakhtin's text, spanning a range of only about a dozen pages in his book on Dostoevsky's poetics. I take this concentration to support the claim that there is a convergence of issues in his discussion of Dostoevsky that demands this reconstruction, that in fact Bakhtin's larger project is incomprehensible without it. (These pages are, so to speak, the crisis-dream of Bakhtin's own text.)

I am not concerned with Bakhtin's Christianity (or Dostoevsky's) but, rather, with the theology that is the requisite condition of (and in this sense implied by) his treatment of the novel genre and the broader theory of the act into which it ideally fits. Camus's discussion of the novel is important because it suggests the possibility of a nontheological alternative. But Camus's account also lacks the specificity and depth of Bakhtin's, and so it is the theological route which is taken here. This theological route is, in any case, the most obvious one when we turn to the genre of the action film, where we must grapple with the manifest presence of superhuman powers and where the theological conditions become an order of magnitude more complex still.

Notes

1. On Alban Berg as the master of the smallest breath of transition, see Adorno 1991, especially 47, 65, 68, 85, 92n1, 107.
2. An interesting point of poetic comparison would be Anna Akhmatova's "Poem Without a Hero" (Akhmatova 1989, 103–49). For an extended description, see Reeder 1994, 369–431.
3. The line is from E. T. A. Hofmann's *Kreisleriana*, which inspired Schumann's composition of the same name; and as Lockspeiser notes, it reappeared in various guises in several passages in Baudelaire, one of which

inspired the prelude by Debussy, "Les sons et les parfums tournent dans l'air du soir" (Debussy 1913, 13–15).

4. A fraction of the manuscript did, however, survive, and has been published in translation as "The *Bildungsroman* and Its Significance in the History of Realism (Toward a Historical Typology of the Novel)," in Bakhtin 1986, 10–59; see also xiv–xv.
5. Bakhtin's idea of creating consciousness merits comparison with L. E. J. Brouwer's idea of the creating subject in intuitionistic mathematics. See van Atten (2000), 64–71, esp. 82–84.
6. Claude Lévi-Strauss has asserted that "we are witnesssing the disappearance of the novel itself" (Lévi-Strauss 1978, 54). I am not convinced this is so, but its place in culture does seem to be changing, and its cultural function taken over by the movie.
7. Michael Holquist and Carol Emerson compare the trajectory of Bakhtin's older brother, Nikolai, who landed a job in a language department in England. Mikhail Bakhtin's early conversations with his brother were formative for his development, but the two brothers had quite different personalities, intellectually and otherwise. The moral is that not even Bakhtin's brother provides a relevant point of comparison. See Clark and Holquist 1984, 16–34. On dissimilarities between siblings, see Sulloway 1996.
8. Bakhtin's notion of developing relation merits comparison with Schoenberg's developing variation.
9. Bakhtin develops the notion of matrices at length in his essay "Forms of Time and Chronotope in the Novel," most especially in relation to the work of Rabelais (Bakhtin 1981, 175–79). Bakhtin's reading of Rabelais is developed in magisterial detail in Bakhtin 1968.

7

An Introduction to *Philosophia Perennis* (Level of Work on Myth and Philosophy: Level 5)

1 Saving Zion: The Waking Dream of Life

"The Waking Dream of Life" is a phrase used on the inside cover of Geoffrey Hill's book-length poem *The Orchards of Syon*, to describe Hill's enterprise in this volume. What is most striking about the orchards that Hill describes in his poem (not the one on the book's cover) is that they are covered with torched maples, their flaming boughs bent, weighted down with burden. Why has there been a Fall (why is it fall) in the orchards of Syon? Whence this autumnal air, filled with smoke? What is the Waking Dream of Life?

To address this question we will have to make our final descent in this Manifest, into the realm of myth, and the *Matrix* movies will continue to serve as our guiding examples. I will not attempt to provide the full "theological conditions for the action film" I referred to in Chap. 6. No doubt my use of the term "theological conditions" will cause some surprise and, with even less doubt, consternation and misunderstanding. What I mean by this are the conditions for the distribution of subhuman, human, superhuman, and divine powers we find in these films. Theology, in this sense, is not a matter of articles of faith or matters of existence but, rather, the distribution of powers of and within an envisioning.

O.B. Bassler, *Diagnosing Contemporary Philosophy with the Matrix Movies*, DOI 10.1057/978-1-137-57889-1_7

Bakhtin traces the prehistory of the novel back to sources in Classical times, and ultimately before that to folklore. As a genre, the movie has been around for a little over a hundred years. Until fairly recently, there were still people in the movie industry who had worked in the era of silent films. Manoel de Oliveira, who debuted *The Strange Case of Angelica* (*O Estranho Caso de Angélica*) at the 2010 Cannes Festival, made his first movie in 1931.[1] As genre, the movie, too, has relevant background that can be traced back to folklore, but in another sense, unlike the novel it is a genre *sui generis*. Perhaps in some ways the movie's closest generic line of relation is to grand opera, the medieval mystery play, and, before that, ancient ritual theater. In a context of such novelty, we can hardly expect to say anything definitive about the film genre, nor should we expect the films we've seen so far will necessarily serve as canonical models for centuries to come. Still, there have been plenty of films produced, and at least in that sense we have as rich a storehouse of examples as in many other "established" genres.

If we were to find an analog for the action film in the tradition of the novel, it would not be in the high novel of Bakhtin's "Second Line," as represented by masterpieces from *Gargantua and Pantagruel* and *Don Quixote* via *Tristram Shandy* to *The Brothers Karamazov*. We would do much better by looking for a positive point of comparison in the tradition of the first-line adventure novel, which stretches all the way back to Classical times. A modern example is *The Count of Monte Cristo*. By comparison with second-line masterpieces, this is a relatively simple affair. Yet as Bakhtin also points out, there are some important ways in which the first-line adventure novel is much closer to Dostoevsky than is the family or biographical novel, which finds high exemplification in works by Turgenev and Tolstoy.[2]

My goal, however, is not to spell out the genre of the action movie, but only to brush up against one instance of it, the *Matrix* trilogy, insofar as it bears on the enterprise of Matrix Philosophy. To this end, I focus on the interaction of two sets of concerns in these movies, which go some way toward indicating why the movies are compelling and important. Whenever I think about these movies I always imagine a ten-year-old boy metaphorically getting his head blown off seeing them for the first time, and I think you should keep that image in mind, too. (More recently I've

added to my arsenal of mental images a scene from Jean-Luc Godard's movie *Éloge de l'amour* (Godard 2003), in which two kids who look to be about ten years old come to their neighbor's door with a petition for *The Matrix* to be dubbed into Breton. *The Matrix* is in the shadows throughout Godard's movie.)

First, I talk about dreaming and waking in the *Matrix* movies as a fundamental, underlying contrast in terms through which we can understand the Matrix/Reality distinction. It might seem like this is taking something less fundamental (dreaming/waking) and using it to understand something more fundamental (appearance/reality). In the end, you'll have to decide what you think about that. Second, I talk about choice versus determinism in the *Matrix* movies, or to put it another way, who is in control. Ultimately, what I do is show how the two sets of issues are intimately connected. This connection identifies a crux that is at the heart of Matrix Philosophy as well.

In the context of the *Matrix* movies, waking and choosing are both forms of dislocation, of wrenching. Dreaming and waking come up at the very beginning and at the very end of the trilogy. At the beginning of first movie, *The Matrix* (*TM*), we see several sequences in which some strange things happen to Neo, and then it seems he might have been dreaming. There is one sequence in particular (when the bug gets put in his chest) that we "write off" as a dream (and so does Neo) and then later (when the bug is extracted) we discover that it actually did happen. (But then, retrospectively, we have to ask: Happened in what sense? For the entire sequence transpires within the Matrix—and then of course within the movie.) The very beginning of the trilogy throws Neo, and by implication us, into a confusion of dreaming and waking. Neo even asks his druggy friend whether he's ever had this experience, and the friend replies: "Of course, it's called mescaline." Generically, this confusion is registered by horror-movie cues: storms and darkness, wind, rain, a phantasmagoric club, and dark rundown buildings.

At the end of the *Matrix* trilogy, at the end of *The Matrix Revolutions* (*TMR*), the sentinel machines first suspend their attack, floating in a listless holding pattern, and then they retreat.[3] Morpheus says that he's imagined this moment for so long, and then he turns to Niobe and asks her if it is real. Yes, they are not dreaming, but the calmness after the

fierce battle is surreal, perhaps even hyperreal, as if Morpheus has finally awoken from a nightmarish dream. What he has believed has finally been made *real*. But it is Neo who is dreaming at the beginning, and Morpheus who "awakes" at the end. Remember, too, how already in the first movie, the oracle said that one of the two—Morpheus or Neo—would have to be sacrificed for the other. There is a *dislocation* in the role of the dreamer/awakener. Of course, there is another sense in which Neo is "awakening" throughout the trilogy: he comes to experience a growing sense of his powers and vocation as "the One." But as identifiable events, it is Neo who dreams at the beginning and Morpheus who awakens at the end.

Of course, it's even more complicated than this—much more complicated. To get into this complication, let's take another example of dislocation, a messier one. When Neo stops the sentinels in the tunnel at the end of the second movie, *The Matrix Reloaded* (*TMR-L*), he is overcome by the stress to his system. As the Oracle says at the beginning of *TMR*, he wasn't ready for that, so that he *should* be dead, but apparently he wasn't ready for that, either (both options have been suspended). Neo's first experience of this deeper level of the Matrix—what he previously took to be the Real—is somewhat forgiving, just as the Matrix Training Program was somewhat forgiving. By the time Neo heads for Machine City in *TMR*, this deeper level hits him with everything it's got.

But I'm getting ahead of myself. The point is that in using his powers to stop the sentinels at the end of *TMR-L*, Neo gets knocked out—knocked out cold, lying on a table head to head with his nemesis, Smith cum Banes, who has just descended into the Real. Neo's use of his powers at this level and the descent of Smith roughly coincide. But Neo doesn't get knocked out of the Real back into the Matrix, nor does he get knocked down to some level of the Real yet more fundamental. Instead, he gets *dislocated*, stuck in transit at the Train Station, a state of limbo between the Machine World and the Matrix (in a way, this echoes Neo's limbo state when he first lands in the soup in *TM*). We are not told specifically, but presumably the Machine World as a reality is centered in Machine City. (This is probably another question we're not supposed to ask.) Morpheus and Trinity re-enter the Matrix to bargain with the Merovingian, who then agrees for the Trainman to take them to the Train Station. They take Neo back to the Matrix, where he sees the Oracle again, and then he

passes back to the Real from the Matrix. When he awakens, he's "jacked in" and Neo and Trinity are already standing there, looking down at him.

But how did Neo get from lying flat on the table to being jacked in? Presumably, his friends had to carry "him"—but what "him"? This, too, no doubt is a question we're not supposed to ask, but it's a different sort of question from the other one. The other question—about the Machine World and its relation to Machine City—is part of *Matrix* metaphysics (I use this phrase to refer to the metaphysics implied by the movies). The other question identifies a fundamental dislocation that has to do with the generic structure of the movies, which in fact lies at the heart of the contrasts between dreaming/waking and Matrix/Reality. What they point to is the central role played by jacking in. "Jacking in" is a way of locating in the Matrix, but it's also a way of dislocating from Reality (like entering a movie theater). And then, when Neo gets lost "between" the Matrix and the Real, he has effectively dislocated the distinction between location and dislocation! To get back to the Real, he first has to "re-connect" to the jacked-in state. Recall how I described, in Chap. 6, the action of someone who is watching an action movie: it's like being plugged into the movie, and to the action by way of the movie—the ultimate flash drive. At the very beginning of the trilogy, Choi (Mescaline Man) tells Neo he needs to "unplug," and Neo accepts his invitation to the club where he'll meet Trinity, following the white rabbit. When he talks to the Oracle about choices he can't understand, the Oracle describes these choices as lacking a "connection." The posture of Neo and his friends as they jack in is the posture we take in our favorite reclining easy chairs. It is not the posture of Neo lying flat on the table after he's been knocked out (which happens twice). But somehow we (and not just Neo) get (in the movie) from the one to the other—from the horizontal posture of nighttime sleeping to the simulated sleeping of movie viewing. This is like the impossible transition from sleeping to waking, which we manage to accomplish all the time.

Going into the Matrix is the ultimate movie experience, and so watching *The Matrix* is the ultimate experience of the movie experience—shades of Löwenheim-Skolem. If I can see the relativization of the Matrix in *The Matrix,* then can I relativize my experience of *The Matrix*? To my mind, this is a much more interesting question than asking if we're living in a

computer simulation. Why? Because as Putnam argued, there's no good sense in which we could speak of "stepping out of our experience," but there are all sorts of ways in which our experience could be constrained by artificial commitments to distinctions that look oh-so-definite: the distinction between the finite and the infinite, for example, or that between appearance and reality.[4] And our commitment to such distinctions could be (and surely is) gravely skewing, and so limiting, our capacity to experience our experience as *our* experience—to en-own it in the only way that is philosophically available to us and for which the picture of stepping outside our experience stands as a kind of tempting metaphysical fix. This is not a diagnosis of some specific position in contemporary philosophy but, rather, the indicator for a strategy of ongoing diagnosis. What I've called *philosophical praxis* is the activity of progressively neutralizing these confining distinctions and seeing what happens. This has a diagnostic function, but should not be reduced to this diagnostic function, either.

But the *Matrix* movies are action movies, and so the narrative movement—or perhaps we should just say the movie-ment—is shaped by action, or at least by "playing at action," and the way this playing at action is floated is in terms of the problem of choice. Like the appearance/reality distinction, the choice/determinism distinction comes in a traditional metaphysical flavor, and the movies play with this distinction mercilessly, tantalizing us just as much as they do with the fundamental gimmick (and I do not use the term pejoratively) of the Matrix. At the level of action, the fundamental gimmick is the *program*, and the basic mystery is this: if Neo is a programming anomaly in the process of working itself out, then how can Neo also be human and make meaningful choices?

There is a version of the L-S–style relativization problem here, and it gets set out explicitly in the Train Station, when Neo talks to Sati's father, Rama-Kandra. When Rama-Kandra speaks of love, Neo is surprised, saying it is a human emotion. Rama-Kandra says that love is a word. But love isn't a word: "love" is a word, according to the distinction between using and mentioning that was discussed in Chap. 5. Using and mentioning are both actions, but they're different actions. And "love" is a word indicating a connection; it is the connection that is important for understanding the choice.

Then, Rama-Kandra speaks of karma, and he says that karma ("karma"?), too, is a word. The point is this: when toward the end of *TMR*, Smith asks Neo why he keeps getting up, Neo says simply, "Because I choose to." Not "Because I choose to get up" or "Because I choose to X" or "Because I choose to Y," but simply "Because I choose to." But if I'm not choosing to X or choosing to Y, then what does it mean to say, simply, "I choose to"? You could say, of course, that this means "I choose to Z" for some Z that hasn't been specified (like getting up), but we all know that's pretty lame. Neo's "I choose to" is close to Camus's notion of revolt. Camus might say, "I choose to because it is absurd." But that, in a certain way, is to say that there is no "sensible" Z for which "I choose to Z" is true. Is Rama-Kandra's acceptance of karma the opposite of Neo's act of volitional revolt? After all, Rama-Kandra is from an older stratum of the Machine World and is about to be terminated.

Inverse to the way Neo dreams and Morpheus awakens, or the way in which Neo is knocked out on the table and then wakes up unjacked from his chair is the way the Oracle and then Neo allow themselves to be assimilated to Smith. Neo dives into Smith once in the first movie, and that only explodes Smith, rather than infiltrating him in any thorough sense. As we learn from the *Matrix* sequels, the explosion of Smith is quite temporary, yet on his return as "so many of me," he comes back with an inexpressible difference. He has clearly been affected by Neo's infiltration; he becomes, as the Oracle puts it, Neo's opposite, his negative or polar complement, the equation trying to balance itself out. Neo is the One and Smith is the Many—not coincidentally these are the two principles underlying Plato's esoteric metaphysical dualism (Gaiser 1962).

Even so, something still more radical happens when the Oracle and Neo assimilate to Smith in *TMR*. One way to understand the more fundamental acts of assimilation we find in *TMR* (though this may be a bit of *Matrix* metaphysics) is to say that the Oracle and Neo risk *bringing Smith back to the Source*. Recall that Smith has grown so powerful by this time that he has become a threat even to the Machine World. He is a machine program that has the power to destroy not just all humans but also all machines; he is capable of turning the Machine World (by way of the human world) against itself. When the Oracle and Neo infiltrate Smith, he is not dislocated, but ultimately *dissipated*. The consequence, then, is

much like what we found resulting from Neo's diving act in *TM*, but it has required penetration to a deeper level and it is hoped correspondingly more enduring. (As more penetrating, it could ultimately have even more deeply disastrous consequences in the long haul, but this is beyond the scale of our trilogy, and moreover would violate the generic closure of the action-movie form. As Bakhtin points out, at the end we traditionally return to "normalcy," at least for the time being. But what happens when crisis becomes the normal state of affairs? (Bakhtin 1981, 152)).

It is here, in fact, that we have an idea close to certain strands of Gnosticism. In Gnosticism, the human, who contains a divine spark, is trapped in the prison-world of an evil demiurge. One strategy for breaking out of this prison-world is to identify with it completely; certain Gnostic traditions practice a thorough identification with all forms of worldly debauchery and debasement, with the end of breaking free from the constraints of the prison-house (shades of carnival). If we understand the assimilation of the Oracle and Neo to the regime of Smith in this way, we must notice two things. First, this is not an identification with the Matrix, which Smith also despises, but with Smith, who is the force of evil both in the world of the Matrix and eventually beyond (though temporarily Neo cannot believe this is possible, and in any case what it points to is the ultimate depth of the prison-house). Second, from a plot perspective, the assimilation into Smith is the inverse of taking the red pill, and each functions as a *deus ex machina* for advancing the story. The first effects the transition from appearance to reality, and the second effects the transition from good to evil—and then back again. So really, the strategy as a *gamble* is one of assimilation, but as a *success* is one of infiltration and overturning. That is to say, what I originally described as assimilation, because it turns out to be successful, is reversible, permitting transportation in both directions (from good to evil, and from evil back to good), with the consequence that the prison-house of Smith (subjective and objective genetive) has been shattered. But not the Matrix. All the Architect concedes is that those who wished to be freed shall be freed.

