

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Metaphysical Realism and Antirealism:
An Analysis of the Contemporary Debate

Deborah Colleen Smith

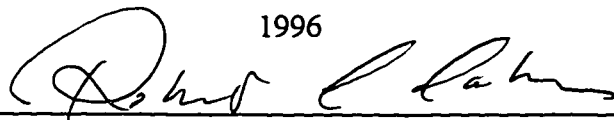
A Dissertation submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

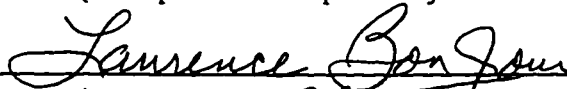
University of Washington

1996

Approved by



(Chairperson of Supervisory Committee)



Program authorized

to Offer Degree Department of Philosophy

Date November 4, 1996

UMI Number: 9716921

**Copyright 1996 by
Smith, Deborah Colleen**

All rights reserved.

**UMI Microform 9716921
Copyright 1997, by UMI Company. All rights reserved.**

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

© Copyright 1996

Deborah Colleen Smith

Doctoral Dissertation

In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at the University of Washington, I agree that the Library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted "the right to reproduce and sell (a) copies of the manuscript in microform and/or (b) printed copies of the manuscript made from microform."

Signature

Date

Deborah Smith

11/25/96

University of Washington

Abstract

**Metaphysical Realism and Antirealism:
An Analysis of the Contemporary Debate**

by Deborah Colleen Smith

Chairperson of the Supervisory Committee: Professor Robert Coburn
Department of Philosophy

The metaphysical realist asserts, while the metaphysical antirealist denies, that there are individuals that exist independently of the existence and workings of any mind or minds. I begin by distinguishing the thesis of metaphysical realism from other theses that are also called 'realism'. Of particular interest in this discussion is the relation between metaphysical realism and views such as moral realism and scientific realism.

Metaphysical realism is commonly thought to be the default position in the debate since it is *prima facie* supported by our commonsense intuitions concerning the nature of reality. I develop several arguments to show that, when we examine our intuitions in detail, metaphysical realism is not the entirely commonsensical view that it at first appears. I argue that if metaphysical realism is true, it is difficult to see how semantic and epistemic access to the mind-independent world is possible and that, even if we do have such access, it will be difficult to rule out very bizarre ontologies. While none of these arguments constitute a decisive objection to metaphysical realism, they do suggest that the metaphysical realist should say something in defense of his view.

I then examine Peter van Inwagen's attempted defense of metaphysical realism. This defense takes the form of a series of arguments designed to show that metaphysical antirealism is incoherent because incompatible with the existence of objective truth. I argue that his criticisms do not constitute an adequate defense of metaphysical realism because certain forms of antirealism, if otherwise philosophically viable, plausibly escape.

This leads to a discussion of the extent to which an antirealist metaphysic can be considered a viable philosophical view. In this context, I examine the views of Hilary Putnam and Paul Grice. While the versions of metaphysical antirealism presented face serious difficulties of their own, I argue that it is a matter of philosophical judgment which view, metaphysical realism or metaphysical antirealism, has the more severe problems.

TABLE OF CONTENTS

	<i>Page</i>
Chapter One: Metaphysical Realism Defined and Distinguished from Other Views called 'Realism'	1
Chapter Two: A Problem for Metaphysical Realism: Skepticism Concerning our Knowledge of the Mind-Independent World	30
Chapter Three: A Problem for Metaphysical Realism: The Possibility of a Very Crowded Ontology	57
Chapter Four: A Problem for Metaphysical Realism: Putnam's Argument Against Semantic Realism	98
Chapter Five: Metaphysical Antirealism and Objective Truth: Is Metaphysical Antirealism Self-Refuting?	147
Chapter Six: How to be a Metaphysical Antirealist?	185
Bibliography	242
Appendix: Some Heroic Attempts to Avoid Skepticism	249

ACKNOWLEDGMENTS

The author wishes to express sincere appreciation to Professor Robert Coburn for his assistance in the preparation of this manuscript. In addition, special thanks to Duane and Shirley Smith for their encouragement and support and to Frederic Pierre for his patience and understanding during the writing of this dissertation and the whole of my graduate career.