The places where we can identify dislocation are important because they figure the experience you have watching the movie: first you fall asleep and then you wake up; first you are knocked out on a table (remember the ten-year-old first-time viewer who gets his head blown off) and

then you take a trip to the Oracle and finally you get unjacked from your recliner (which we hope doesn't resemble a dentist's chair so much as the jacking-in chairs in the *Matrix* trilogy do). And what happens in between is the *jolt* (here, the dental resonances may be more appropriate!). But the jolt is not just in falling asleep, having a fun dream, and then waking up. The jolt is starting off somewhere and ending up somewhere completely different—not getting how you went from the one place to the other. This is the power of the action movie: while in one sense it is a "perfectly harmless" entertainment, in another sense it is rewriting the most fundamental levels of our experience—perceptually, aesthetically, culturally, even paraphysically! Kiarostami may or may not be right in his suggestions for capacities for evil, or even how dangerous it is (he poses this as a question), but I am convinced he's dead right about the power of the action film as a *carrier of cultural experience*.

Is there anything analogous to be learned from the "inverse" move we find in the movies—the move of assimilation? The Oracle's assimilation into Smith may be even more fundamental, and in any case, it is a precondition for Neo's assimilation. When the Architect meets the Oracle at the end of *TMR*, he says to her, "You played a very risky game." Perhaps it is precisely the power of the action film that we can recognize by thinking of this move of assimilation. Although there are some obvious things to consider, assimilation quickly raises some difficult questions—issues that for any extensive treatment would require provision of the sort of action-movie theology I noted at the beginning of this chapter. The ground for extrapolation is shaky, too, given that it is confined to the third *Matrix* movie; to do justice to these issues would require looking beyond this specific context to a range of action movies, which is something readers can do individually.

There are, however, still some things to be said about the connection between the dream/reality register and the program/choice register. One of the really great works about life as dream is the seventeenth-century Spanish playwright Calderón's drama *La vida es sueño* (*Life is a Dream*) In the introduction to his translation of the play, Edwin Honig makes a simple, but helpful point. For those who are living life-as-a-dream in the pejorative sense, the self-counsel will always be *against action*, because action threatens to wake us up. To choose—in the radical sense, really to

choose—not for or because of X or Y or Z but, rather, "Because I choose" is a vote *for action*; and as such, this is always against the pejorative sense of life-as-a-dream. But let's distinguish between this pejorative sense (call it "dreaming life away" and the positive sense (call it "the waking dream of life"). What is a waking dream if not a contradiction in terms?

Strangely, there is another description for the pejorative sense of life-as-a-dream that seems almost in contradiction to the phrase "dreaming life away." Caldéron uses the term *desengaño*, which Honig translates as "disillusionment." But isn't disillusionment the removal of the condition of illusion, the condition of dream? Our terms are not only slippery; they also seem to be collapsing under the cumulative metaphysical weight of the problem. But what "dreaming life away" and disillusionment have in common is that they both register a refusal of engagement, a refusal of action, a refusal of radical choice.

Paraphysics can help us here. It's easy enough, conceptually, to recognize that the distinction between dreaming and waking is not so entirely definite as it may at first seem; one thinks of daydreaming, for example. But paraphysics supplies us with a philosophical strategy for investigating more radically the indefiniteness of this distinction. And so we begin by bracketing the way we've traditionally taken the distinction to hold, by building up an agnostic defense against our preconceptions of how this distinction between dreaming and waking rules our lives. Watching the *Matrix* movies gives us a jolt, whose charge lies somewhere near this distinction and the associated one concerning action. Paraphysics should not just give us a bigger or longer jolt, it should make the empowerment that comes from confronting this distinction an enterprise that can be sustained and deepened indefinitely: a *move-ment*, serially and collectively, beyond "movie solipsism." Husserl spoke of phenomenology as an infinite, unending task. Paraphysics is an indefinite, open-ended one, too; perhaps we are bugs on an island and one day it will come to an end, and not foreclosing the possibility of its ending is part of its aspiration (or nightmare), its "dream," but it is still open-ended for us *now*. Questioning the distinction between dreaming and waking is not the only avenue into paraphysics, though; we've seen how consideration of the finite/infinite distinction can also make inroads, and not unrelated ones. But the dreaming/waking distinction does lie at the root of the modern philosophical

tradition in Descartes's *Meditations*, and more broadly, it extends to the ancient traditions in both the East and the West (one thinks especially of Zhuang-zi's dream of the butterfly). More focally, I've shown how we can trace a line back from the paraphysical methodology of bracketing to the Cartesian methodology of doubting, for which the distinction between dreaming and waking becomes a primary locus.

Of course, we still have not resolved the potential contradiction of the "waking dream." (This could occupy us indefinitely.) Here, too, Edwin Honig says something especially helpful. His point is that it is precisely the conversion of the life-as-a-dream formula from its negative sense as *desengaño* to its positive sense as waking dream that is the active choice. Camus, no doubt, would simply chalk up the contradiction in the idea of a waking dream to the absurdity of the choice. There is something to this, something that should caution us against attempting to resolve the contradiction prematurely. Another possibility is to insist that in both the positive and the negative sense of life-as-a-dream we are dealing with something radically metaphorical, what Hans Blumenberg calls an "absolute metaphor" (Blumenberg 2010). This, too, is an orientation worth pursuing.

The choice at issue, then, has to do with a change in the valence of this absolute metaphor, but we should not expect the choice to change this valence to be any easier to accomplish than it is to choose absurdly on Camus's view or to engage the change of valence necessary for phenomenological bracketing. Honig goes on to say that, in the case of Caldéron's hero, Segismundo, this choice can only be effected by recognizing that the formula life-as-a-dream refers to his *unborn condition* (but how grounded is Honig's claim in the drama of the play?). This unborn condition is what the Gnostics refer to as the "divine spark within." Further, this active recognition comes by way of a perversion of the natural order, in which Segismundo rebels against his father, the king.

Saving Zion is saving the remnant of God's eternal spark among his chosen ones, who in this non- or postcovenant context are themselves the ones who have chosen, who have been chosen in their having chosen. Postcovenant theologies tend to have an air of Gnostic anarchism about them, with Harold Bloom's Gnosticism serving as a recent example (Bloom 1996, 233–53). Blumenberg already identifies

anarchist tendencies in Paul, whose God promoted a law with which it was impossible to comply and who therefore "makes a new form of constitutional lawlessness necessary" (Blumenberg 1985, 23). Saving Zion is casting off the outer garment of dream that is the prison-world of the Matrix world, the land of imperfect, not-too-complacent complacency, where the rub is just right to keep us jacked in, to keep us jacked in without even knowing we *are* jacked into anything. It would be a mistake to think that movies are potentially dangerous only because they string us along in this way. *Movies are much more powerful and potentially dangerous than that*. The mistake would be to confuse what the movie is ostensibly about, its gimmick—and in a very basic sense, maybe all movies are about this—with what the movie does. What the movie *does* is more powerful than a gimmick, for good or for ill. The movie does something that is unprecedentedly modern. The action movie, in particular, *carries* contemporary American culture. That is the weight of its "burden."[5]

2 The Matrix Is Work on Myth and Philosophy

While there is something very American, very Hollywood, about the *Matrix* movies, we are obviously also circling around a crux that descends into the depths of our folklore tradition. The generic name given to such folk tales by anthropologists is *myth*. To borrow a definition from Blumenberg, myths in this sense are "stories which are distinguished by a high degree of constancy in their narrative core and by an equally pronounced capacity for marginal variation" (Blumenberg 1985, 34).[6] This definition of myth could equally serve for Claude Lévi-Strauss, in his four-volume *Introduction to the Science of Mythology* (Lévi-Strauss 1969, –81).

Blumenberg begins his book *Work on Myth* as follows: "To those who are bored with this success, the mastering of reality may seem a dream that has been dreamed out, or was never worth dreaming" (Blumenberg 1985, 3). In the first *Matrix* movie, Smith tells Morpheus how the Matrix had originally been designed to perfection, eliminating all friction and suffering, but that humans had rebelled against such a simulation, repeatedly arriving at the conclusion that it was all *only* a dream; consequently,

"many were lost." As a result, this perfect simulation was replaced by a simulation of humanity "at its height," and it is to just this sense of height that Blumenberg's idea of mastery refers. The dream of mastery can be an exhausting and ultimately a boring dream, and the venture of improving our control over reality can come to seem a venture in reality's gaining a new form of control over us. When he first meets Morpheus, Neo tells him he doesn't like the idea of fate, because he doesn't like to think that he's not in control of his own actions—that something else is doing the controlling. But as Blumenberg points out, traditional ideas like fate, before becoming the demonized bugbears of our ultra-voluntarism, were in fact an integral part of our repertoire for dealing with a reality that was manifestly—and much more directly—not under our control.

The mythical gods were once agents for the mediation of a sense of chaos that the world presented. The gods began close to us, then distanced themselves so that, as Kafka put it, "human lungs could have air" (quoted in Blumenberg 1985, 3). We find such a distancing in Epicurus's withdrawal of the gods to the interzones between worlds; why, then, does Epicurus retain them? It is, as Blumenberg puts it, because "it would be better to accept the myth about the gods than to become a slave of [*sic*] the necessity of the physicists" (Blumenberg 1985, 13). By the time we have Neo and the modern age, the myth of these distant gods has been replaced with a myth about the *freedom to choose*, to escape the absolutism of necessity without the interposition of divine forces. We have moved radically from a condition under which our actions needed to be attributed a significance they did not possess, to a situation in which the most significant acts have no discernible impact, but for an antithetical reason: our acts have been so inundated with the myth of personal significance that they are drowned in a cumulative sea of significance that can only be referred to as some anonymous "other." In the passage from the personal to the anonymous, "significance" is converted into "information." The necessity we would combat, the "control" at issue, is no longer so much physical or even social as it is conspiratorial: *as long as the Matrix exists, humans will not be free.* The dream of mastery has become the dream of the Matrix. But this leaves us in a paradoxical position: In a condition of having mastered reality, what action retains significance—knowledge, community, and possibly love? If so, then specifically what sort of actions do these entail?

For those who have not attained the mastery of reality, Blumenberg asserts, the dream of mastering reality remains paramount, and though he does not say so explicitly, we may understand the reception of First World culture by Third World culture in these terms. Indeed, Blumenberg's insistence on this dream of mastering reality, and that these youthful cultures would "snatch its realization away from those who think they have already awakened from it," advertises a First World fear—call it "outsourcing"—that may be all too grounded in reality. Cultures on different sides of the mastering-reality divide will necessarily receive a movie like *The Matrix* on different terms. On the far side of this mastery of reality, the movie will be about our postapocalyptic sense of needing to reaffirm the human capacity for choice; on the near side, it will be about wanting to be like Neo, the new American hero, or maybe even just about being Keanu Reeves. This is the Gnostic version of our very own culture wars: while "we" are plugging in to break out, "they" are plugging in to break in. Movies, like weapons, can be used for opposing, even antithetical, ends.

If Blumenberg is right, and the dream of mastering reality is the master dream of our reality, and we stand on either the waking or the dreaming side, then the fundamental link between dreaming and choosing is immediate. *The quintessentially modern dream is the dream of human self-assertion,* just as Bakhtin identifies the crisis-dream as central to the genre of the modern novel. But how did we get to such an advanced (perhaps chronically advanced) stage of dreaming and waking? To address this question, Blumenberg turns to philosophical anthropology, drawing conclusions from what he calls "the common core of all currently respected theories on the subject of anthropogenesis" (Blumenberg 1985, 4). On this basis, he posits the limit-concept of superior power (*Übermächtigkeit,* "overpowering"), and asserts that archaic man "interpreted this circumstance of superior power of what is (in each case) other by assuming the existence of superior powers" (Blumenberg 1985, 4). The concept of superior power becomes a way of handling superior power. ("Love" is a word; "karma" is a word.) To interpose the concept of superior power between oneself and superior power is to mediate this superior power by an act of naming.

This disjunction between superior power and a concept of superior power is repeated in the description of dreams, as well. Dreaming is at once "pure impotence with respect to the content of the dream" and at the

same time "pure dominion of wishes, which makes waking up the epitome of disappointment" (Blumenberg 1985, 10). To place the emphasis on awakening is, then, to defy the domain in which dreams retain their function as *agent*! But this defiance can be pointed in two directions: inward, so to speak, as a remythicization of dreaming, or outward, in the direction of surpassing the dream agency by real agency. (The philosophical Matrix is the locus of Real Action.) It is this latter sense of agency that the dream of mastering reality invokes. But how can we distinguish between awakening as remythicization of dream-agency and awakening as the dream of mastering reality *outside our dreams*? Everything we've done so far suggests that we must not take this as presenting separate options and that, rather, we should engage in paraphysical bracketing with respect to these two seeming options. This would push us even farther down the rabbit hole.

Instead of that, for now let's think about how this compares to two different ways of viewing the *Matrix* movies. First, the "lazy" way to take in the movies is to plug in to them in a purely impotent way—zone out for two to six hours and experience the "pure dominion of wishes" and the consequent disappointment (or even relief) when the movies are over. But, in fact, how purely can we do this, actually? Aren't there limits on the extent to which even I can find the passive reception of these movies an experience of the "pure dominion of wishes"? For in fact, these movies are constantly reminding me (as any good nightmare will do) how claustrophobic I find the reality from which I've withdrawn, not just because I've temporarily escaped from it but also because an image of that reality has been internalized in the movie (or dream, for that matter). Notoriously, Freud insisted on something close to the idea that all dreams be understood as wish fulfillments, but this obviously downgrades the internalization of that experience in a dream that makes the fulfilment of our wish possible.

In the limiting case at the other end of the spectrum, then, the active way of taking the movie rejects the movie experience in terms of a false literalism about reality—you wouldn't really waste your time on that movie, would you, if you were really serious about engaging in radical existential choice? But at this end of the spectrum, we are relying on a hard-and-fast distinction between reality and dreaming that is only applied in an inverse fashion. Now, instead of failing to recognize the internalization of a projection of our reality inside the dream, we fail to recognize an internalization

of dreaming inside our reality. Although it doesn't (yet?) have much of interest to say about it, even our current neuroscientific image of human experience wouldn't deny that dreaming is something humans do.

But the way dreaming factors into the *Matrix* movies, and into action movies more generally, cannot be understood solely in terms of the quintessential dream of the mastery of reality. In fact, there is a privileged Classical stratum, very roughly contemporaneous to Epicurus's withdrawal of the gods to interstellar space. As I've mentioned, Blumenberg emphasizes the root power of naming as a way of rendering the indefiniteness of reality at least conceptually definite; archeologically, naming also possesses the tremendous virtue of allowing us to reconstruct much of this process by utilizing the relative historical stability in the employment and transmission of names. As the gods become more remote, there is a consolidation of the divine into transcendental power, and the power of naming then becomes concentrated in the name or names of the divine. We met this mythology of divine names in regard to the movie π, discussed in Chap. 5.

> In the classical writings of the Jewish Cabala, the statement is continually repeated that "the entire Tora is nothing but the great Name of God." But these names are not only appellations, but also designations of the various ways in which God operates and is active. When he speaks, he acts, as the account of the Creation shows, and since he is not a demiurge his action consists exclusively in naming the effects that he wants to achieve. For the Cabala, again, that means that "the language of God has, in fact, no grammar. It is composed entirely of names." (Gerschom Scholem, quoted in Blumenberg 1985, 37)

In Plato's cosmology, the language of names spoken by the demiurge "here already—and momentously—has that of numbers and geometrical figures superimposed on it" (Blumenberg 1985, 37). This language of pure symbols is reflected in the "code" of the Matrix; as Cipher (the reader of script, and our bad guy) explains, the image translators are built into the simulation, and there's too much information to run them "on-screen." Cipher no longer sees the code, only "blonde, brunette, redhead," and the transparency of the code for him may reflect his bare anxiety in the face of reality, with no meaningful escape from the dreadful "desert of the real." As a general matter, the convergence of mythology on a purely

combinatoric play of names seems to be correlated with the removal of the source of divinity into an absolute transcendence. There is a reason why, at the end of the film π, reality becomes unbearable for Max. To experience the meaning of divine names would be to stand in absolute contact with divine transcendence, which is "the madness of the day," staring straight into the sun (Blanchot 1981). After Max points the drill at his head and we see a tremendous splatter, the movie ends with an unreal image of Max sitting on a park bench, enjoying "little knowledge," community, and (possibly) love. The image escapes trivializing sentimentality only to the extent that it feels somehow dreamy, unreal, or hyperreal.

The bracketing of our limiting stances toward reality and dream is the paraphysical entry into work on myth and philosophy. On the one hand, we must neither reduce our images to nominal constructs—"love is just a word"—nor reduce them to conceptual realities, a domain of pure but meaningful symbols, whether linguistic or mathematical. On the other hand, we must not reduce our reality to a dream play of images. Strindberg's introduction to his *A Dream Play* approaches this limit when he says: "Time and space do not exist; on a slight groundwork of reality, imagination spins and weaves new patterns made up of memories, experiences, unfettered fancies, absurdities and improvisations" (Strindberg 1955, 193).[7] (Strindberg is speaking, of course, of a play—or is he?). And yet, at the same time this bracketing must open the opposing extremes to the potential for radical interpenetration. In terms of work on myth and philosophy, this means we must neither exclude the internalization of myth from the philosophical domain nor exclude the internalization of philosophy from the mythical domain. As Blumenberg insists, myth is always a "piece of high-carat 'work of logos'" (Blumenberg 1985, 12), and philosophy is ineliminably riddled with absolute metaphors inherited from mythological tradition. But Blumenberg speaks only of work on myth, not of work on myth and philosophy, which suggests that his insistence on the ineliminability of absolute metaphor from the philosophical domain is not yet sufficient for the paraphysical investigation of the relation between philosophy and myth. Early in his career, Blumenberg admitted that his focus on metaphor is a pragmatic one, yet a thorough pursuit of his project would require that he look at the transmission history of myth internal to the philosophical enterprise, and he certainly does so in major ways both in his *Work on Myth* and in the magisterial, late *Höhlenausgänge* (*Exits from the Cave*) (Blumenberg 1989).

Why, then, does Blumenberg speak of work on myth and not work on myth and philosophy? He takes unequivocal exception to Nestle's project *From Myth to Logos*, citing it as an instance of neo-Kantianism's historical back-reading, which Whiggishly identifies Plato as the founder of a theoretical tradition of which Kant and subsequent neo-Kantianism stand as the culmination, and therefore marginalizing the "art myths" that constitute an integral part of Plato's corpus. Nonetheless, Blumenberg seems to focus on the capacity for a philosophical working-over of myth to the exclusion of a mythical working-over of philosophy. Yet it is clear that this latter is *precisely what happens in the* Matrix *movies*. At the very beginning of *TM*, Neo pulls out a fake book used to hide things; it is a copy of Baudrillard's *Simulacra and Simulation*, and the inside compartment is opened to the final chapter, "On Nihilism." The incessant name-dropping (see if you can spot Schopenhauer's *Die Welt als Wille und Vorstellung*) has become a collision-point for the defenders and denigrators of these movies. In any case, the failure to recognize mythological work on philosophy can only have spiteful consequences, and so I hope I have managed to balance philosophical work on *Matrix* mythology with consideration of the mythologization of philosophical work in the *Matrix* movies. This is perhaps not so far in spirit from Blumenberg as it may seem, and if I have any criticism to offer, it is that *Work on Myth and Philosophy* would have made a better (though less catchy) title for his book.

Most intriguingly, Blumenberg identifies myth as a strategy for a kind of *affective bracketing*, but this is a bracketing together rather than a bracketing out:

> Even when it is still a matter of being on one's guard for the invisible and evading it by observing its rules, affect is the inclusive bracket that unites partial actions that work against the absolutism of reality. Intentionality—the coordination of parts into a whole, of qualities into an object, of things into a world—may be the "cooled-off" aggregate condition of such early accomplishments of consciousness, accomplishments that had led the way out of the bracketing together of stimulus and response and that were at the same time the outcome of this exodus. To that extent there is something in the classical idea that emotion is the unclarity of the mind in the process of feeling its way forward. This schema of accomplishment is filled

> not only by sensation and perception but also by the names, figures, and stories, the rituals and machinations that are bracketed together by the one still-undefined affective condition of overwhelming power that Rudolf Otto entitles the "numinous." (Blumenberg 1985, 21)

Here, Blumenberg's philosophical anthropology meets the theological conditions for myth as story, in the way that individual stories (along with sensations, perceptions, names, and the like) articulate a prior bracketing of stories that collectively "represent" the indefinite experience of the numinous as "overwhelming power." Each story is a chiseling out of story-matter from the collective repository. In relation to this idea of bracketing we would, in fact, need to understand Husserl's epochē as a "bracketing apart," a dispersing of the collectivization of affective contents that allow us to discern the fundamental intentional structures that remain invariant with respect to these contingent affective bases. And yet Blumenberg suggests instead that we think of intentionality itself, the "directedness" of consciousness to objects and aims, in terms of a phase-transition whereby these "early" affective accomplishments are sedimented into less affectively laden orienting structures. This casts doubt on the extent to which we can expect a "bracketing apart," a setting aside, of these affective contents in the course of seeking to identify underlying intentional invariants. Nonetheless, whatever invariance we could disclose in these intentional structures would be a genetic function of the underlying affective stabilizations, and we may wonder if Blumenberg's proposal constitutes in this regard something like a radicalization of Husserl's late program of genetic logic (Husserl 1973).

Howsoever this may be, Blumenberg's key insight is, as he states immediately before the paragraph quoted, that "attention, which is the difference between perception and observation, is stabilized most of all by affect" (Blumenberg 1985, 21). Even if emotion is classically understood as the "unclarity of the mind feeling its way forward," we must not expect that this underlying affective mechanism is eliminable, only that it is refinable. The first way to ask about this refinability in a contemporary context would be in terms of the modern dream of mastering reality.

Paraphysical bracketing proposes a methodology for the management of this refinement of affective stabilization. No matter how "pure" the concepts are that I consider—take the distinction between the finite

and the infinite—they are always surrounded by a penumbra of affective association. First, there is no reason to believe that this affective association will be held in common among even the closest group of philosophical partners in dialogue, and there is much reason to assume it will not be. Second, there is no reason to believe that this affective association can be rendered largely explicit, and there is much reason to assume that it cannot. The levels of historical, personal, sociological, cultural, religious, and political association are labyrinthine, and while I am only counseling realism in identifying them, not despair, Blumenberg's own "metaphorologies" demonstrate just how challenging even the most rigidly demarcated of such investigations can—indeed, must—be. This context is tailormade to the recommendation that, instead of resolving conflicts in our respective orientations, we develop a methodology for *floating* them. Indeed, much of the democratic political theory of the public domain may be seen as a highly circumscribed, mostly formal attempt to do just that. A more thoroughgoing attempt in the spirit of paraphysical philosophy would probe the neutralization of our conceptual distinctions, with consequences for the refinement of our control over affective stabilization.

Just as Blumenberg identifies names as an important source of transmission for studying myth reception and transformation, so the identification of persistent antinomies in the history of philosophy can serve as a starting point for a paraphysical approach to work on myth and philosophy (Kolakowski 1988, 1990). This insight, in fact, lies at the heart of Kant's philosophical revolution, and it is only in the wake of neo-Kantianism that Kant's "analytic" has come to overshadow the influence of his "dialectic." Kant's major innovation was precisely to identify a way to float the antinomies he considers in the dialectic rather than deciding in favor of one side or the other, and although this enterprise does depend on his distinction between appearance and reality, it is committed to this distinction methodologically rather than metaphysically. The metaphysical consequences of this adoption tend to press themselves on Kant much more, both in the analytic of the First Critique (which culminates in the tortuous "refutation of idealism") and in the practical philosophy of the Second Critique. Kant's approach to the antinomies is a prototype for work on myth and philosophy. He does not attempt to dissolve the mythological background behind these metaphysical speculations, only to dislodge reason itself from

the entanglement it inherits from the variety of affective stabilizations for these contested claims. Furthermore, the methodological distinction between appearance and reality is just the right one for distinguishing these affective commitments along just the right conceptual lines.

Viewing matters along these lines, we see that the affective stabilization of the conceptual distinction between appearance and reality lies in the *experiences of dreaming and waking.* Just as Wittgenstein asked the question whether a language not shared between at least two speakers was possible, so we could ask whether the conceptual distinction between appearance and reality would be possible without an experience of dreaming and waking. I am not so interested in securing the answer in either case, for such answers are only to be had in affectively invested metaphysics, but the exercise is canonical.

***

As I finish drafting this volume, my dog sleeps across from me, dreaming.

***

I've dealt a good deal with Father Sleep and the new-day Son, but the last word belongs to Trinity. In the first movie, when Neo meets her at the rave, she pops the question, "What is the Matrix?" Neo is there because he knows the question, just as she did, and because the question is driving him mad. And then she tells him that he will find the answer, or rather, that the answer will find *him*, if he wants it to. At any time of day, in the impossible transition from sleeping to waking or from waking to sleeping, it is a matter of what you want. It is an astonishing matter, a matter of desire. In *Kreisleriana*—Hofmann's or Schumann's—it is a matter of delirium. What is it for you? What is it for me? What is it for us?

Notes

1. The date 1931 is given twice in Pym 2010, 789 (where the title of this film, *Douro*, *Faina Fluvial*, is given) and the date 1929 is given once, 551; Buscombe 2003 dates Oliveira's first feature film, *Aniki-Bóbó,* from 1942. I leave it for the interested reader to sort all this out.
2. "The adventure plot cannot therefore be the ultimate binding force in Dostoevsky's world, but as a plot it offers favorable material for the realization of Dostoevsky's artistic designs" (Bakhtin 1984, 105). Hence, we

have a triangle linking the action movie, the adventure novel, and Dostoevsky's polyphonic dialogism. What links the adventure plot to this dialogism is that it "uses any stable social localization not as a finalizing real-life form but precisely as a 'position'" (Bakhtin 1984, 104). It is just such "positions" that will grow into the dialogical "worlds" of Dostoevsky's characters.

3. This moment broadly recapitulates, but with the roles of Neo and Morpheus inverted, the moment at the end of the first movie when the sentinels are stilled by Morpheus firing the EMP. It also inverts the moment at the end of the second movie, when Morpheus's ship is bombed and he declares that he has lost his dream.
4. This point, already implied by his paper on brains-in-vats, is taken up more explicitly and at length in his paper on the Löwenheim-Skolem Theorem, discussed in the Appendix.
5. My remarks aren't intended as cultural or political analysis, but they are intended as pro-cultural and pro-political. That is, I view them as establishing broad conditions for addressing the cultural and political dimensions of the action movie in general, and the *Matrix* trilogy in particular. What these conditions make possible is generally foreclosed in narrowly cultural and political analyses by virtue of the ways in which they are typically vested. I'm unapologetically taking the long view; Godard's *Éloge de l'amour*, mentioned above, is a movie that also tries to take on these issues. I hope it is clear as well that I am not forgetting there are action movies made in many countries other than the United States.
6. For a recent, book-length treatment of Blumenberg's *Work on Myth* in English, the reader might consult Nicholls 2015.
7. These lines are taken up by Bergman at the conclusion of the film *Fanny and Alexander* (Bergman 1982). In the screenplay, however, these lines are not given.

From the Director's Cut (Level of Transit: Level 1.5)

Löwenheim-Skolem

> In some way, it really seems that the Skolem paradox underlies the characteristic problems of twentieth-century philosophy.
>
> —Putnam, 1983, 15

It is at this point that matters are about to become really excruciating. But fear not. Have faith that if you've made it here, you're ready for it. Go ahead, take the red pill.

Remember how Neo's face twists and writhes just before he pops through the mirror-window and wakes up in the pod? And all those strange people that he's just met are standing around saying "Relax, just try to relax" (they say this with appalling regularity just before something really wrenching is about to happen). The image of the passage down the shoot suggests that *everything happens to Neo in a flash, a tremendously jolting flash*.

In philosophy, the flash gets stretched out over a long discursive trek. If we take the sum of the impact over the entire path, it may not be any less wrenching, but there will be a drastic change in feeling. This

O.B. Bassler, *Diagnosing Contemporary Philosophy with the Matrix Movies*, DOI 10.1057/978-1-137-57889-1

is the difference between the *Matrix* movie experience and the Matrix Philosophy experience.

In movies generally, you can fast-forward (at least at home), but in the case of the *Matrix* movie experience, the movie does the fast-forwarding for you (and most ironically, without any choice on your part). In a flash you've gone from Matrix to Reality. Is it reasonable to think that such an experience of reality-conversion could happen in a philosophical flash? Well, maybe so, but surely not without a good deal of antecedent preparation. A story is told that Wittgenstein "saw" the transition from his early view of language in the *Tractatus* to the view he would develop in the *Philosophical Investigations* on a train when his friend Piero Sraffa made an obscene gesture and said, "What is the logical form of *this*?" (Malcolm 1984, 57–58). But although that was a critical turning point, and maybe even a case of philosophical "conversion," there was plenty of philosophical preparation behind it.

The textual analog of the fast-forward function is *skipping ahead*. This is what we did in the Theatrical Version of this Manifest. It's really only at Level 4 that Matrix Philosophy comes back into any alignment with the *Matrix* movie experience. But in the Theatrical Version we just substituted a "quick and dirty argument" for Level 1.5 (which is the "worst"), and then jumped to Level 2. Maybe some of you skipped down past Level 2, or even Level 3, while others (the brave or foolhardy) are here without having skipped anything at all, in the middle of their first pass through the Manifest. Is there a best strategy? The best strategy, I think, is the one of your own choosing: poke around and see what interests you. Poking around first and then reading in a linear fashion later is something you have to learn how to do when you study higher mathematics. It's a skill that can help in a lot of other walks of life, too.

Putnam's "Models and Reality": A Preliminary Presentation

In this first subsection, I go through the rudiments of Putnam's argument in the paper "Models and Reality," trimming as much as possible subject to the concerns at hand. Strictly speaking, the argument is presupposed by every-

thing that follows in this Manifest, both in this section and beyond, and so it could not be more central to the enterprise than anything else at this point. The reader who knows Putnam's paper may want to reread the final subsection, "Solving the Skolem (Near) Antinomy," before jumping into section "Transition: From Putnam to Yessenin-Volpin" of this Appendix.

Countable versus Uncountable Infinity

The first requisite we need is the distinction between the *countably* and the *uncountably infinite*. It may seem—indeed, I think it *should* seem—that anything which is infinite is just plain infinite: bigger than anything finite and that's the end of the story. But the infinite is a strange territory, and one way to map this territory is to draw a distinction between two basic orders of infinity. The most obvious, and really the most canonical, example of the mathematical infinite is provided by the counting numbers: 1, 2, 3, . . ., as George Gamow used to say. For obvious reasons this establishes the canonical example of a *countably* infinite collection. What, then, is an *uncountable* one?

An uncountable collection of the infinite is one that cannot be paired with the counting numbers. You might think that any collection that includes the counting numbers plus something else would, therefore, be uncountable, but you would be seriously mistaken. For example, suppose we add the number zero to the collection of counting numbers. This can be paired up in the following way: pair 0 with 1, 1 with 2, 2 with 3, et cetera. It turns out that even collections that (in some sense) include a lot more than the counting numbers can be paired with them in this way, like, for example, the rational numbers. We say, therefore, that the positive rational numbers, and also the rational numbers, are *countably infinite*. So, what is an example of a collection that *cannot* be paired up with the counting numbers in this way?

For my money, the best example to look at is the collection of all infinite strings of 1s and 0s: we can call these *binary strings,* since they are built out of the *two* symbols 0 and 1. We can easily build up the collections of binary strings of a given finite length. For example, the binary strings of length 1 are just 0 and 1. The binary strings of length 2 are 00, 01, 10, and 11. The binary strings of length 3 are 000, 001, 010, 011,

100, 101, 110, and 111. Notice that there are two strings of length 1, four strings of length 2, and eight strings of length 3: in general, there will be 2^n strings of length n, where n can be any finite number, and where 2^n means 2 multiplied by itself n times; in the expression 2^n, n is called the *exponent*. This is to say that the number of binary strings is growing *exponentially*. Using an argument usually called the *diagonal argument*, we can show that the collection of infinite binary strings (i.e., strings of 0s and 1s that have a countable infinity of entries and so "go on forever") *cannot* be paired with the natural numbers. We therefore say that the collection of infinite binary strings is *uncountable*.

The Skolem Paradox

Putnam says that "in some way" the "Skolem paradox" underlies the "characteristic" problems of twentieth-century philosophy; the "in some way" allows a lot of leeway, but I'll just say here that I think there *is* a way in which what Putnam says is right, and I offer my gloss on that way a bit further down the road, or way; (count the number of occurrences of "way" in this sentence and include this one (!?): count away!).[1] But before we can say what the Skolem paradox is we first have to meet the Löwenheim-Skolem (L-S) Theorem, and the presentation of the distinction between the countably and uncountably infinite was in preparation for that.

The L-S Theorem is usually paraphrased by saying that any theory of a certain type that has an infinite model has a countable model.[2] Everything is going to hang on what is meant by "model," though, so this paraphrase doesn't really get us very far. Consequently, I'll stick fairly closely to Putnam's presentation, since the whole point is to get at the philosophical point he's trying to make. For our purposes, we focus on the *force* of the L-S Theorem in its generation of a purported "paradox," and then see how Putnam responds. As much as possible, I let this "force" of the L-S Theorem represent the meaning of the theorem (for us), rather than any paraphrase as in the form(s) given above. This is not so much a principled commitment on my part as a strategic, pragmatic one, but as we'll see, a lot of the force of the entire set of issues has to do with the

extent to which we need to abandon a principled for a more pragmatic orientation.

Here's how Putnam presents the purported paradox. Consider, he says, the (quasi-formal) sentence:

$$\neg(\exists R)(R \text{ is one to one. The domain of } R \subset N. \text{ The range of values of } R \text{ is } S)$$

In something like English, the sentence says "it is not the case that there is a relation R which is a unique pairing, where the domain of the relation is contained in the natural numbers and the range of values in the relation is all real numbers." This amounts to saying that the real numbers are uncountable (instead of "countable" and "uncountable," Putnam uses the terms "denumerable" and "non-denumerable"). Since the natural numbers and the infinite binary strings are both uncountable and can themselves be paired, if you like you can substitute "the collection of infinite binary strings" every time you see "the real numbers" and form the paradox in these terms.

A model, roughly, is a certain sort of thing, an "ontological domain," if you will, that either *satisfies* or *fails to satisfy* a particular theory. If it satisfies the theory, which means that all the statements of the theory are "true in the model," then we say that the model is a model *of* the theory. So, in all models of the real numbers, the sentence given above, which says that the reals are uncountable, will be true in the model—since I can, indeed, *prove* that the reals are uncountable. But the L-S Theorem says that if the theory has infinite models, then it must have a *countably* infinite one. So, in this countable model, there is some set that is "the real numbers in this model," and since the *whole* model is countable, the set which is "the real numbers in this model" must be countable (since this set is a subset of the whole model, and an infinite subset of a countably infinite set must itself be countably infinite). But the sentence above says that this set is *uncountable* (in the model), and because this sentence is part of the theory this model satisfies, it must be true. So, how can it be true of this countable set that it is uncountable in the model?

The key observation is that the definition of "uncountability" is a *negative* one: a set is uncountable if there is *not* any unique pairing of the collection with the counting numbers. But in the model, we're not just given an ontological domain in the sense of the underlying objects being specified; we're also told what relations are and aren't available. So, even if we, "outside" the model, can see a possible relation that establishes a unique pairing of the set S with N (which is the natural numbers *in the model*), if no such relation is given as part of our ontological domain (i.e., the model), then *in the model,* the set S will be uncountable, even if "from outside" we can see that it is countable. At this point, the reader should start thinking about the similarities to the brains-in-vats scenario discussed in the text. The explanation I've just given tempts us to draw the conclusion that "what is reality *for us* is not reality *in the model.*" This tempts us to think there might be something like "being a brain-in-a-vat" in which reality would be different (and somehow self-contained) from what reality is for us. Notice, however, that the notion of model is *not* really something "self-contained," since we are "in reality" the ones specifying it. And therein lies the trouble.

Skolem Paradox versus Skolem "Antinomy"

The so-named Skolem paradox does not trouble Putnam, and he's happy to resolve it along the lines indicated above. What troubles Putnam is something he refers to as the Skolem "antinomy, or close to an antinomy" (Putnam 1983, 2), and this is where the philosophical problem sets in. Putnam frames this near-antinomy, in particular, as a problem for the moderate realist, who desires to avoid "mysterious perceptions of mathematical objects" but remains nonetheless committed to "a classical notion of truth" (Putnam 1983, 4). The pejorative reference to "mysterious perceptions" is aimed at Platonists who are committed to mathematical objects inhabiting some realm beyond our natural world, and yet of which we are somehow capable of direct inspection. In contrast to the Platonists, Putnam does make it fairly clear what he means by the classical notion of truth: this is the commitment to something being true without its necessarily being knowable *as* true. Putnam's argument

is designed to show that there is no sense in which one can be committed to such a classical notion of truth without being committed to the mysterious perceptions the moderate realist would want to reject. That is, Putnam's argument presents an antinomy, or something like it, for the moderate realist. Putnam uses a stronger version of the L-S Theorem, the "downward" L-S Theorem, to argue against the moderate metaphysical realist, and I think his argument is best viewed as one in which we use the L-S Theorem to identify a problem with what the moderate metaphysical realist will identify as a "gap" between model and reality.