CHAPTER ONE: METAPHYSICAL REALISM DEFINED AND DISTINGUISHED FROM OTHER VIEWS CALLED 'REALISM'

Intuitively, we believe various things and kinds of things to exist. We believe that there are rocks and trees and people and chairs and perhaps even electrons. We also believe that there are things that do not exist, such as goblins, unicorns, and golden mountains. There are still other things whose existence is more controversial: extra-terrestrial visitors, gravitons, God. Ontology is the branch of metaphysics that studies existence; it asks what individuals and kinds of individuals exist and what it means to say of something that it exists. This dissertation will focus primarily on the second of these two issues, examining in detail various elements of one contemporary philosophical debate concerning the nature of existence: the debate between metaphysical realists and antirealists. In particular, I will examine what is at issue between the two positions in this debate, evaluate some of the traditional defenses of each position, and suggest some new and more fruitful directions the debate might take. In the sections of this chapter that follow, I will first formulate and explicate the position of the metaphysical realist. Second, I will explore a variety of ways of denying metaphysical realism as formulated, identifying the view for which I reserve the term 'metaphysical antirealism' and with which I will be primarily concerned. Third, I will distinguish what is at issue in the debate between metaphysical realists and antirealists in my usage from what is at issue in other debates whose opponents have been called 'realists' and 'antirealists'. Fourth, I will explore in detail the commonsense intuitions that are thought to support metaphysical realism and which appear to give the realist the default position in the debate. This chapter will conclude with an overview of the entire dissertation.

1. The metaphysical realist claims that the various individuals that exist, do so and have (at least some of) the features¹ that they have, independently of the existence of any mind or minds and the collective activities of any group of minds.² That is, the existence and (at least some of the) features of the objects that there are are independent of not only *our* social and linguistic conventions, *our* ways of perceiving and conceiving the world, *our* interests and wishes, *our* hopes and dreams, etc., but of the social and linguistic conventions, interests, etc., of any other mind or group of minds as well. It is clear that, on the face of it, the metaphysical realist's thesis is quite plausible, highly intuitive, and very much in line with our commonsense view of the world. After all, we believe that

¹ I want to construe 'features' very broadly. It is intended to pick out not only the sortal properties, but any non-sortal properties, if any there be, of an individual. Further, if it is metaphysically possible for there to be bare particulars, the very bareness of the individual should also come under the heading of a 'feature' of the individual as I will use the term.

² This formulation is somewhat problematic in that the reader may be led to infer that traditional theism is not a form of metaphysical realism. After all, if traditional theism is true, that is, if there is a God in the sense of an intelligent being who, in virtue of being the creator of the world, is responsible for the existence and features of all other individuals, then those individuals will exist and have (at least many of) the features that they do in virtue of some facts about the mind of God. They are mind-dependent in the sense that they would not exist and have (at least many of) the features that they do if a certain mind had not existed. However, the metaphysical realist does not deny that minds can causally affect the world. Even finite minds such as ours are capable of bringing things into existence. But, once they are brought into existence, the continued existence of the objects we create is, according to the metaphysical realist, independent of what we or anyone else thinks about them. Similarly, the sense in which God's creation is dependent upon the mind of God may not be a sense of mind-dependence which the metaphysical realist denies. Even so, there may still be those who worry that the traditional doctrine that God "sustains" the world would seem to entail that none of God's creation is metaphysically real given this formulation of metaphysical realism. As I do not intend this result, I will take the thesis of metaphysical realism in such a way that the mind of God (if there is one) is exempted from the minds that I am talking about when I speak of mind-dependence and mind-independence.

there are entities in the Gobi desert or on the dark side of the moon that exist and have various features even when no one is perceiving them or thinking about them, even if no one has ever perceived or thought about them, and even if no one ever does perceive or think about them. And, we believe all this not only as regards human minds, but as regards any minds that exist. Thus interpreted, the claim of the metaphysical realist may seem to some to amount to little more than the rejection of Berkeleyan idealism, the thesis that everything that exists is either a mind or a modification of a mind.³ However, while most contemporary metaphysical realists reject Berkeleyan idealism, the denial of Berkeleyan idealism will be seen to be neither necessary nor sufficient for metaphysical realism when the realist's intended sense of 'mind-independence' is more fully explained.