Perhaps, one might suggest, the problem that the L-S Theorem poses has to do with the fact that not enough has been pinned down by the original theory to fix the notion of the countably versus uncountably infinite, and that if we simply amplify our theory, then we will be able to pin down such distinctions. The downward L-S Theorem, Putnam argues, shows that this is not so. Appealing to it, one can argue (I leave out the details) that even a formalization of all of science (if such were possible) or even beyond that a formalization of all our beliefs, scientific or not, could not pin down such notions. Furthermore, even if we add all the empirical measurements we could make in the world, what Putnam calls "operational constraints," that won't help, either. The moral, as Putnam frames it, is: "there certainly seems to be a *countable* model of our *entire body of belief* which meets all operational constraints" (Putnam 1983). "Operational constraints" here refers to everything we could ever do to measure things empirically in the world. The point is that what we can do is finite, even if it is not fixed by a definite finite bound in advance. The countable domain, therefore, serves as a sufficient model for all operational constraints, and so (roughly speaking) adding in operational constraints could never suffice to pin down the distinction between the countable and the uncountable.

As Putnam goes on to say, this philosophical state of affairs can pull in either of two directions: either we admit a mysterious "intuition" of mathematical objects like the uncountable collection of real numbers, or we claim that all it *means* to say that the real numbers are uncountable is that we know how to give a proof of the sort given above to show that the infinite binary strings are uncountable. In the first case, we fall into the radically metaphysical camp of the Platonists, and in the other we fall into the camp of what Putnam calls the verificationists (realist or

otherwise). Putnam himself will go on to lobby for an empirical realism grounded in the commitment to meaning-as-verification.

The Status of the Notion of "Collection"

I have been using the generic term "collection" for what Putnam refers to more technically as a set. Putnam focuses the problem which the whole Skolem crux discloses for philosophy by honing in on the question: "Is there a problem with the notion of a 'set'?" (Putnam 1983, 11). Beyond what has already been pointed out, Skolem went on long ago to declare, as Putnam points out, that "even the notions of 'finite,' 'infinite,' 'simply infinite sequence,' and so forth turn out to be merely relative within axiomatic set theory" (quoted in Putnam 1983, 2). But what if the very notion of *set itself* turned out to be relative? That is, what if what set theory (as a sort of "mathematical theory of everything") was trying to formalize *itself* turned out to be relative? If we have some independent, platonic intuition of "set," then clearly this problem is avoided, but such an intuition is "mysterious" and "radically metaphysical." If we don't, what then? It seems we're sunk, but miraculously, Putnam draws the line right here: "But this is not so" (Putnam 1983, 11). It will, however, come at a philosophical price: we will have to give up our traditional, "moderate realist" conception of truth.

The way Putnam argues may at first blush look circular; in fact, it is an instance of the form of transcendental argument we've encountered in the brains-in-vats context in Chap. 3. Just as Putnam there *assumed* the notion of brains-in-vats to show that there was a problem with this assumption, so Putnam assumes here, not just the formal concept of model but also a particular philosophical commitment to what such models are and do. Then, he shows that on such a view there just isn't anything it could mean to ask about the referent of the term "set" in the model, and this will cast aspersions on the way of thinking about the model that was assumed in the first place (by the moderate realist, essentially). Two possible "patch jobs" are investigated: first, to "push the problem back," which corresponds to trying to make an argument about "brains-in-vats-in-the-image"; and second, to adopt an external theory of reference, which fails for a somewhat more sophisticated reason. Since the patch jobs don't work, we're forced to conclude that there isn't anything to the problem beyond what we can do

with these models, and there can't be any problem there, because it would have to amount to a problem of not-being-able-to-do-what-we-are-able-to-do, which is just plain contradictory. The upshot is that, according to Putnam, there is no good reason to espouse moderate realism.

Drawing inspiration from Michael Dummett, Putnam recommends that we understand meaning as the use to which we put language. Roughly, to put language to use *just is* to mean. To many, this will seem less occult only in the sense that it occultly rules out any further appeal to occult explanations—it is second-order occult. We are simply supposed to see (in a metaphorical sense?) that any appeal to occult explanations cannot help, much as the scales fell from Wittgenstein's eyes in his exchange with Sraffa on the train. Whether that constitutes a significant improvement is a matter for philosophical debate, but not one that greatly interests me.

Solving the Skolem (Near) Antinomy

How does this "solve" the Skolem near-antinomy? As is generally the case with antinomies, it solves it in a dissolutive rather than a resolutive sense. What I mean is that it calls into question the terms according to which the so-called problem emerged—really, seemed to emerge—in the first place. As Putnam puts it, Skolem showed that model theory could only fix its (internal) objects of reference up to isomorphism. In other words, anything that fit the structure of the theory was good enough. As Hilbert, a forerunner of the model-theoretic orientation in some ways, once said, geometry can be about beer steins and coffee mugs if you can get them to follow the axioms. But then, Putnam says, it starts to seem like reference must be something occult, and that there must be some sort of appeal to non-natural mental powers in order to pin down this referential indeterminacy. Various sorts of appeals can be made, but they all land themselves in the same philosophical morass.

Putnam concludes that the problem, then, must be "with the predicament itself" (Putnam 1983, 23). He frames the predicament most starkly as follows:

1. On any view, truth is truth in the (intended) interpretation (i.e., our understanding).
2. Understanding fixes the reference.
3. If use doesn't determine reference, then use isn't understanding.
4. Language (on the model-theoretic view) has a full program of use, but lacks an interpretation.
5. Therefore, we must make an occult appeal (i.e., to something outside of use) to fix the truth.

What Putnam urges us to reject is (4): "To adopt a theory of meaning according to which a language whose whole use is specified still lacks something—namely its 'interpretation'—is to accept a problem which *can* only have crazy solutions" (Putnam 1983, 24). With this moral in mind, we can see that the point of assuming the model-theoretic approach to meaning was to explode premise (4) *from the inside*.

Notice, finally, that both the Platonist and Putnam accept premise (3), but for antithetical reasons. The Platonist believes that use doesn't determine reference, so use isn't understanding. Putnam thinks that use does determine reference, and so is understanding. Both hold a referential conception of truth, but for different reasons, according to their differing conception of reference. From a perspective external to this debate, one might suggest that the Platonist and Putnam options are disagreeable for antithetical reasons. Putnam is right that the Platonist's occult fixation of reference is disagreeable. But on Putnam's view, though we manage to get by without any such occult appeal, it is unclear that at the end of the day we have any better understanding of reference than the disagreeable metaphysicians.

Transition: From Putnam to Yessenin-Volpin

Maybe—but just maybe—we are now in a position to intimate (but just barely intimate) a sense of what I want to call the *Matrix Problem*. The first moral I want to promote is that *the destabilization of our sense of reality may not come from where we're first inclined to look for it.* Putnam's paper already teaches us this lesson. First, we preoccupy ourselves with

a potential derealization scenario: that although we walk about in the world and it all feels quite real (even hyperreal, like that woman in the red dress), we imagine the dreadful possibility that in fact we're pathetic pod creatures whose life force is being sucked out by some giant mechanism. That doesn't necessarily sound so far off metaphorically (think about the cubicles that are peopled by all the Thomas Andersons in the world), but we'll get back to that later, when we descend to Level 5. On the one hand, if Putnam teaches us anything, it is that this is not, or at least maybe is not, the real threat. On the other hand, maybe the Platonists are right, and this *is* the real threat. On the third hand . . . but we'll have to wait for the third hand. In fact, the preoccupation with this scenario reinforces just those prejudices that underwrite our *real* experience of (philosophical) unreality—our sense that there is some "extra" world out there beyond what we engage in our uses and actions. Why do we fall into this metaphysical illusion? And more pressingly even, how do we escape this Metaphysical Matrix?

The ironies abound. For it is just that insistence on an "independent" reality that Putnam tries to teach us is most *unreal*. But if we aren't inhabiting that thing that we've always taken to be the "real" reality, then from that previous perspective, what we are inhabiting *is a kind of Matrix.* This is not the Matrix a brain-in-a-vat inhabits, since (according to Putnam) that couldn't happen, anyway. This is a Matrix we've been philosophically and metaphysically running away from. This Matrix is the *really real world that we always took to be only a projection!* And once we recognize our previous imprisonment by the image of a world-beyond-the-projection, then we can begin to ask the question: What is the real world? And we see that from the previous perspective, the question that we are asking would have the form: What is the Matrix? What we previously took (in an incoherent way) to be the Matrix has become our real world, and so we are indeed living in a sort of Matrix—the real and only world we have always inhabited. This may or may not make your head swim, but the language is certainly doing some swimming. That's because we're beginning a process of reorientation, and when our orientation changes, our commitment to terms changes, too. This is an important fact of life that is part and parcel of what's so difficult about significant, permanent reorientation.

More questions emerge as we consider the possibility of the process of reorientation that Putnam's argument suggests. Perhaps the most important question is: How can I know if the reorientation I've effected so far is sufficiently thorough? It's very hard to know how to answer this question—indeed, how it could be answered in a *general* way, because this is necessarily a question that is posed from within the context of reorientation. This suggests that we have to look at the particular context of reorientation and evaluate it with great attention and care. These questions will only become thematic at Level 4, but everything we do from now to then serves as a focal object lesson as we attempt to evaluate the particular context of the reorientation Putnam's argument initiates, and it will be equally important to remain mindful of the steps that attempt to deepen this reorientation.

As a preliminary tool, we can focus on the particular concepts that we will subject to scrutiny and use these as wedges to promote the progressive levels of our deepening. Roughly speaking, we will go from the distinction between the countably and uncountably infinite to the distinction between the finite and the infinite, and from there to the distinction between the surveyable and the unsurveyable, and then from there, finally, to the investigation of the nature of proof and the revisionist attitude to proof that is required to take account of the genuine depth of the challenge lying behind the Skolem paradox. Although this trajectory takes us beyond Putnam, serving in fact as a transition from the preoccupations of Putnam to those of Yessenin-Volpin (and David Isles), we never really leave behind Putnam's concerns. For Putnam himself recognizes proof, in the domain of mathematics, and the concomitant abilities of language use in the domain of language, as the ground for his account of meaning, understanding, and reference. What we will see, simply, is that Putnam has not taken these commitments nearly far enough.

We can begin to sense that all is not right in the house of Putnam if we pay attention to Skolem's remark, quoted by Putnam[3] in section "The Status of the Notion of 'Collection'," that not only the countable/uncountable distinction is relativized by the L-S Theorem but also the much more basic distinction of finite/infinite. In fact, this claim that the finite/infinite distinction is relativized is controversial,

and Hartry Field argues that in any case it should be jettisoned in favor of a consideration of the possibile *indeterminacy* of the conception of the finite.[4] I am not myself committed philosophically to this particular model-theoretic version of relativization, but I do want to look at what happens when Putnam tries to take this relativization seriously. Right about here is where things start to get steep: maybe we're finally falling into the world of real philosophy. As Prince says, things are much harder in this world; in this world, you're on your own (Prince 1984). Here we go.

Specifying the Observational Base

If the finite/infinite distinction is relativized, this leads to a big problem (to put it mildly). There are junior and senior versions of this problem. I describe the senior problem here, but then explain why I'm leaving it aside for the moment. The senior problem is that the very notion of model depends on our ability to draw on a countable collection (at least)—namely, the countable collection of symbols that make up the language of the model. If the distinction between the finite and the infinite is relativized, then we have relativized the very notion of the underlying language for our model. To put it in the language of metaphysical mythology—the sort of language the moderate and Platonic realists use, and which Putnam has been working to combat—we don't know if what looks like a collection of symbols that goes on indefinitely is actually a collection of symbols that, "from the outside," actually comes to an end. We're not so much brains in vats here as we are what I like to call bugs on an island, marching along thinking that the island goes on forever. But that's only because our legs are so short. Could we be bugs on an island? Does Putnam's argument against brains-in-vats exclude that as a possibility on purely philosophical grounds? To my mind, at least, the idea that we are bugs on an island, at least metaphorically speaking, is a lot easier to swallow than the idea that we might be brains in vats. Indeed, in some sense we are manifestly, if metaphorically, bugs on an island. We live on a tiny planet in a tiny solar system in a regular-size galaxy beyond which lie whole expanses of space. A lot of space is observable to us, but what's beyond that? This is not the

imagination of something implausible—although the more one thinks about it, the weirder it may seem—it's what we're firmly committed to being the case, according to our best scientific theories. There are no monkeys sitting in a room waiting to type out Shakespeare here.

The reason I'm leaving the senior problem aside for now is that all along Putnam has been engaged in a process of looking at the model-theoretic perspective from the inside. Although he has his own revisionist ideas about what needs to be done with or to this perspective, he's perfectly happy to explain what needs to be done in terms he borrows from this model-theoretic perspective. At the end of his paper, Putnam confidently asserts that "from the viewpoint of non-realist semantics"—that is, from his very own viewpoint—"the metalanguage is completely understood, and so is the object language" (Putnam 1983, 24–25). Putnam is giving up the *viewpoint* of realist semantics, but he's not giving up the *project* of formalized semantics, and in particular he's still fine with the distinction between object language and metalanguage. In fact, the program seems to be to retain as much of the model-theoretic orientation as is compatible with the abandonment of semantic realism, which essentially amounts to the rejection of premise (4) described in section "Solving the Skolem (Near) Antinomy."

So, playing along with Putnam for the moment, I defer concerns about the viability of the very framework of non-realist semantics and concentrate on the junior problem. The junior problem has to do with the formalization of the observational base within the theory as Putnam describes it in the course of his argument that there is not a problem with the notion of set and in particular, that the L-S Theorem does not lead to the relativization of this notion. Since it's Putnam's argument we're focusing on now, and he uses the term set rather than collection, I transition over to his vocabulary on this point.

The first obvious, but fundamental thing to note is that the notion set is the very thing that set theory, as a theory, is intended to model. If we give up the philosophical claim that there is a categorical sense of 'set' that is captured by our theory, then we've really relativized away the whole shooting match. This is a whole different, and much bigger, thing than conceding the relativity of the countable/uncountable distinction, for example.[5] The second thing to note is that Putnam is arguing for

the fixation of the notion of set not within set theory itself but within a much more ambitious theory, T_I, which is a kind of idealized theory-of-everything (at "the end of time"). That means, to begin with, that the argument for the fixation of the concept of a set is extremely idealized. But more relevantly for our purposes, it means that the observational base has been fixed (all the way out to the end of time, whatever that means). For all we know, that observational base is unending; and even if it's not, we certainly can't fix any definite finite bound on it in advance. That means we have to include it in our theoretical language as what is technically known as an *ω-sequence* (omega sequence)—that is, a countably infinite list of observational specifications.

Without even digging into the guts of Putnam's argument, we can see that Putnam is committed to the ontological status of such ω-sequences, since they are in fact the very constraints within the theory T_I that Putnam uses to tie down the notion of set. That is to say, they are what makes it possible for Putnam to argue (in this context) for an empirically realist conception of set. This point has been noticed, and stressed, by Hartry Field in his book *Truth and the Absence of Fact* (Field 2001). In the chapter "Radical Indeterminacy," Field questions the determinacy of the notion of ω-sequence, which as a mathematical concept certainly falls prey to the relativization accruing from the L-S Theorem. Here, the larger context of discussion is the extent of indeterminacy of our language vocabulary, so Field is already looking at the senior problem I've mentioned. Nonetheless, some of his remarks can be restricted to the junior context, although we will see that the solution he proposes cannot.

Field considers two possible characterizations of ω-sequence. In the first, an ω-sequence is defined as "a linear ordering which has a first member and no last member, and where each member has only finitely many predecessors" (Field 2001, 263). To make this definition definite, the notion of finite would already need to be definite, for starters. If the L-S Theorem leads to the relativization of the finite/infinite distinction, or if the finite/infinite distinction is indefinite on independent grounds, that just means that (in our current context) Field's approach is out (or at least needs independent defense); in that case, we just can't specify the notion of ω-sequence in that way. The second definition of an ω-sequence is that "the ordering is just like that of the natural numbers" (Field 2001,

263). That definition, of course, relies on the definiteness of the natural numbers. We can characterize the set of natural numbers formally, in which case it falls prey to L-S style relativization, but foundationally (as opposed to formally) this definiteness is usually argued for on the basis of the definiteness of the notion of set and the associated notion of member (what it means for something to be a member of a set). *But the definiteness of the notion of set is just what Putnam is arguing for*. So, we clearly can't appeal to it, in this case, to argue for the definiteness of the notion of an ω-sequence, on pain of circularity.

Field wonders, in fact, whether any of these notions can be made sufficiently determinate to determine the notion of an ω-sequence. I think he's right to wonder. Field thinks Putnam is in trouble, and goes on to argue, modifying Putnam's argument, that the most reasonable hope is that we can pin enough of this down by constraining our theory through reference to physical objects and vocabulary, which in turn relies on the appeal to constraints stemming from observational practices (Field 2001, 266). Field calls this "impure set theory," and asks the question: Are there any "bad"—that is, nonintended models of this theory? If there are, then once again we haven't fixed the notion of set categorially—that is, definitely—and we're back to the problem of the indefiniteness of the notion of set.

What Field suggests is that if we are "empirically lucky," our acceptance of this theory will serve to rule out nonintended models. The major point here is that whether this will work or not, according to Field, will be an empirical matter: "it is not *a priori* obvious that any physical ω-sequences can be singled out in a (sufficiently) determinate way" (Field 2001, 268). Field is taking seriously, in a way it seems Putnam really is not, what obligations accrue on a commitment to empirical realism. The point is not that there are no a priori conceptual truths, but that these will not be enough to fix our most basic commitments—in particular, the commitments we face in arguing for the definiteness of our most fundamental mathematical notions. As Field goes on to insist, if the notions of finite and ω-sequence cannot be made sufficiently definite, this poses problems for the determinacy of basic logical notions like not and there is as well (Field 2001, 269). Relativization at the root of mathematics generates relativization at the roots of formal logic, too. To make this argument, Field does graduate from the junior to the senior problem, but the senior problem is one we need to face eventually in any case, so this advertise-

ment is certainly relevant. Field ends the section under consideration by remarking "all I really want to claim is that there is a worry here which deserves serious discussion" (Field, 2001, 269). Indeed there is, and in a later chapter Field goes on to discuss it.

In this later discussion, in the chapter "Undecidable Sentences," Field directly responds to Putnam's paper "Models and Reality," and specifically on the issue of the concepts of finiteness and natural number (Field 2001, 338–42). Here, Field spells out certain "cosmological assumptions" that undergird the sort of argument for the determinacy of the notion of set that Putnam makes in his paper. I won't bore (in both senses) into the gory details of Field's argument, since I'm more interested in what he thinks such reasonable "cosmological assumptions" might be. Field enlists two basic assumptions (Field 2001, 340)

(A) Time is infinite in extent.
(B) Time is Archimedean.

Adding these two assumptions to set theory, Field gets a theory he calls S, which, he argues, will allow us *"to extend the determinacy in the physical vocabulary to the notion of* [mathematical] *finiteness*" (Field 2001, 341).

It will be helpful to introduce a predicate, call it $\Phi(Z)$, which will allow us to frame the two assumptions in common terms. Let $\Phi(Z)$ mean: Z is a set of events that (i) has an earliest member and a latest member; and (ii) is such that any two of its members occur at least one second apart (Field 2001, 340). Then (A) says that there is no finite bound on the sets satisfying Φ, and (B) says that only finite sets satisfy Φ. We need the two of these together so that our mathematical conception of the finite can be "inherited" from the physical base. Notice that the predicate Φ depends on our physical capacity to measure time intervals of (greater than) one second. That is, it depends on establishing and being able to measure a basic unit of time.

Now notice that (A) makes the determinacy of the mathematical conception of finite depend on the indefinitely remote condition of our universe—namely whether it continues to exist or not, and whether it continues to make sense under indefinitely remote conditions to continue to speak of an "arrow of time" (a notoriously vexing problem in the

philosophy of science). However, (B) is the real kicker. What (B) is ruling out is that there could be any *physically* infinite sequences with a smallest discrimination ("one second apart") and which are bounded (have a first and last member).

But notice that this is exactly how we have described the bug's trajectory above: from the bug's perspective, the trek across the island "goes on forever," and yet from our perspective, the bug's trajectory has a first member (its first step) and a last (the last before it dies), and yet *we* can see that all of this takes place from our perspective within a finite number of steps. If there are "bug histories" that hand down the treks of previous generations, and these generations are cumulative, we may need a somewhat larger island, but that's all. This is all starting to sound very much like the sort of relativization that goes on as a consequence of the L-S Theorem.

After much philosophical reflection, you might decide that it is or it is not like that, but my opinion, which I will not defend in detail here (but which is at least supported by the discussion of Isles's program in section "Yessenin-Volpin and the L-S Theorem: The Work of David Isles"), is that it is like the L-S relativization in all the relevant regards. Therefore, Field's assumptions beg the question, albeit in a more sophisticated way. That is, these assumptions simply presuppose (rather than identify) the distinction between the finite and the infinite in the physical base. When we incorporate these assumptions into our larger theory S, they can still be interpreted differently internally from the "bug's perspective" ("this is an ω-sequence") than from the external perspective ("this is a finite sequence"). The problem is twofold. First, the assumptions may be physical in the sense that they are about physical conditions, but there is another sense in which the physicality of the assumptions is dubious: What would it even mean for the assumptions to be physically true? To think that we can answer this question is already to assume that there is a determinant distinction between the "bug's perspective" and the "external perspective" and that we are inhabiting the latter. But effectively, that is just to assume what we are seeking to prove. One can try to haggle over the interpretation of the model theoretic details, but even if these details can be smoothed out, that

should just make us more dubious about the model theory. The second problem is that even if we could say what would make the assumptions physically true, that would manifestly not be something we could determine empirically. Not that it's Field's concern, but that means that even if his argument were somehow to work, it wouldn't support Putnam's defense of empirical realism per se.

Field responds to several other objections. First, although these objections concern me less, vetting them briefly will allow me to bring some of my own concerns into greater relief. Field agrees that there is something troubling about using physical hypotheses to secure the determinacy of mathematical concepts, commenting "I sympathize—I just don't know any other way to secure their determinacy" (Field 2001, 342). This is a strategy that Field inherits from Putnam, more or less: Putnam appeals to the observational base, which Field finds question-begging, and so he appeals instead to "physical hypotheses." Yet the empirical content of Field's assumptions is dubious, and specifically on the point at issue, which they beg![6]

The second potential objection is more interesting because this time Field does have a response: Isn't it especially objectionable to rely on physical hypotheses that use the very mathematical terms these hypotheses are then intended to support? (Notice that this is not the criticism I have given above, which is, rather, that the physical hypotheses do not secure the determinacy of the distinction *even in the physical domain*.) Field responds to this potential criticism by noting that, on his view, we needn't secure a concept before using it in reasoning: the point of securing its determinacy is to reinforce our confidence in using it, not to sanction our using it in the first place. To this I am generally sympathetic, and only reiterate that using a concept to justify mathematical determinacy does require that its appearance in the physical context be determinate to secure determinacy in the physical domain, and this is just what I have contested. In other words, Field is arguing that if we can secure the determinacy of the mathematical concept in a limited domain (the physical), then this justifies its determinacy in the general domain of the mathematical. I grant the conditional claim, but deny the antecedent: we

have not secured the determinacy of the concept in this limited domain in a non-question-begging way.

Field's intuition that Putnam cannot establish the determinacy of the notion of set (or of finite or the natural numbers) simply by specifying the observational base strikes me as correct and hard-hitting, but when he tries to find "physical" assumptions that would fix this determinacy in the physical realm so that it can be imported into the mathematical domain, his assumptions ultimately assume the determinacy in the physical domain that he seeks in the mathematical. Instead of trying to push the problem back into the metalanguage, Field has tried to embed a miniature version of its solution within the theory S in the form of the hypotheses (A) and (B). The intuition that the physical notion of finiteness implied by these specifications is determinate is really just another version of metaphysical realism, and one that begs the question "from the inside." I doubt that Putnam would have any truck with it. We are left in the skeptical position of concluding that we have no good reason to take these basic mathematical notions as determinate.

Is there *any* way out of this? One possibility we might explore is that, so to speak, Putnam is right but not for quite the right reasons. But right about what? My suggestion is that Putnam is right that we do need something like a "non-realist semantics" based, for mathematics, in the appeal to proof; I'll leave the case of "ordinary language" aside for the time being. But for philosophical purposes, for us to look at such a proof-based semantics in the context of axiomatic-set theory, with its distinction between object language and metalanguage, we need at a minimum the determinacy of the notion of set, since this is what axiomatic set theory is a theory of. *And we don't have that*. So, what are we going to do? It seems we somehow have to "redo" proof theory, and we will do it by *relativizing the concept of a set in the relevant context*.

That, in fact, is exactly what we need to do, and in particular, we need to redo proof-theory in such a way that we can keep track of those places where the indeterminacy in such fundamental concepts as set, finite, and natural numbers becomes an issue. But what, exactly, have we been relying on in our prior commitment to the determinacy of these notions? Well, generally speaking, what we've been relying on is that every time we appeal to them they mean the same thing. And this only really causes a

problem if there are places where the assumption that they mean the same thing each time begs the question: that is, if this assumption assumes what, in any given case, we are trying to prove. So, if we want to diagnose the extent of the problem, we should begin by trying to identify such instances of question-begging. There is little hope, really, that we would be in a position to know whether we had identified all instances, or all types of instances, of question-begging. That's just philosophical life, but it's also true that every little bit helps. It helps, at least potentially, by allowing us to deepen our understanding of the relevant concepts.

What I'm advocating is a certain way of "turning the philosophical enterprise around," not so much on its head, as in 180 degrees. Instead of trying to rule out problems in advance, let's try to identify problems as we go along and understand what can be done about them. Suppose you and I have an argument, which we resolve, and then we find out later that we were using a key term in two different ways. So far from being science fiction, this is something that happens all the time. A simple, but far from the most interesting case is, say, that we might have been referring to two different people. You say Bart is tall, and I say Bart is tall, and so we agree, but then it turns out we're talking about two different Barts. Of course, we don't know for sure that our discovery of this difference isn't predicated on some further equivocation; that's just life. But we can ask the question: What does this say about the agreement we previously reached? Does it stand anyway—as it will if both Barts are tall; or if it doesn't stand, can we revise it, or do we need to jettison it? As we know from life, there will not be one uniform answer to this question. Progress, to the extent we can even identify it, is ongoing.

The next step, the step into the land of Yessenin-Volpin, attempts to bring mathematics in line with this practical wisdom. It makes mathematics more complicated, and so it's not likely to be endorsed by the mathematical community until its members have good practical reasons to do so, which may be a long time in coming or never, but in any case it doesn't seem likely to happen soon. (Then again, it didn't seem likely in 1989 that the Berlin Wall would come down anytime soon.) But that's not the point here: the point here is philosophical, and nobody ever said philosophical life is simple (though a few diehards have insisted it should be).

Yessenin-Volpin and the L-S Theorem: The Work of David Isles

If the finite/infinite distinction is relativized—if we fundamentally can't tell whether or not we're bugs on an island—then doesn't this induce skepticism about whether we are really specifying the observational base? Yes, but all is not lost: this just pushes the general problem down to a level where Putnam isn't yet considering it. To address the problem now we first have to "redo" our orientation, moving from model theory to proof theory, but this is really where Putnam says we should be anyway. However, the problem reappears in the context of proof theory as well: this is Yessenin-Volpin's critique of "proofs without soul." And this, in turn, leads us to Level 2, where we consider the finite/infinite distinction and its significance for the foundations of mathematics. Our goal is to see how this effects what we can call the transition from the logical to the mathematical.

The first thing we have to do, then—and it will perhaps be the single biggest thing we have to do in this book—is plunge into the bowels of proof theory. In particular, we need to understand Yessenin-Volpin's idea of "proofs without soul." Because Yessenin-Volpin's own writings are extremely difficult to understand, I'm relying heavily on work done by David Isles to promote Yessenin-Volpin's general program.

To begin with, let's reframe the general reason Putnam gives for moving to a proof-based semantics. It seems that for Putnam, the basic idea is that what we *mean* is something that is rooted in how we back it up. In a general context, if you ask me a question—like, Why did you go to the grocery store?—I give you an answer that explains the justification for my action: I needed some onions. A lot of the time our answers to questions about knowledge may eventually throw us back on appeals to beliefs. If you ask me why I'm working on this Manifest, I'll tell you that I believe it's important, or that it whiles away the hours, or that I find it amusing, or something else. In any of these cases, not just the Manifest (perhaps not even the Manifest) has a value but also the working on it has a value. That value is something I believe in, anyhow. It may not ever have even occurred to me explicitly, and I may not be able to (or need to) formulate

it, but to the extent that I can (or even could) provide a justification, what I'm articulating is some sense of *value*. That, at least, is one way of looking at it; I don't pretend to be solving any philosophical problems here. What is peculiar about the situation, philosophically speaking, is thrown into dramatic light when we start thinking about knowledge claims. In particular, how could a claim to *knowledge* ever be backed up in any really satisfying fashion by an appeal to *belief*, when belief seems quite generally to be something with a weaker justification status than knowledge? Even if I might be able to justify my actions generally by appealing to beliefs, how am I going to make this stick in the case of the very particular sort of action that is making a knowledge claim? What good, for example, does it do to say that the reason I *know* there is a stone standing at my feet is because I *believe* it to be there? That doesn't seem to provide any sort of help at all.

For my purposes, I want to suggest that there are two general ways of trying to address this problem. One way is to recognize that there's something special about knowledge claims, and that therefore the justification of them requires something different, maybe even radically different, from what is usually required when we are asked to justify our actions generally. Call this, if you like, the "metaphysical route." We reencounter it later, when we turn to Descartes. One of the major problems with this route, at least to my mind, is that as a historical matter of fact, the strategy of finding a radically different way to deal with the justification of knowledge claims almost inevitably ends up appealing in some way or another to the determinacy of the finite/infinite distinction. I'm not claiming that this is conceptually inevitable, since such a claim would likely involve me in just the sort of circularity that metaphysicians find themselves tangled up in with their tacit or explicit reliance on the determinacy of the finite/infinite distinction. Rather, I'm just telling you that this is the result of my own historical investigations (Bassler 2015, 104–109).

Putnam is coming from somewhere similar, at least in this regard; he's convinced that there's an endemic problem with what I've called the "metaphysical route." And he also has the virtue of seeing that dislocating ourselves from this metaphysical strategy (to the extent we are even able) requires us to get clear about the fixity (or lack thereof) of basic notions like set, the natural numbers and the infinite (and hence,

by implication, the distinction between the finite and the infinite). So far, so good. The problems begin when we ask: What is the alternative? Finding an answer to this question is not the problem—indeed, the only apparent answer to the problem is readily forthcoming: we have to treat the justification of knowledge claims the way we treat the justification of claims about our actions more generally. The problem comes when we try to figure out how to do this.

We can begin by asking: What is the analog for "backing up actions" in the domain of truth? If we take mathematics as a laboratory for thinking about this question, then at least the answer is readily forthcoming: what backs up a mathematical truth-claim (in general) is a proof. It seems that roughly, at least, this is Putnam's motivation for turning to a proof-based semantics. One reason to be skeptical of this maneuver is that Putnam already finds himself in the context of a highly mathematized treatment of truth to begin with—namely, the formalized semantics of the model-theoretic tradition. I don't want to underestimate the extent to which this context may trump up the sorts of answers that are being suggested. But in any case, the appeal to basic mathematical distinctions (as part of the general strategy I've referred to as "metaphysical") does, as a matter of historical fact, go back well beyond the twentieth-century tradition of model-theoretic semantics; it goes back to Descartes, and to Plato. (The other, and more challenging, problem, which I will not to try to hit with any force until we descend to Level 3, is how this approach might be extended beyond the mathematical domain, beyond the pseudo-Matrix of number strings streaming down a "moving screen." It may seem that the larger domain of action, or as I come to call it, *praxis*, should be more amenable to this sort of treatment. But that intuition, I believe, changes once we have appreciated the problems exposed in our mathematical laboratory.)

If proof is what backs up claims to mathematical truth, then the question of clearing ourselves of charges of circularity devolves into the domain of proof. Is there a way we can escape the charges of circularity in our appeal to the definiteness of notions of set, natural number, and the finite/infinite distinction if we shift our focus from the domain of models to the domain of proofs? In the first instance, an analogous problem of circularity can be identified in the way the domain of proofs is typically conceived—a charge of circularity which is not only equally serious to

that faced in the model-theoretic domain but also arguably one which is a residual reflection of model-theoretic thinking in the way that the domain of proof is conceived. To explain this, it is best to proceed by example, and the examples that most naturally exemplify concerns about the status of set, natural number, and the finite/infinite distinction are claims about the basic operations on natural numbers and the associated proofs backing up the properties of those operations. The most basic operations on the natural numbers are (arguably, and typically taken to be) those of succession, addition, and multiplication. While the latter two reach back into the darkest recesses of our earliest mathematical education, the former, the operation of succession, is not usually singled out as such early on, and would instead simply be described as "adding 1." In other words, the successor of 1 is 2, the successor of 2 is 3, and so on.

Notice how this characterization of the successor operation almost amounts to a characterization of natural numbers themselves. If the successor operation is not definite, then certainly natural number will not constitute a definite mathematical concept. But there's more to the concept of natural number than just the successor operation, because of the nefarious "and so on" that comes at the end of the sentence ending the preceding paragraph. How are we to understand this "and so on"? In the bugs-on-an-island sort of way, or in the "genuinely going on forever" sort of way? Or is there even any conceptual fact of the matter distinguishing what these two different "ways" are trying to point out? In the idea of the successor operation and the application of it over and over, the problem we've been zooming in on is already implicated.

If we are going to make any progress thinking about this in a non-question-begging way,[7] we had better proceed by asking ourselves what the "and so on" could mean without prejudicing whether we ourselves have a bugs-on-an-island perspective or a genuinely unending one. That is, we must assume, provisionally, that when we say the natural numbers "go on forever," we don't really know which of these two perspectives this unendingness describes. Because it is often confused with a form of finitism, let me stress that the methodological orientation I insist on is one of *agnosticism*, not either finitism (there is nothing truly infinite) or infinitism (there is something truly infinite). We have seen that we're not in a good position to answer this question, so we had better proceed in

some fashion that brackets it. (This particular act of bracketing serves as an exemplification for a more general methodology of bracketing, which I call *paraphysical bracketing*, and which modifies the concept of bracketing found in the phenomenological tradition.)

So here's how we proceed. Since the idea of applying the successor operation "over and over"—whatever that means—is tantamount to the notion of natural number, we start with this idea and then ask: What does it mean to perform addition and multiplication on these natural numbers? In particular, we can ask the question: If it's possible to apply the successor operation "over and over," does that mean that when we perform addition or multiplication we can always find a sum (addition) or product (multiplication) that is one of these successive results of applying the successor operation? If we can (or could), then we say that these operations are *total*. The claim, for example, that addition is total takes the form: For all natural numbers *l* and *m*, when we add them together, we get another natural number *n*. A little bit more formally, we can write this as:

$$\forall l \forall m \exists n (l + m = n)$$

where the quantifiers ∀ ("for all") and ∃ ("there exists") range over the natural numbers. That just means that when I say "for all" I mean "for all natural numbers" and when I say "there exists" I mean "there exists a natural number." In talking this way, therefore, we are assuming that it makes sense (in some way) to talk about "ranging over" the natural numbers, though there are various more specific things this could mean (like, "every time I pick something out I should be able to tell whether it's a natural number or not" versus "there is a giant collection of natural numbers all gathered together and I can pick things out of it at my whim"), and I insist that we remain agnostic with respect to these more specific things as well. This is another instance of what I call *paraphysical bracketing*; more on that as a general topic at Level 3.

So, we are supposing it means something for these quantifiers to range over the natural numbers, and we're considering the claim that addition is total (this claim is therefore relative to our commitment to quantifying

over the natural numbers, whatever they are, and whatever that means). Now, how might we back up such a claim? We need a proof. The first thing to recognize about proofs is that *there can be many different proofs of the same truth*. So, we can't just make claims about all proofs of a particular fact (unless we want to commit to quantifying over proofs, which, for various reasons, I promise you we don't want to do, at least not any time soon). What we can do is look at a particular proof that might be supplied for this truth and consider its structure. How might we try to prove that the addition operation is total? Basically, we have to come up with a candidate proof and then analyze how convincing we find it. In other words, we can supply a candidate proof and then try to look for potential ways in which it may or may not be begging significant questions. This is what David Isles does in his paper "What Evidence Is There That 2^65536 Is a Natural Number?" Most of what follows is simply a partial precis of what he discusses there (Isles 1992).