What exactly does the metaphysical realist mean when he says that something exists mind-independently? One way to get at the intended sense of 'mind-independence' is by considering a story similar to the one that Descartes entertains in the First Meditation. Consider a world inhabited entirely by nonmaterial Cartesian souls who believe that they live in a world very much like our own. They experience the world as containing many kinds of enduring material beings: tables, chairs, lumps of gold, lilies, human beings, etc. They have also developed several well-confirmed scientific theories to explain the behavior of the material beings they believe themselves to experience. These

³ See Berkeley, *A Treatise Concerning the Principles of Human Knowledge* (London: 1910) and *Three Dialogues between Hylas and Philonous* (London: 1713). Actually, Berkeley called his view 'immaterialism' rather than 'idealism', since it amounted to a denial that there were material substances.

scientific theories postulate still other kinds of physical entities: strands of DNA, black holes, quarks, electrons, etc. All of these entities (and perhaps some non-material entities like numbers and sets) are taken to exist and have definite features by the Cartesian souls. If we suppose that these souls further believe that all, or at least most, of the entities of their commonsense world view and best scientific theories exist and have the features that they seem to have independently of minds and the workings of minds, then they are sadly mistaken about the content of their world. The only things that exist in the metaphysical realist's sense in this world are the Cartesian souls themselves,⁴ for only they exist and have (for the most part) the features that they do independently of the workings of minds. All of the other entities ostensibly experienced or postulated by these souls (including their own bodies) exist only as sensations and ideas in their minds.

It is interesting to note at this point that since Cartesian souls are essentially minds, this is a world in which idealism, at least one version of it, is true: everything that exists is either a mind or a modification of a mind. However, that *this* is how things are is, in the metaphysical realist's sense, *independent* of the mental workings of the Cartesian souls. The souls do not think themselves into existence; they have ontological status independently of their own perceptions and theorizing. If this story represents a genuine metaphysical possibility, then we can see that the rejection of idealism is not necessary for metaphysical realism. Indeed, Berkeleyan idealism⁵ would seem to entail the claim that

⁴ Unfortunately, the Cartesian souls have had no reason to postulate the existence of Cartesian souls given the success of their neuro-physical theories.

⁵ Or, anyhow, Berkeley's idealism minus Berkeley's God.

the existence (and many of the features) of minds is mind-independent in the realist's sense. So, if Berkeleyan idealism is a metaphysically possible view, it is possible to be both a metaphysical realist and a Berkeleyan idealist. On the other hand, if Berkeleyan idealism entails metaphysical realism, someone who denied metaphysical realism (the so-called 'antirealist') would necessarily be committed to a rejection of Berkeleyan idealism as well. Further, it is clear that several philosophers who have held some version of metaphysical antirealism have also rejected Berkeleyan idealism.⁶ So, a rejection of this type of idealism will not be sufficient for metaphysical realism either. The fact that the rejection of Berkeleyan idealism is neither necessary nor sufficient for metaphysical realism (at least in its most general form) is important to note since Berkeleyan idealism has traditionally been held to amount to a denial of another metaphysical view often called 'realism'.⁷ To be sure, the Berkeleyan idealist does deny that the objects that we commonsensically take to be material objects exist independently of the mind. However, he need not deny that there is anything at all that has mind-independent existence.

To return to our explication of the metaphysical realist's sense of 'mind-independence', consider another possible world inhabited by normal human beings, who believe that the entities of their commonsense world view and those postulated by their best scientific theories exist and have (at least many of) the features that they do in the

⁶ See, Hilary Putnam, *Reason, Truth, and History* (Cambridge: Cambridge University Press, 1981), p. xi and pp. 58-9.

⁷ See, Paul Edwards ed., *The Encyclopedia of Philosophy*, vol. 4 (New York: Macmillan, 1967), pp. 110-111. Here, 'realism' is defined as the view that "material objects exist independently of being perceived."

metaphysical realist's sense. Suppose that in a great many instances their beliefs about the world are correct. That is, suppose that metaphysical realism is true in this world and that many of the entities of their commonsense world view as well as many of those postulated by their best scientific theories exist and have (at least many of) the features that they do independently of any minds. Intuitively, what makes their beliefs about the existence and (at least many of the) features of these entities true when they are true is (at least in part) that there is a similarity or other correspondence relation between the world as they experience it and describe it in their scientific theories, and the world as it *really* is, independently of the workings of minds.⁸ In most cases, when an individual from this world has the experience as of a tree in front of him, it is (at least in part) because there really is, independent of the collective activity of minds, a tree in front of him. When an individual from the Cartesian soul world ostensibly experiences a tree, it is always entirely because his mind is modified in a certain way, and not because there really is, independently of the activity of minds, a tree in front of him. In the one world, there are metaphysically real trees, and rocks, and perhaps even electrons; in the other world only

⁸ It is the role of a theory of truth and reference to attempt to specify necessary and sufficient conditions for what is to count as the appropriate correspondence between our beliefs or experiences and the world, such that our beliefs and experiences can be said to be *true of* the world. As we will see in Chapter Four below, the mere fact that there is a correspondence relation which holds between the metaphysically real world, and the world as experienced is probably not sufficient for commonsense and scientific beliefs to count as knowledge *of* the metaphysically real world. For such beliefs to count as knowledge of the metaphysically real world, the metaphysically real world would have to be the *cause* of their experiencing the world the way they do, and of the success of their scientific theorizing.