Isles's presentation is axiomatic in the sense that we start off with certain axioms, or unquestioned assertions, that describe the basic properties of the system we are considering. These assertions tell us, in particular, properties of the basic operations of succession, addition, multiplication, basic facts about the number zero, and how the operations of succession, addition, and multiplication relate to each other. The ideas are not hard (at least on the face of it), and the axioms are only intimidating if they are given in formal terms to someone unfamiliar with the formalism. For this reason I state them in as close to plain English as is reasonably possible, and for completeness I include all the axioms.

There are six basic axioms involving the operations of succession, addition, and multiplication, and this group establishes the most basic system of arithmetic. Since Isles calls the first group of axioms "Q axioms," I label them that way, too. To denote the successor of x, I use the notation $x|$.

Q1: For every natural number, the successor of that natural number is not 0.
(0 is the name that we give to the very first natural number.)

Q2: For all natural numbers x and y, if the successor of x is equal to the successor of y, then x is equal to y.

Q3: For all natural numbers x, if I add x plus 0, I get x.
Q4: For all natural numbers x and y, $x + y = y + x$.
Q5: For all natural numbers x, x multiplied by 0 is equal to 0 (i.e., $x \cdot 0 = 0$)
Q6: For all natural numbers x and y, x times the successor of y is equal to first multiplying x times y and then adding x (i.e., $(x \cdot y|) = x \cdot y + x$). For example: $2 \bullet (3 + 1)) = (2 \bullet 3) + 2$.

That concludes the basic axioms of our arithmetic of natural numbers. Now, consider the claim that for all natural numbers x and y, there is a z such that $x + y = z$. We supply what in mathematics is known as a *proof by induction*. There are plenty of issues surrounding the status of the induction principle that underwrites such proofs, but I leave those to one side for now. The induction principle says that if I can show something for the first relevant value (most often 0 or 1), and if I can show that if this same something holds for an arbitrary value n, then it holds for the value $n|$, the successor of n, then I can conclude that it holds for all the natural numbers. The idea behind the induction principle is just a form of the same idea expressed earlier about the natural numbers: you get them by starting with something, namely zero, and then whenever you have a natural number n, you get the next one by taking the successor, which is the same thing as adding 1. And that's the way you get all the natural numbers. So, if we look at whatever the "something" is that we want to claim holds for all natural numbers, what we need to do is show that it holds for the first one, and then show that if it holds for some arbitrary n, it holds for $n|$; by the very characterization of the natural numbers that means our "something" holds for all of them.

In the proof I offer, I hold a particular value of x fixed and then show that no matter what the value of y is, there is some z that is the sum $x + y$. How do I do that? Well, for y equal to 0, it's clear that $x + 0 = x$ by Q3. If we call $0|$ by the name 1, it's also clear that $x + 1 = x + 0| = (x + 0)|$ (by Q4) $= x|$ (since $x + 0 = x$ by Q3 as above). (We don't really need this latter fact now, but below it will be convenient to start the induction with 1 instead of 0.) Now suppose for some arbitrary value of y, call it $y = n$, $x + n = z$. (This supposition is sometimes called the "inductive hypothesis," and I use this name to refer to it in general below.) Then the successor of

y is $n|$, and $x + (n|) = (x + n)|$, by Q4. But $x + n = z$ by the inductive hypothesis, and so $(x + n)| = z|$. It is part of the definition of the natural numbers that every number has a successor—this is just what it means to apply the successor relation indefinitely—and so we have shown that there is a number, namely $z|$, that is the sum of x and $y|$. We have verified the existence of a sum for $y = 0$, and we have shown that assuming there is a sum for $y = n$, there is a sum for $y|$. By the induction principle, this means that for all values of y, $x + y$ has a sum—that is, the addition operation is total.

Both to condense the proofs to follow and also so that we may talk about the operations of addition and multiplication directly, it will help to add additional axioms about addition and multiplication as named operations—that is, properties of the natural numbers. This is not as innocuous or innocent as it perhaps looks, for in doing so we are in fact introducing axioms for *expressions* rather than for the natural numbers themselves. That is, something of the form $x + y$ is an expression in the system of arithmetic rather than being a natural number itself (of course, so was $x|$). To be sure, it names a number. Perhaps the easiest way to see the difference is to notice that $2 + 5$ and 7 are different expressions but they both name the same number (just like $6|$ and 7 also do). Canonically, we refer to the number that both these expressions name as 7, but really that's just the canonical *name* for the number, and we mustn't confuse the name with what it names, any more than you (who are a person) should be confused with your name (which is a piece of language, but which of course becomes a deep part of your social identity). In the context of arithmetic, we refer to these names as *numerals*, and the following axioms are about numerals—that is, expressions naming numbers, rather than about numbers themselves. Beyond that, they are also about arithmetic expressions that state arithmetic truths.

A1: For all natural numbers x, $x + 1$ and $x|$ are names for the same number.

A2: For all natural numbers x, y and z, the truth of the expression $x + y = z$ implies the truth of the expression $x + y| = z|$.

M1: For all natural numbers x, the expression $x \cdot 0| = x$ is true. (If, as above, we substitute the name 1 for the name 0|, then we can write the expression in the form $x \cdot 1 = x$.)

M2: For all natural numbers x, y, z and w, the truth of the expressions $x \cdot y = z$ and $z + x = w$ implies the truth of the expression $x \cdot y| = w$. For example: $4 \bullet 2 = 8$ and $8 + 4 = 12$ implies $4 \bullet (2 + 1) = 12$.

What we showed before was that if x and y are natural numbers, then there is a number z that is the sum of x and y. Now, however, we are able to express the truth: *Arithmetic is a total operation.* For all natural numbers x and y, there exists a natural number z such that the *expression* $x + y = z$ is true.

Let's run the argument for the claim that arithmetic is total. Fix a particular value for x, and call it a. A1 gives the truth of $a + 1 = a|$, which establishes the claim for the value $y = 1$. Now we assume (this is the inductive hypothesis) that for $y = n$ there is a z such that $a + n = z$ is true. From A2 it follows that $a + n| = z|$ is true, and so appealing to the induction principle we have the truth of the expression $a + y = z$ for *any* natural number y, where the value of z will depend on the value of y, of course. But then we note that a was chosen completely arbitrarily, so we have that the expression $x + y = z$ will be true in general, for a value of z which now depends on the values of x *and* y. If we give the expression $x + y = z$ the name A(x, y, z), we can then say that the addition relation A is total.

Now give the expression $x \cdot y = z$ the name M(x, y, z), and call this latter the multiplication relation. We can then prove that the relation M is total by relying on our proof that A is total. In order to consolidate further I present this proof in terms of the relations A and M. Be careful not to confuse these relations with the associated axioms A1, A2, M1 and M2! If you don't want to chase through the details of the next paragraph, you can just take my word that we have such a proof.

Fix $x = a$. M1 gives us that M(a, 1, $a|$); hence $(\exists z)$A(a, 1, z) (i.e., let $z = a|$). Let $y = n$ and assume (inductive hypothesis) that $(\exists z)$M(a, n, z). From the conclusion that A is total we conclude that there is some c such that A(z, a, c), and then from M2 we can conclude M(a, $n|$, c). (Here, let c play the role of w in the axiom M2 and let a play the role of x.) Then,

using induction on y, we can conclude $\forall y \exists z \mathrm{M}(a, y, z)$. But since a was chosen arbitrarily, we can conclude $\forall x \forall y \exists z \mathrm{M}(x, y, z)$. That is, M is total.

So far we have only been looking at the operations of succession, addition, and multiplication, but we could also introduce the next operation down the line, which is exponentiation. You might say, okay, we could do that, but *why* do that? And if we do that, why stop there? Why not introduce the next operation down the line (which is called hyperexponentiation) and the next (which is called hyperhyperexponentiation), and on and on? Well, indeed we could, but once we get to exponentiation (which I define in just a moment), something new and interesting happens. It turns out, at least according to Isles's reconstruction of Yessenin-Volpin's program, that the proof that exponentiation is total is question-begging in a way that the proofs of the previous claims (that addition and multiplication are total) are not. This will serve as our principal example of a question-begging proof—what Yessenin-Volpin calls a "proof without soul." Therein hangs a tale, and pursuant to that tale, we will be in a position to plumb the very depths of the Löwenheim-Skolem Theorem.

Losing One's Soul in the Land of Exponentials

The operation of exponentiation can be thought of as repeated multiplication, just as multiplication is repeated addition. Our goal is to look at the purported proof that exponentiation is total, but along the way we have to consider the problem of how proofs of the totality of the relations of addition and multiplication might also be considered more carefully to determine if they're question-begging or not. Considering this has to do with a certain set of issues that so far we have completely overlooked. This has to do with the way that *values* of natural numbers are being drawn to play the roles of *a, b, c, n, x, y, z,* and *w* in the proofs given in the previous section. In particular, we haven't tracked how the choices of particular values have depended on each other. To take one simple example, but already an important one, when M2 asserts that if $x \cdot y = z$ and $z + x = w$, then the value of z depends on the values of x and y, and the value of w depends on the values of z and x. But in a certain

sense, the value of *w* depends on *z* and *x* in different ways, because *z* already depends on *x*! These relations of dependence will only become more subtle when we put the different parts of the proof together, and it is out of this tangle that the subtle presuppositions in the proof-structure will manifest themselves.

Given that exponentiation stands in the same relation to multiplication as multiplication stands in relation to addition, it is perhaps not surprising that we can offer a proof of the totality of the exponentiation relation by appealing to the totality of multiplication in exactly the same way that we offered a proof that multiplication is total based on the totality of addition. Note that this makes the proof of the totality of exponentiation dependent on the totality of both addition and multiplication, even if we only mention the totality of multiplication explicitly in the proof, since the proof of the totality of multiplication itself depends of proving the totality of addition.

Of course, to make the analogy explicit, we have to introduce axioms for the exponentiation relation analogous to those for addition and multiplication. The exponential, or "power" relation (hence the letter "P" in the following axioms) will be written P(*x*, *y*, *z*), which should be read "*x* raised to the *y* power is equal to *z*," or equivalently, "the multiplication together of *y* factors of *x* is equal to *z*." Here are the two new axioms:

P1: For all natural numbers *x*, P(*x*, 1, *x*).
P2: For all natural numbers *x*, *y*, *z*, and *w*, P(*x*, *y*, *z*) and M(*z*, *x*, *w*) implies P(*x*, *y*|, *w*).

For Example $4^2 = 16$ and $16 \cdot 4 = 64$ implies $4^3 = 64$.

P1 says that *x* raised to the first power is simply *x*. P2 says that if *x* raised to the *y* power is equal to *z* and *z* multiplied by *x* is equal to *w*, then *x* raised to the power of *y*| is equal to *w*. The reader should think about why these axioms make sense in terms of the idea of exponentiation as repeated multiplication. Although I have written these in slightly more compact forms than the axioms above, the reader should pause to think about just how similar P1 and P2 are to M1 and M2.

Given that the relation of exponentiation to multiplication is so similar to the relation of multiplication to addition—even to the point that

the proof of the totality of exponentiation depends on the totality of multiplication in the same way that the proof of the totality of multiplication depends on the totality of addition—it may be quite surprising that there should be any deep difference in the statuses of these two proofs. Where does the *disanalogy* lie?

Into the Labyrinth: Analyzing the Proof of Arithmetic Totality (Caution: Minotaur Within)

Let's start at the beginning (a very good place to start), with the proof for the totality of addition. Actually, this isn't the beginning, because we've already moved from the level of numbers to the level of the relation A(*x*, *y*, *z*), and as I already remarked, this move is anything but innocuous. (Isles just comments that at the level of these relations the issue is "particularly clear" [Isles 1992, 468], but I think that's some serious soft-pedaling.) For now I focus on the issue of the sort of dependencies intimated above in section "Losing One's Soul in the Land of Exponentials," and that certainly is easier to talk about in this latter context. The main point is to come to understand how dependencies arise between the ranges from which both constant values and values of variables are being drawn, and we do this by focusing on the proof that the relation A(*x, y*, *z*) is total. In other words, we look at the proof of the proposition that for all values of *x* and *y*, there is a value *z* such that $x + y = z$. The tricky question will be: What sort of values are we talking about for *x, y,* and *z*?

The proof that addition is total uses axioms A1 and A2 to make an inductive argument. It is helpful to write these axioms somewhat more formally than we have above in section "Yessenin-Volpin and the L-S Theorem: The Work of David Isles" in order to keep things straight. Since these are exactly the same axioms, I give them the very same names and hereon just use these new expressions of them. A1 is now given this way:

A1 $(\forall x)\mathrm{A}(x, |, \mathrm{x}|)$.

This can be read: "For all natural numbers *x*, $x + 1$ equals the successor of *x*." A2 is now given as:

A2 $(\forall x)(\forall y)(\forall z)[\mathrm{A}(x, y, z,) \Rightarrow \mathrm{A}(x, y|, z|)]$.

This can be read: "For all natural numbers x, y, and z, x plus y equals z implies that x plus the successor of y equals the successor of z." (The parentheses and brackets are being used to keep things straight; brackets are used the same way as parentheses; using both just looks a little nicer.)

When we actually use these axioms in the proof, we have to replace the variables (x, y, and z) by particular values. We didn't see this earlier, and this is the fine detail of the argument that we overlooked. Say, we replace x by the value a in A1; we represent this by $x \rightarrow a$. (Note that the single arrow ($\rightarrow$) is doing a very different sort of work than the double arrow ($\Rightarrow$). In the first case, the single arrow indicates the range of substitution-values for a variable; in the second case, the double arrow represents a logical implication.) In A2, we need to make three replacements—values for the variables x, y, and z, which appear there. Since the x that appears there is a different variable (after all, it appears in a different formula!), but since it must be assigned the same value, we need to do something with our notation to respect this difference and identification. So, instead of calling the three variables in A2 x, ,y and z, we call them x^*, y^*, and z^*; that way, we are able to keep track of the fact that these are the variables assigned to A2. Since x^* must be assigned to the same value as x, and since we have written $x \rightarrow a$, we have also $x^* \rightarrow a$. We must also assign values for y^* and z^*; since y^* will ultimately need to play the role assigned earlier to n in the inductive step, I write $y^* \rightarrow n$. For z, we just need some generic letter for the value substituted, and the next one after a is b, so I write $z^* \rightarrow b$. In writing this we are assuming that $a + n = b$; call this *Assumption One*. Later on, we must find a way to "discharge" this assumption, but for the moment we are relying on it. On the basis of this assumption, and using the particular values we've selected for application in A2, we have $a + n = b$ and $a + n = b \Rightarrow a + n| = b|$, from which we can conclude that $a + n| = b|$. This means that for a variable u there is some value (namely, $b|$) such that $a + n| = u$—that is, $(\exists u)\mathrm{A}(a, n|, u)$. Since here we have "back-substituted" a variable u for a value $b|$, we write $u \rightarrow b|$.

Now, for the inductive step, we assume that there is some z such that $a + n = z$; call this *Assumption Two*. This may seem a very strange thing to do, since we already have that $a + n = b$; why would we need to assume

such a z when we already have the value b? Note, too, that this is different from what we did earlier with u, where we let the variable u stand for the value $b|$. The answer to the question is that we are trying to keep straight the various parts of the argument, and the assumed variable z has to do with the inductive hypothesis—in fact, it is "representing" the role of the inductive hypothesis in the argument. But since z and u are now playing the same role, we write the identification as $z = u$. But also (and this is perhaps even harder to wrap one's head around) we had earlier that $a + n = b$ and we are now assuming that $a + n = z$, so we have to write the substitution as $b \rightarrow z$. Why do we do this? Because it allows us to discharge Assumption One and replace it with Assumption Two. (Later, we get rid of Assumption Two by appealing to the induction principle.)

So, where do we stand? On the one hand, from A1, we have the assertion A(a, |, $a|$), and from our analysis stemming from A2, we have ($\exists u$) A(a, $n|$, u). If we "back-substitute" a variable u' for $a|$ in the first assertion, we get ($\exists u'$)A(a, |, u'), and if we identify $u' = u$, then these are the same assertion. This allows us to discharge the inductive hypothesis—that is, Assumption Two (since, with the equality $y = u$, it is now supported independently from A1), and applying the induction principle we get that for the "arbitrary" values a and n, ($\exists$u)A(a, n, u). But since a and n were arbitrary, we can replace them with universally quantified variables, noting the substitutions $a \rightarrow v$ and $n \rightarrow w$ and write: $(\forall v)(\forall w)(\exists u)$A($v$, w, u). This says that the addition relation A(v, w, u) is total!

Now we have analyzed the entire proof for the totality of the addition relation A, paying full attention to the ways values are drawn for variables and variables are drawn to represent values. Let's list all the relations involved: $x \rightarrow a$, $x^* \rightarrow a$, $y^* \rightarrow n$, $z^* \rightarrow b$, $u \rightarrow b|$, $z = u$, $b \rightarrow z$, $u' = u$, $a \rightarrow v$ and $n \rightarrow w$. What we need to do now is look for potential "circularities" in these substitution relations. Such circularities can arise when the same letter appears on both the left-hand side and the right-hand side of (different) relations, taking identifications (equalities) also into account. For example, x appears on the left-hand side of an arrow, but it doesn't appear on the right-hand side of any arrow and it's not identified with anything else. On the other hand, a appears on the right-hand side of the first substitution, $x \rightarrow a$, and on the left-hand side of the substitution $a \rightarrow v$. This means we can get from x to v. But v doesn't appear on the left-hand side of

any substitution, and it's not equated with anything, so there's no "circle" here. I won't go through all of the analysis exhaustively, but we can find one circle, and it's a circle that exemplifies an interesting feature. Using the three pieces $u \rightarrow b|$, $z = u$, $b \rightarrow z$, we can get from z, which is equal to u, to $b|$, and since b then goes to z, we can get $z = u \rightarrow b| \rightarrow z|$ (since $b \rightarrow z$). This means that, cumulatively, we have $z \rightarrow z|$. So, if z is contained in the range over which we are drawing, then so is $z|$. This implies in turn that we must be drawing the variable $u = z$ over a range of values that is *extended by the successor function*. To start with, this means that proof of the totality of the addition operation assumes the totality of the successor operation.