Cartesian souls are metaphysically real. But, in both worlds there are entities that exist and have (at least many of) the features that they do independently of the workings of minds. It is this sense of ‘mind-independence’ that is the hallmark of the thesis of metaphysical realism.

2. Although more remains to be said, the reader should by now have a fairly clear grasp of what the thesis of metaphysical realism amounts to. What the reader may wonder now is why anyone would have any doubts about such a plausible-sounding thesis as metaphysical realism. This is a question to which I will turn in chapters two through four of the dissertation. I want here to address a slightly different question: roughly, what is it to deny metaphysical realism? Given the above account of the metaphysical realist’s thesis, there are a number of different views that entail the falsity of metaphysical realism. One way of denying metaphysical realism would be to deny that anything exists at all. Obviously if nothing whatever exists, nothing exists mind-independently. Another way of denying metaphysical realism would be to hold that, although there are things that exist, neither their existence nor any of their features are independent of the existence or operations of any mind or minds.⁹ It is this second way of denying metaphysical realism that is of interest to this dissertation; I shall reserve the term ‘metaphysical antirealism’ for the position of those who deny metaphysical realism in this way. However, in the next

⁹ It is important to note that if this view is to genuinely differ from Berkeleyan idealism, the existence (and many of the features) of minds themselves must be mind-independent.

paragraph I shall briefly examine the first way of denying metaphysical realism. My reason for doing so is that it seems to me that these two ways of denying metaphysical realism are occasionally conflated by proponents of metaphysical realism. And, once this confusion has occurred, it is easy to see why metaphysical antirealism might be thought obviously unattractive. The discussion in the paragraphs that follow is intended to demonstrate to the reader that such a confusion is just that and to facilitate the critical evaluation of the debate between metaphysical realists and metaphysical antirealists that will comprise the remainder of the dissertation.

Let us consider the claim that metaphysical realism is false because nothing whatever exists, a view which might be called 'nihilism'.¹⁰ Now, many have thought that there is at least (exactly?) one metaphysically possible world that is completely empty. Even if this is true, however, to claim that the *actual* world is in fact an empty world, is obviously false! Nihilism, taken as a thesis about the nature of the actual world, is what is sometimes called a self-refuting supposition, since, that anyone should suppose the nihilist's thesis to be true entails its falsity. In other words, if one can claim or believe that nothing exists, then something (beliefs and believers, claims and claimers) must exist. It may be true to say that there are possible worlds in which metaphysical realism is false because nothing whatever exists. However, none of these will be the actual world, as the

¹⁰ This view must be distinguished from another view also sometimes called 'nihilism' which claims that the only things that exist are metaphysical simples. For such a use of the term 'nihilism', see Peter van Inwagen, *Material Beings* (Ithaca: Cornell University Press, 1990), pp. 72-3.

actual world is one in which one can claim (though falsely) that nothing whatever exists. If metaphysical realism is false of the actual world, there must be some reason other than the truth of nihilism why it is false.

So, the first way of denying metaphysical realism (the adoption of nihilism) is not even a genuine contender. Any real threat to metaphysical realism must come from some other way of denying that thesis, such as the adoption of what I am calling 'metaphysical antirealism'. (More will need to be said, of course, if we are to assess the extent to which this view is a threat to metaphysical realism.) If metaphysical antirealism is true, it is not because there are no entities that can truly be said to exist, but rather because all of the things that can be truly said to exist are in some nontrivial sense dependent on the existence and activities of some mind or minds for their existence and features. Although, in this chapter, I will say little else to fill out the different forms that metaphysical antirealism can take (that task is primarily reserved for the third part of the second chapter and the bulk of the final chapter), it should be clear that the basic dispute between the metaphysical realist and antirealist concerns the ontological status of the entities and of (at least many of) their features that make up the world and not over which entities may be said to exist. To assume that metaphysical antirealism simply collapses into nihilism is to assume from the outset that the notion of the existence of an object presupposes that the object exists independently of any mind or minds. But to assume from the outset that an object that is in any way dependent on the existence and activities of any mind or group of minds would have no genuine ontological status at all is to beg the question against the