This illustrates the general principle that the availability of certain expressions implies the commitment to the availability of certain other expressions, but in fact something much more potent is involved once we start looking at the induction principle. Since the induction principle allows us to draw a conclusion about all values n if we can show something for the case $n = 0$ and the conditional implication (true for $n \Rightarrow$ true for $n + 1$), we are presupposing that each number n does have a successor $n|$, which means not only that we are presupposing the totality of the successor function for this particular position in the argument but also, crucially, that all non-numeral numerical terms can be reduced to this basic numeral form (i.e., a sequence of strokes: the numeral name for the number six is ||||||). This is because whenever the induction principle is applied, we are making the assumption that any numerical term (non-numeral or numeral) refers to a natural number in the natural number sequence to which the induction principle applies, and so must be nameable by some numeral. David Isles describes this by saying that the range of this position "must be the numerals." Let me explain in more detail what that means.

Suppose I am in a situation where I am committed to the totality of the successor function but not (yet) committed to the totality of the addition function. It may be that in particular cases I can convert addition expressions, like $4 + 3$, into the language of succession—that is, in this case seven strokes: |||||||. But asserting that this is something that can always be done is something more. And, indeed, to assert without further argument that this is something that can always be done would beg the question of the totality of the addition operation, since it amounts to a

procedure for performing the addition operation *at the level of expressions*. And so, if our argument for the totality of the addition function tacitly relied on such a conversion from additive notations like $4+3$ to stroke notations like |||||||, then we could accuse it of being question-begging. But our argument will rely on this for the specific variable-slot in the argument over which induction is being performed, and so we need to make sure that this doesn't make the entire argument question-begging. To anticipate, it won't in the cases of the proofs for the totality of addition and multiplication, but it will in the case of the (attempted) proof for the totality of exponentiation.

To see this, we need to go beyond looking at the substitution relations and equalities considered earlier. Once again, I do not present the whole analysis but, rather, focus on a chunk of it sufficient for the purposes at hand. In the proof for the totality of addition, the inductive variable is *y*, and so by applying the induction principle to it we have effectively assumed that the range of this variable is the entire range of numerals, in the sense that there will be a reduction procedure for any non-numeral (arithmetic or successor) term to the corresponding numeral term. On the other hand, if we have an expression like $4+3$ in the *z*-slot, then the expression $(4+3)|$ will be valid for the *z*-slot, since we have the substitution relation $z \rightarrow z|$; but unless we already have the expression ||||||| (i.e., 7 strokes) in the *z*-slot, we can't conclude that we have the expression |||||||| (i.e., 8 strokes) as part of our reference-range there. And in general, even if we do have both $4+3$ and ||||||| as expressions in the range of the *z*-variable, we will not in general be able to reduce the one to (i.e., replace the one by) the other. It is only in the case that such a reduction is possible in general that we say that "the range must be the numerals."[8]

The totality of multiplication is expressed by the formula $(\forall x)(\forall y)(\exists z)$ M(*x*, *y*, *z*), and here, too, the argument proceeds by induction on the second variable, *y*. This means that the range of *y* must contain the entire range of numerals here as well. But in the proof of totality for multiplication, we make use of M2. In this axiom, *y* occupies the second slot in the expression M(*x*, *y*, *z*) and *x* occupies the second slot in the expression A(*z*, *x*, *w*). Since the proof for the totality of multiplication assumes the proof for the totality of addition as a prerequisite (in fact, the latter is

embedded as a part of the proof for the totality of multiplication), we now have induction over both the variable *y* (for multiplication) and the variable *x* (for addition). So, the range of the *x* and *y* variables must be the numerals. That's still not a problem, because it's *z* that is the variable for the multiplicative sum, and for that we're not assuming that the range must be the numerals (which would then beg the question about the totality of the multiplication relation). Note, however, that because *x* also occupies the first slot in M(*x*, *y*, *z*), we now have that both the first and second slots in the M relation must have the numerals for their range. But in the case of the exponentiation relation, we make use of P2. In this axiom, *z* is in the first slot of M, and this means that now we have that *x*, *y*, and *z* must have the numerals for range. But that means that the very inductive conditions *presuppose* that the relation P(*x*, *y*, *z*) is total. The proof for the totality of exponentiation is therefore question-begging.

This argument is very subtle,[9] and it may be helpful to go over it briefly again in slightly different language. The point is really to keep track of which slots in the various relations must have the numerals as their range in the different proofs. In the proof of the totality of addition, only the second slot must have the numerals as its range. In the proof of the totality of multiplication, the second slot of A must have the numerals as its range in the first and second slot of M (but crucially not the third slot). In the proof of the totality of exponentiation, the second slot of A must have the numerals as its range, the first and second slots of M must have the numerals as its range, but *all three* slots of P must have the numerals as its range. That means the proof for the totality of P is question-begging. In some sense, we can say that the proof of the totality of P is question-begging because it comes third in line.[10]

Löwenheim-Skolem, Finally, Again: Liberation Mathematics

Having practiced looking at the dependencies among values and variables in the context of the totality proofs, we are now in a position to draw some general philosophical morals that promote Isles's analysis of the L-S Theorem. In the section "Losing One's Soul in the Land of Exponentials" discussion, we drew a distinction between numeral and

non-numeral terms. This is a distinction within the domain of numerical notations. Here's where the fact that we've been considering mathematical expressions comes in: numeral and non-numeral terms are types of mathematical expressions, and we can go down one of two routes, which I call Option 1 and Option 2.

We can assume from the get-go that all non-numeral terms can be converted into numerals, and then we can assume that when we plow through a proof the ranges from which we draw our numerical notations all coincide, since they can all be "reduced" to the numerals. Call this Option 1. Isles's point is that if you do that, then there's really no reason to prove that the operations of addition, multiplication, exponentiation, et cetera are total in the first place, since you've effectively already assumed it, at the level of numerals. The other option, Option 2, the one Isles has been introducing us to, is that the numerical notations can vary over the course of a proof.

Isles says Option 1 makes what he calls a "semantic assumption," which is over and above the commitment to the proof being something that preserves truth (i.e., converts true premises into true conclusions). This seems right, since we've seen that there is a very real sense in which Option 2 retains a commitment to the preservation of truth in the proof-theoretic structure without assuming that the range of reference remains fixed over the course of the proof. Although not assuming that all values and variables are drawn from the same range does make our life more complicated, it is more general than the typical orientation, and Isles argues (convincingly, I think) that it is more purely syntactic (i.e., purely oriented to truth-preserving) than Option 1, which overlays the semantic assumption of range equality over the structure of truth-preservation.

But if we adopt Option 2, we have to face a radical implication of our change in stance, which is that what we are talking about (i.e., the semantics) changes over the course of the proof. As the proof-structure gets laid out, more and more constraints are imposed on the way the values and variables can be drawn from their respective ranges. We've seen just exactly how this happens in the analysis above in section "Into the Labyrinth: Analyzing the Proof of Arithmetic Totality (Caution: Minotaur Within)," and in the case of the application of the induction principle in particular, this has some rather drastic consequences. Indeed, we constrained ourselves to such an extent that by the time we reached the proof for the totality of exponen-

tiation we had already committed ourselves at the level of expressions to what we were claiming to prove at the level of numbers.

Isles points out, however—and, again, I think rightly—that the phenomenon of variation of reference is ubiquitous in both the sphere of ordinary language use and in mathematics. To idealize away from this in the mathematical domain is to pretend that mathematics is a special case where we're equipped in advance with an eternal and unchanging collection of objects. That's a form of the Platonism Putnam has discussed. And, anyway, the semantic assumption, which is the relevant residue of an implicit Platonism in the context we're dealing with, can be considered as a limiting case of Isles's program in which all constraints on the ranges of values and variables are exhaustively stipulated in advance. So, in one sense, in pursuing the second option, we're just looking at something more general. But in another sense, the shift registers a conceptual commitment to the variability of the objects underlying the process of proof understood as a syntactic (truth-preserving) venture.

Isles works up the notion of this variable range of reference supporting a syntactic proof-structure and calls it a "Buridan-Volpin structure," after the medieval philosopher John Buridan and the twentieth-century poet-mathematician-philosopher-political iconoclast A. S. Yessenin-Volpin, whose work serves as the proximate inspiration for Isles's program. I won't go through the technical details of a Buridan-Volpin (BV) structure; suffice it to say that it charts the transformation of commitments of the sort we looked at in the proofs of totality of the relations A, M, and P.

A proof, considered as a syntactic object which is truth preserving, takes a collection of premises and transforms them, by pre-approved rules, into a conclusion. If we represent the list (or "sequence") of premises by Γ and the conclusion by Δ, then a derivation D of Δ from Γ can be expressed as $\Gamma \vdash_D \Delta$. A Buridan-Volpin structure, on the other hand, is a dynamic reference assignment for values and variables. It is dynamic in the sense that reference evolves over the course of the proof. What we want is to show that whenever a BV structure V is compatible with the derivation D and the mapping of values and variables into V—that is, whenever it respects the structure of the derivation (in a way which Isles makes precise) and the way values and variables are being assigned—then if it makes all the premises true (i.e., satisfies the premises), it will make the conclusion true (i.e., satisfy the conclusion) as well. Isles is able to show that his definition

of BV structures does just this, and so the *syntactic* structure of the proof is reflected in the *semantic* (reference) BV structure.

On the basis of this definition of BV structure, Isles is able to present an analog of the L-S Theorem that specifically applies to BV structures. First, we must (roughly speaking) pick out the proofs that aren't question-begging in the sense that we saw earlier that the proof for the totality of exponentiation was. Our commitment to Option 2 requires us, in particular, to reject all question-begging proofs as vacuous, but even if we didn't make this commitment, it's clear that we shouldn't expect (nontrivial) BV structures for such vacuous proofs—indeed, that's just the point! In the context of the natural numbers, such proofs require us to fix our reference frame in advance as the range of all numerals, and so in particular our frame will be *infinite*. What is profoundly interesting is that in non-question-begging contexts, we can always find finite BV structures that satisfy the proof-structure![11]

Call a non-question-begging proof a "quantifier-normal" (or "Q-normal") derivation. Then Isles's analog of the L-S Theorem says that if a sequence of formulas Γ has a model in the "regular" (i.e., Option 1) sense, and we have a Q-normal derivation D such that $\Gamma \vdash_D \Delta$, then there is a BV substructure of M that is a *finite soul* for D. This means that BV is a finite (reference) structure and is compatible with the derivation D in the sense indicated earlier. It gives us a reference frame for making sense of the result in a non-question-begging way. It liberates our mathematical soul from the oppression of fixed reference and exorcises the infinite specters of standard model theory.

Recall that the original L-S Theorem said that if a syntactic structure had an uncountably infinite model (i.e., reference structure), then it already had a countably infinite one. This led to the relativization of the distinction between the countably and uncountably infinite. Isles's version of the L-S Theorem says, roughly, that if a syntactic proof structure (of the right, non-question-begging type) has a model, then it already has a finite "model with soul." Although this could be used to argue for a philosophical version of finitism (or in Yessenin-Volpin's case, "ultrafinitism"), this is not the philosophical moral I wish to draw. Rather, it seems that this exemplifies the relativization of the distinction between the finite and the infinite in a very deep sense. But for now I leave this only as a tantalizing suggestion. Instead of Putnam's advocacy of a proof-based

semantics, it advocates a new dynamic semantics that does justice to the syntactic constraints of proof. Perhaps this is the royal route out of the twentieth-century philosophical morass, during which the Löwenheim-Skolem "problem" hung over the entire philosophical arena like an ominous cloud. Perhaps the sky is beginning to peek through (Girard 2015).

As Isles says, his version of the L-S Theorem has some curious (and surprising!) consequences. First, assuming that set theory (which is the reigning "foundation of mathematics" in the standard, Option 1 sense) is consistent—that is, it contains no contradictions—Isles's L-S Theorem implies that any theorem in the first-order theory of sets (i.e., quantifying only over sets, not over properties of sets) has an equivalent form with a finite BV model. This is a sweeping result, indeed. The analogous result also holds for the first-order theory of the natural numbers, and there is even an extension to a weak system of second-order logic.[12]

Isles ends his paper with a provocative comment. From the traditional (Option 1) perspective, if we group together the two formulas $(\forall x)F(x)$ and $(\forall y)\neg F(y)$, where the symbol ¬ stands for negation, we have an immediate contradiction. For the first statement says, "for all natural numbers x, $F(x)$ is true," and the second statement says, "for all natural numbers y, it is not the case that $F(y)$ is true." Since from the traditional perspective, both x and y range over all natural numbers, the two statements cannot be compatible. But from the perspective of Option 2, we can find BV structures that support both statements simultaneously. We just have to make sure that the reference ranges for x and y don't overlap. If we adopt this perspective, it seems to present a picture in which mathematics could be full of all sorts of "potential contradictions" that we simply haven't met yet because the reference ranges haven't come into collision. But recall that in the context of some future proof, the two reference ranges may be constrained to interact and clash; in that eventuality, we will not say, "Oh dear, there's a contradiction," throw our hands up in the air, and give up. (Neither will Neo. Neither will the Oracle. Eventually they will both penetrate the alien reference range.) Call such a collision, when it first manifests itself, an *anomaly*. (This is precisely what Neo is called at the end of the first and the beginning of the second *Matrix* movie.)

Isles takes Option 2 to support, and to be supported by, a passage from philosopher Ludwig Wittgenstein:

> One would like to say: the proof changes the grammar of our language, changes our concepts. It makes new connexions, and it creates the concept of these connections. (It does not establish that they are there; they do not exist until it makes them.) (quoted in Isles 1992, 527)

Isles's program gives us a limited formal model of Wittgenstein's point, though Wittgenstein's point is ultimately deeply yet subtly different from the moral Isles himself draws. This leads to problems in the philosophy of mathematics that are beyond our bounds here (Bassler 2015). For now, our consideration of the Löwenheim-Skolem Theorem draws to a close, and we are ready to (re)turn to Level 2.

Notes

1. A way a lone a last a loved a long the (Finnegans Wa(y)ke, where the end is the beginning of the way/why/y. . . Let me count the y's . . .) (Joyce 1967, 628).
2. Putnam remarks: "The Löwenheim-Skolem Theorem says that a satisfiable first-order theory (in a countable language) has a countable model" (Putnam 1983, 2).
3. Though without giving a citation. The matter is cleared up, and then some, by Lavine 1994, 127n26.
4. Field 1998, 116–17, argues especially against a particular way of construing Skolem's and Putnam's claim; see also Buena 2005.
5. Though not the finite/infinite distinction; if we relativize that distinction, we arguably have problems with the very notion of proof itself. Field discusses these issues and searches for some possible responses in Field 1998. This set of issues leads from a focus on the finite/infinite distinction to the surveyable/unsurveyable distinction; see Bassler 2006.
6. This question-begging has also been identified, in model-theoretic terms, by Buena 2005, 76.

7. By stipulation, the use of more than two hyphens in a single term means that we are writing in something I give the technical name of *philosophical German*. We are close to it here. The reader is invited to confirm that we have passed into this earlier in the Manifest, but the author pleads that it was not his fault, as he was supplying the hyphens that had been artificially suppressed by philosophers pretending to write in philosophical English. Pretty soon, though, it will be his fault.
8. Isles's language is potentially misleading here, but I'm doing my best to clear it up. Note how at Isles 1992, 472, he specifies the range of u as including all the numerals, but also other non-numeral numerical expressions. To say that the range "must" be the numerals (effectively) means that all non-numeral numerical expressions are eliminable from the range—that is, that the range is (after reduction) *only* the numerals.
9. If my reading is correct, Isles has misspoken at the top of Isles 1992, 473, when he says that in the context of the proof of the totality of the multiplication relation M, the conditions permit the inclusion of "non-numeral numerical exponential terms." That is true, but what Isles really needs to say to point out that the proof is not question-begging is that it permits the inclusion of non-numeral numerical *multiplicative* terms, and I read him accordingly.
10. The reader should compare the "quick and dirty argument" by Nelson, given in Chapter 3.
11. Since these finite structures grow as a function of the complexity of the proof and are therefore unbounded, they are *parafinite* in my sense of the term.
12. In this weak, second-order system, the only second-order terms are second-order parameters or constants. In this system, impredicatively defined sets typically have different finite interpretations in different derivations (Isles 1994, 526–27).

Bibliography

Adorno, Theodor. *Alban Berg: Master of the smallest link*, trans. and ed. Juliane Brand and Christopher Hailey (Cambridge: Cambridge, 1991).

Adorno, Theodor. *Philosophy of New Music*, trans. Rubort Hullot-Kentor (Minneapolis: University of Minnesota, 2006).

Akhmatova, Anna. *Poem Without a Hero and Selected Poems*, trans. Lenore Mayhew and William McNaughton (Oberlin: Oberlin, 1989).

Aronofsky, Darren. *π*. DVD Artisan 60494 (1998).

Auerbach, Erich. *Dante: Poet of the Secular World*, trans. Ralph Manheim (Chicago: Chicago, 1961).

Badiou, Alain, et alia. *"Matrix": machine philosophique* (Paris: Ellipses, 2013).

Baird, Bryan Neal. *Transcendental Arguments and the Call of Metaphysics*, University of Georgia Dissertation (2003).

Bakhtin, M. M. *The Dialogic Imagination*, trans. Caryl Emerson and Michael Holquist, ed. Michael Holquist (Austin: Texas, 1981).

Bakhtin, M. M. *Problems of Dostoevsky's Poetics*, trans. and ed. Caryl Emerson (Minneapolis: Minnesota, 1984).

Bakhtin, M. M. (1968). *Rabelais and His Word*, trans. Hélène Iswolsky (Cambridge: MIT, 1968; repr. Bloomington: Indiana).

Bakhtin, M. M. *Speech Genres and Other Late Essays*, trans. Vern W. McGee, ed. Caryl Emerson and Michael Holquist (Austin: Texas, 1986).