metaphysical antirealist. Indeed, it is intuitive to suppose that there would be a genuine metaphysical difference between a world that contains only mind-dependent entities and the empty world, at least on the assumption that both are metaphysically possible worlds. The two ways of denying metaphysical realism that I have distinguished should not be conflated. Both the metaphysical realist and the metaphysical antirealist might agree that, e.g., dogs and cats exist; their disagreement would be over the metaphysical import of the existence of such entities. In the analysis of the debate that follows, I will provisionally assume that both metaphysical realism and antirealism in this broad sense are philosophically viable theses.¹¹

3. Although this may be clear from what has been said above, I believe that it is important to explicitly distinguish between what is at issue in the metaphysical realism/antirealism debate and what is at issue in other debates concerning views which have traditionally been called ‘realism’. We have already seen that the debate between metaphysical realists and antirealists does not merely collapse into the debate between the so-called realists and idealists; nor does it collapse into a debate between metaphysical realists and nihilists. But, there is yet another type of view that has commonly been called ‘realism’ and which we must take care to distinguish from metaphysical realism. This is

¹¹ There will, no doubt, be those who have worries at this point about the intelligibility of metaphysical antirealism. Such worries will be raised and addressed in chapters five and six. For now, I ask the reader to grant me the provisional assumption that there is at least one philosophically viable version of metaphysical antirealism.

the type of realism at issue in debates concerning, for example, moral realism and scientific realism. So it is to a discussion of the relation between realism in this sense and metaphysical realism that I will now turn.

As I mentioned at the beginning of this chapter, we intuitively believe that there are (at some time or other) instances of certain kinds—people, chairs, lumps of gold, and possibly numbers—but that there are no instances (at any time) of various other kinds—witches, unicorns, or even prime numbers greater than 2. Thus, relative to any sortal term Φ , there is the question as to whether or not there are, at any time, individual Φ s. The non-existence of objects of some kinds, red-headed-Eskimos for example, will be a purely contingent matter. The non-existence of objects of other kinds, four-sided-triangles for example, will be a matter of necessity. Anyone who asserts, for some sortal term Φ , that there are (at some time) Φ s is commonly said to be a realist about Φ s, or more simply, a Φ -realist.¹² Anyone who denies that there are (at any time) Φ s or who denies that we are warranted in asserting that there are Φ s is said to be a Φ -antirealist, or an irrealist about Φ s. Although there is much agreement among philosophers concerning many of the kinds of things that exist (most philosophers believe that there are people, trees, rocks, and planets¹³ and deny that there are unicorns or demons), the debates between Φ -realists and

¹² Obviously, Φ -realism for any sortal term Φ will be true only if there actually are Φ s whether or not anyone asserts that there are.

¹³ Although, see van Inwagen, *Material Beings* for a notable exception to Φ -realism concerning rocks and planets.

Φ -antirealists with respect to the entities postulated by science, as well as moral properties or facts, are still very much alive.

Now, it seems to me that the debate between metaphysical realists and antirealists and debates between Φ -realists and Φ -antirealists are debates concerning the answers to distinct questions; they are differently focused and have different scope. While various attitudes with respect to the debate concerning metaphysical realism and antirealism may have an influence on one's attitudes with respect to any particular debate concerning Φ -realism and Φ -antirealism (and vice versa), the two types of debates are, to a large extent, independent. Provided that both metaphysical realism and antirealism are intelligible views, both the metaphysical realist and antirealist can consistently hold either a Φ -realism or Φ -antirealism for any given sortal term Φ . Although this claim may strike some as obviously correct, it is clear that there have been many philosophers who have been inclined to view the truth of metaphysical realism as a necessary condition for the truth of Φ -realism with respect to all, or at least many important, sortal terms Φ .¹⁴ For this reason, it will be important to say a bit more in defense of the claim that the truth of

¹⁴ See, for example, Richard Boyd, "On the Current Status of Scientific Realism," in *The Philosophy of Science*, Richard Boyd, Philip Gasper, and J. D. Trout, eds. (Cambridge: MIT Press, 1991), esp. p. 195, and Boyd, "How to be a Moral Realist" in *Essays on Moral Realism*, Sayre-Mc Cord ed. (Ithaca: Cornell University Press, 1988), esp. pp. 181-2. "Scientific realism is the doctrine that scientific theories should be understood as putative descriptions of real phenomena...and that the reality described by scientific theories is largely independent of our theorizing. Scientific theories describe reality and reality is 'prior to thought'. By 'moral realism' I intend the analogous doctrine about moral judgments, moral statements and moral theories. According to moral realism...[t]he truth or falsity...of moral statements is largely independent of our moral opinions, theories, etc."

metaphysical realism is not a necessary condition for the truth of Φ -realism for any particular sortal term Φ .