O.B. Bassler, *Diagnosing Contemporary Philosophy with the Matrix Movies*, DOI 10.1057/978-1-137-57889-1

Bakhtin, M. M. *Toward a Philosophy of the Act*, trans. and notes Vadim Liapunov, ed. Michael Holquist and Vadim Liapunov (Austin: Texas, 1993).
Bassler, O. Bradley. *Kant, Shelley and the Visionary Critique of Metaphysics*, manuscript in preparation.
Bassler, O. Bradley. "Marlene and Maria," *Southern Humanities Review* 39, no. 2 (2005): 167–70.
Bassler, O. Bradley. *The Pace of Modernity: Reading with Blumenberg* (Melbourne: re.press, 2012).
Bassler, O. Bradley. *The Long Shadow of the Parafinite: Three Scenes from the Prehistory of a Concept* (Boston: Docent, 2015).
Bassler, O. Bradley. "The Surveyability of Mathematical Proof: A Historical Approach," *Synthese* 148, no. 1 (2006): 99–133.
Baudrillard, Jean. *Seduction*, trans. Brian Singer (New York: St. Martin's, 1990).
Baudrillard, Jean. *Simulacra and Simulation*, trans. Sheila Faria Glaser (Ann Arbor: Michigan, 1994).
Baudrillard, Jean. *The System of Objects*, trans. James Benedict (London: Verso, 1996).
Bergman, Ingmar. *Fanny and Alexander*, trans. Alan Blair (New York: Pantheon, 1982).
Beyssade, Jean-Marie. *La philosophie première de Descartes. Le temps et la cohérence de la métaphysique* (Paris: Flammarion, 1979).
Blanchot, Maurice. *The Madness of the Day*, trans. Lydia Davis (Barrytown: Station Hill, 1981).
Bloom, Harold. *The Anxiety of Influence: A Theory of Poetry* (New York: Oxford, 1973).
Bloom, Harold. *Omens of Millennium: The Gnosis of Angels, Dreams, and Resurrection* (New York: Riverhead, 1996).
Blumenberg, Hans. *Höhlenausgänge* (Frankfurt am Main: Suhrkamp, 1989).
Blumenberg, Hans. *Paradigms for a Metaphorology*, trans. Robert Savage (Ithaca: Cornell, 2010).
Blumenberg, Hans. *Work on Myth*, trans. Robert M. Wallace (Cambridge: MIT, 1985).
Borel, Émile. *Les Nombres Inaccessibles* (Paris: Gauthier-Villars, 1952).
Borel, Émile. *Valeurs pratique et philosophie des probabilités* (Paris: Gauthier-Villars, 1937).
Bos, Henk J. M. *Redefining Geometrical Exactness: Descartes' Transformation of the Early Modern Concept of Construction* (New York: Springer, 2001).
Bostrom, Nicholas. *Superintelligence* (New York: Oxford, 2014).

Bostrom, Nick. "Are you living in a computer simulation?" *Philosophical Quarterly* 53, no. 11 (2003): 243–55.
Bostrom, Nick. "The Simulation Argument: A Reply to Weatherson", *Philosophical Quarterly*, 55 (2005): 90–97.
Bostrom, Nick. "The Simulation Argument: Some Explanations", *Analysis* 69 (2009): 458–461.
Bostrom, Nick, and Kulczycki, Marcin. "A Patch for the Simulation Argument," *Analysis* 71, no. 1 (2011): 54–61.
Brandist, Craig; Shepherd, David; and Tihanov, Galin, eds. *The Bakhtin Circle: In the Master's Absence* (Manchester: Manchester, 2004).
Brueckner, Anthony. "Brains in a Vat," *Journal of Philosophy* 83 (1986): 148–67.
Brueckner, Anthony. "If I am a Brain in a vat, then I am not a Brain in a Vat," *Mind* 101 (1992):123–128.
Brueckner, Anthony. "The Simulation Argument Again", *Analysis* 68 (2008): 224–226.
Buena, Otávio. "On the Referential Indeterminacy of Logical and Mathematical Concepts," *Journal of Philosophical Logic* 34 (2005): 65–79.
Buscombe, Edward. *Cinema Today* (London: Phaidon, 2003).
Calderón de la Barca, Pedro. *Life is a Dream*, trans. with a foreword and introduction Edwin Honig (New York: Hill & Wang, 1970).
Calude, Cristian S., ed. *Randomness and Complexity, From Leibniz to Chaitin* (New Jersey: World Scientific, 2007).
Camus, Albert. *The Rebel: An Essay on Man in Revolt*, trans. Anthony Bower (New York: Vintage, 1956).
Cavell, Stanley. *Pursuits of Happiness* (Cambridge: Harvard, 1981).
Cavell, Stanley. *The World Viewed: Reflections on the Ontology of Film*, enlarged edition (Cambridge: Harvard, 1979).
de Cervantes, Miguel. *Don Quixote*, trans. Edith Grossman (New York: Harper Collins, 2003).
Chaitin, Gregory. *Conversations with a Mathematician: Math, Art, Science and The Limits of Reason* (London: Springer, 2002).
Chaitin, Gregory. *Exploring Randomness* (London: Springer, 2001).
Chaitin, Gregory. *Meta Math!: The Quest for Omega* (New York: Pantheon, 2005).
Chaitin, Gregory. *Thinking about Gödel and Turing: Essays on Complexity, 1970–2007* (New Jersey: World Scientific, 2007).
Chaitin, Gregory. *The Unknowable* (Singapore: Springer, 1999).
Chalmers, David. "The Matrix as Reality," in Grau, 2005, 132–176.

Chihara, Charles. "Wittgenstein's Analysis of the Paradoxes in his *Lectures on the Foundations of Mathematics*," *Philosophical Review* 86, no. 3 (1977): 365–81.
Clark, Katerina, and Holquist, Michael. *Mikhail Bakhtin* (Cambridge: Belknap, 1984).
Coates, Ruth. *Christianity in Bakhtin: God and the exiled author* (Cambridge: Cambridge, 1998).
Conway Morris, Simon. *Life's Solution: Inevitable Humans in a Lonely Universe* (Cambridge: Cambridge, 2003).
Debussy, Claude. *Préludes pour Piano* (Paris: Durand, 1913).
Dennett, Daniel. "Artificial Life as Philosophy," in Christopher G. Langton, 1995, 291–92.
Derrida, Jacques. *Of Grammatology*, trans. Gayatri Spivak (Baltimore: Johns Hopkins, 1976).
Descartes, René. *Meditations on First Philosophy, with Objections and Replies*, trans. John Cottingham, in *The Philosophical Writings of Descartes*, 3 vols. (Cambridge: Cambridge, 1984), II 3–383.
Dostoevsky, Fyodor. *The Brothers Karamazov*, trans. Richard Pevear and Larissa Volokhonsky (New York: Vintage, 1991).
Ehrenzweig, Anton. *The Hidden Order of Art: A Study in the Psychology of Artistic Imagination* (Berkeley and Los Angeles, 1967).
Epstein, Richard L., and Carnielli, Walter A. *Computability: Computable Functions, Logic, and the Foundations of Mathematics*, 2nd edition (Belmont: Wadsworth/Thomson Learning, 2000).
Fadely, J. Patrick. "Fantasy Baseball," unpublished poem.
Field, Hartry. "Do We Have a Determinate Conception of Finiteness and Natural Number?" in Matthias Schirn, ed. *The Philosophy of Mathematics Today* (Oxford: Clarendon, 1998), 99–130.
Field, Hartry. *Truth and the Absence of Fact* (Oxford: Clarendon, 2001).
Fletcher, Angus. "Allegory Without Ideas," *boundary 2* 33, no. 1 (2006): 77–98.
Fletcher, Angus. *A New Theory For American Poetry: Democracy, the Environment, and the Future of the Imagination* (Cambridge: Harvard, 2004).
Frisch, Walter. *The Early Works of Arnold Schoenberg, 1893–1908* (Berkeley: California, 1993).
Gaiser, Konrad. *Platons Ungeschriebene Lehre* (Stuttgard: Ernst Klett, 1962).
Gibson, William. *Neuromancer*, 20th annivesary ed. (New York: Ace, 2004).
Girard, Jean-Yves. "Transcendental syntax I: deterministic case," revised November 24, 2015, available at http://iml.univ-mrs.fr/~girard/trsy1.pdf and listed as under consideration for publication at *Math. Struct. in Comp. Sci.* as of December 29, 2015.

Girard, Jean-Yves. *The Blind Spot: Lectures on Logic* (Zürich: European Mathematical Society, 2011).
Godard, Jean-Luc. *In Praise of Love*, New Yorker DVD B00009YXIS (2003).
Grau, Christopher, ed. *Philosophers Explore the Matrix* (New York: Oxford, 2005).
Haber, Karen, ed. *Exploring the Matrix: Visions of the Cyber Present* (New York: St Martin's, 2003).
Haight, Amanda. *Anna Akhmatova: A Poetic Pilgrimage* (New York: Oxford, 1976).
Hardin, Garrett. *Stalking the Wild Taboo*, 3rd ed. (Petosky: Social Contract, 1996).
Harnad, Stevan. "Levels of Functional Equivalence in Reverse Bioengineering," in Christopher G. Langton, 1995, 293–301.
Haydn, Joseph. *The Symphonies*, cond. Antal Dorati, Philharmonia Hungarica, 33 discs + booklet, Decca 448 531–2 (1996).
Hickey, Lance P. "The Brain in a Vat Argument," *Internet Encyclopedia of Philosophy*, http://www.iep.utm.edu/brainvat/.
Hill, Geoffrey. *The Orchards of Syon* (Washington: Counterpoint, 2002).
Husserl, Edmund. *Cartesian Meditations*, trans. Dorion Cairns (The Hague: Martinus Nijhoff, 1960).
Husserl, Edmund. *Experience and Judgment: Investigations in a Genealogy of Logic*, revised and ed., Ludwig Landgrebe, trans. James S. Churchill and Karl Ameriks (Evanston: Northwestern, 1973).
Husserl, Edmund. *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy, First Book*, trans. F. Kersten (Dordrecht: Kluwer, 1982).
Husserl, Edmund. *Philosophy of Arithmetic*, trans. Dallas Willard (Dordrecht: Kluwer, 2003).
Isles, David. "A Finite Analog to the Löwenheim-Skolem Theorem," *Studia Logica* 53 (1994): 503–32.
Isles, David. "Freeing the Bound Variable: The Influence of Context on Reference in First-Order Derivation," *Journal of Non-Classical Logic* 3, no. 2 (1986): 37–90.
Isles, David (1992). "What Evidence is There That 2^65536 is a Natural Number?," *Notre Dame Journal of Formal Logic* 33, no. 4 (1992): 465–80.
Johnstone, Peter. *Stone Spaces* (Cambridge: Cambridge, 1982).
Jones, David. *The Anathemata: fragments of an attempted writing* (New York: Viking, 1965).

Joyce, James. *Finnegans Wake* (New York: Viking, 1967).
Kant, Immanuel. *Critique of Practical Reason*, trans. Werner S. Pluhar (Indianapolis: Hackett, 2002).
Kant, Immanuel. *Critique of Pure Reason*, trans. Werner S. Pluhar (Indianapolis: Hackett, 1996).
Kiarostami, Abbas. *10 on Ten*, included as special feature documentary with *Ten*, Zeitgeist Video Z1057 (2004).
Klein, Jacob. *Greek Mathematical Thought and the Origin of Algebra*, trans. Eva Brann (repr. New York, Dover, 1992).
Kolakowski, Leszek. *Metaphysical Horror* (Oxford: Blackwell, 1988).
Kolakowski, Leszek. *The Presence of Myth*, trans. Adam Czerniawski (Chicago: Chicago, 1990).
Lakatos, Imre. *Proofs and Refutations: The Logic of Mathematical Discovery* (Cambridge: Cambridge, 1976).
Langton, Christopher G., ed. *Artificial Life: An Overview* (Cambridge: MIT, 1995).
Lapidus, Michel. *In Search of the Riemann Zeros* (Providence: AMS, 2008).
Lavine, Shaughan. *Skolem Was Wrong*, unpublished manuscript.
Lavine, Shaughan. *Understanding the Infinite* (Cambridge: Harvard, 1994).
Lévi-Strauss, Claude. *Introduction to a Science of Mythology*, 4 vols.: *The Raw and the Cooked* (New York: Harper & Row, 1969), *From Honey to Ashes* (1973), *The Origin of Table Manners* (1978), *The Naked Man* (1981).
Lévi-Strauss, Claude. *Myth and Meaning* (New York: Schocken, 1978).
Llinás, Rodolfo R. *I of the Vortex* (Cambridge: MIT, 2001).
Lockspeiser, Edward. *Music and Painting: A Study in Comparative Ideas from Turner to Schoenberg* (New York: Harper & Row, 1973).
Malcolm, Norman. *Ludwig Wittgenstein: A Memoir*, 2nd ed. (Oxford: Oxford, 1984).
Mandelbrot, Benoit (1963). "The Variation of Certain Speculative Prices," *The Journal of Business* 36, no.4 (1963): 394–419.
Mandelstam, Nadezhda. *Hope Against Hope: A Memoir*, trans. Max Hayward (New York: Atheneum, 1970).
Mayberry, John. *The Foundations of Mathematics in the Theory of Sets* (Cambridge: Cambridge, 2000).
McLarty, Colin. "What does it take to prove Fermat's Last Theorem? Grothendieck and the Logic of Number Theory," *Bulletin of Symbolic Logic* 16 3 (2010), 359–377.
Mulhall, Stephen. *On Film* (London: Routledge, 2002).

Nelson, Edward. "Mathematical Mythologies," in Jean-Michel Salanskis and Hourya Sinaceur, eds., *Le Labyrinthe du Continu* (Paris: Springer-Verlag France, 1992), 155–67.

Nelson, Edward. "Warning Signs of a Possible Collapse of Contemporary Mathematics," in Michael Heller and W. Hugh Woodin, eds., *Infinity: New Research Frontiers* (Cambridge: Cambridge, 2011), 76–85.

Nicholls, Angus. *Myth and the Human Sciences: Hans Blumenberg's Theory of Myth* (New York: Routledge, 2015).

de Oliveira, Manoel. *The Strange Case of Angelica* (*O Estranho Caso de Angélica*), film (2010).

Omnès, Roland. *Understanding Quantum Mechanics* (Princeton: Princeton, 1999).

Preston, Richard. "Profiles: The Mountain of PI," *The New Yorker* (2 March 1992), 36–67.

Prince. *Purple Rain*. DVD Warner Brothers 9 25110–2 (1984).

Putnam, Hilary. "Brains in a vat," in Hilary Putnam, *Reason, Truth and History* (Cambridge: Cambridge, 1981), 1–21.

Putnam, Hilary. "Models and Reality," in Hilary Putnam, *Realism and Reason: Philosophical Papers, Volume 3* (Cambridge: Cambridge, 1983), 1–25.

Pym, John, ed. *Time Out Film Guide 2011* (London: Time Out Guides Limited, 2010).

Reeder, Roberta. *Anna Akhmatova: Poet and Prophet* (New York: Saint Martin's, 1994).

Rorty, Richard. *Achieving Our Country* (Cambridge: Harvard, 1998).

Rubin, William, ed. *Picasso and Portraiture: Representation and Transformation* (New York: Museum of Modern Art, 1996).

Scheler, Max. *Formalism in Ethics and a Non-Formal Ethics of Values*, trans. Manfred S. Frings and Roger L. Funk (Evanston: Northwestern, 1973).

Schmidt-Biggemann, Wilhelm. *Philosophia perennis: Historical Outlines of Western Spirituality in Ancient, Medieval and Early Modern Thought* (Dordrecht: Springer, 2004).

Schoenberg, Arnold. *Style and Idea: Selected Writings of Arnold Schoenberg*, trans. Leo Black, ed. Leonard Stein (Berkeley: California, 1975).

Schweber, Silvan S. *QED and the men who made it: Dyson, Feynman, Schwinger, and Tomonaga* (Princeton: Princeton, 1994).

Schweitzer, Viktoria. *Tsvetaeva*, trans. Robert Chandler and H. T. Willetts, poetry trans. Peter Norman, ed. Angela Livingstone (New York: Farrar, Straus and Giroux, 1992).

Shanker, S. G. *Wittgenstein and the Turning-Point in the Philosophy of Mathematics* (Albany: SUNY, 1987).
Shaw, Mary Louise. *Performance in the Texts of Mallarmé: The Passage from Art to Ritual* (University Park: Pennsylvania State, 1993).
Simmons, H. *Derivation and Computation: Taking the Curry-Howard correspondence seriously* (Cambridge: Cambridge, 2000).
Smolin, Lee. *The Life of the Cosmos* (New York: Oxford, 1997).
Stern, David. *Wittgenstein on Mind and Language* (New York: Oxford, 1995).
Strand, Mark. *Selected Poems* (New York: Atheneum, 1980).
Strindberg, August. *Six Plays of Strindberg*, trans. Elizabeth Sprigge (Garden City: Doubleday, 1955).
Sulloway, Frank. *Born to Rebel: Birth Order, Family Dynamics, and Creative Lives* (New York: Pantheon, 1996).
Tait, W. W. "Finitism," *Journal of Philosophy* **78** (1981) 524–547, errata 618.
van Atten, Mark. *On Brouwer* (Toronto: Wadsworth, 2004).
van Gennep, Arnold. *The Rites of Passage*, trans. Monika B Vizedom and Gabrielle L. Caffee (London: Routlege, 1960).
Vuillemin, Jules. *Le Dieu d'Anselme et les Apparences de la Raison* (Paris: Aubier, 1971).
Wachowski Brothers. *The Ultimate Matrix Collection*, dir. Josh Oreck, prod. Eric Mattheis 10 DVD's + booklet (Burbank: Warner Brothers, 2004).
Watson, Ian. "The Matrix as Simulacrum," in Haber, 2003, 149–67.
Weatherson, Brian. "Are You a Sim?" *Philosophical Quarterly*, 53 (2003): 425–431.
West, Cornel. *The American Evasion of Philosophy* (Madison: Wisconsin, 1989).
Wind, Edgar. *Art and Anarchy* (London: Faber and Faber, 1963).
Wittgenstein, Ludwig. *On Certainty*, ed. G.E.M. Anscombe and G.H. von Wright, trans. Denis Paul and G.E.M. Anscombe (New York: Harper, 1972).
Wittgenstein, Ludwig. *Philosophical Investigations*, 3rd ed., trans. G.E.M. Anscombe (Oxford: Blackwell, 2001).
Wittgenstein, Ludwig. *Tractatus Logico-Philosophicus*, trans. D. F. Pears and B. F. McGuinness (London: Routledge, 1974).
Wittgenstein, Ludwig. *Wiener Ausgabe: Studien Texte*, Band 1–5, hrsg. Michael Nedo (Austria: Zweitausendeins, 1994–96).
Wittgenstein, Ludwig. *Wittgenstein's Lectures on the Foundations of Mathematics, Cambridge, 1939*, from the notes of R. G. Bosanquet, Norman Malcolm, Rush Rhees, and Yorick Smythies, ed. Cora Diamond (Ithaca: Cornell, 1976).

Witzel, E. J. Michael (2012). *The Origins of the World's Mythologies* (Oxford: Oxford, 2012).

Yessenin-Volpin, A. S. "The Ultra-Intuitionistic Criticism and the Anti-Traditional Program for the Foundations of Mathematics," in J. Myhill, A. Kino, and R. E. Vesley, eds., *Intuitionism and Proof Theory* (Amsterdam: North-Holland, 1970), 3–45.

Index[1]

[1] Note: Page numbers followed by 'n' denote notes

O.B. Bassler, *Diagnosing Contemporary Philosophy with the Matrix Movies*, DOI 10.1057/978-1-137-57889-1