Recall that Φ -realism amounts to the claim that there are individual Φ s. So, if the truth of metaphysical realism were a necessary condition for Φ -realism's being true with respect to every (or any) sortal term Φ , the metaphysical antirealist would seem to be committed to claiming that for every sortal term Φ , there are in fact no Φ s. If this were the case, metaphysical antirealism would seem to simply collapse into nihilism (provided, of course, that there are no particulars that instantiate no kind properties whatever). But, as we saw above, there is no reason to suppose that metaphysical antirealism collapses into nihilism. The claim of the metaphysical antirealist is instead that the things that exist are mind-dependent, or at least can't be said to be completely mind-independent. This would seem to imply that there are, on such a view, things that have genuine ontological status, things that exist. Given that there are things that exist, and given that all particulars are of some kind or other, there is no reason to suppose that all kind concepts are empty, i.e., that Φ -realism fails for every sortal term Φ . In short, the truth of metaphysical realism is not a necessary condition for the truth of all versions of Φ -realism because metaphysical antirealism is perfectly consistent with at least some forms of Φ -realism.

Although many will concede that the truth of metaphysical realism is not a necessary condition for the truth of all forms of Φ -realism, it may yet be thought that with respect to the more contentious debates concerning moral realism and scientific realism, the Φ -realist is of necessity a metaphysical realist. However, even this does not seem to be

correct. First, consider the debate between moral realists and antirealists. The moral realist is someone who claims that there are moral properties or moral facts and that it is the existence of these moral properties or facts that makes moral statements (for example, the statement that the torture of babies purely for entertainment purposes is wrong) true or false. Broadly speaking, there are two main ways to be a moral antirealist. The first is to deny that what appear to be moral statements have any genuine propositional or cognitive content, deny that they are either true or false. This view is held by the moral noncognitivists: the emotivists,¹⁵ prescriptivists,¹⁶ and others. These philosophers all hold that ethical statements do not purport to refer to or describe facts in the world. For example, emotivists claim, roughly, that ethical statements express approval or disapproval of various actions; prescriptivists claim, roughly, that ethical statements express recommendations or prescriptions. Both views express moral antirealism rather than realism. Another way to be a moral antirealist is to hold what is often called an error theory with respect to ethical statements. An error theory is one which, while agreeing with the moral realist that moral statements have a truth value, holds that the truth value that all moral statements possess is falsehood. In the first chapter of *Ethics: Inventing Right and Wrong*,¹⁷ J. L. Mackie argues that ethical statements *do* purport to refer to objective moral values or properties, but that there are no such moral properties; all moral statements are, in fact, false. One of Mackie's arguments for denying the existence of

¹⁵ See C. L. Stevenson, *Ethics and Language* (New Haven: Yale, 1944).

¹⁶ See R. M. Hare, *The Language of Morals* (New York: Oxford University Press, 1952).

¹⁷ J. L. Mackie, *Ethics: Inventing Right and Wrong* (London: Penguin Books, 1977).

objective moral values and adopting moral antirealism (or, as he calls it, moral skepticism) turns on his claim that such essentially normative entities would have to be very strange entities indeed and that knowledge of them would require “some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else.”¹⁸ Now, it seems quite clear that if we were suddenly confronted with a decisive *a priori* argument in favor of metaphysical antirealism (perhaps we have all been persuaded by Michael Dummett’s argument that various conditions on an acceptable theory of meaning will entail the adoption of an antirealist metaphysic¹⁹), this would in no way settle *without further argumentation* the questions concerning whether moral statements have cognitive content and purport to refer to objective moral values (or perhaps properties or facts), and concerning whether or not such objective moral values (properties or facts) would be too queer to be accorded genuine ontological status. The debate between moral realists and antirealists would still be very much alive. This suggests that the truth of metaphysical realism is not a necessary condition for the truth of moral realism.

But what of scientific realism, the view that the theoretical entities that would be postulated by an ideal science exist? Won’t *scientific* realism, so understood, necessarily entail metaphysical realism? Again I think that the answer is, perhaps surprisingly, ‘no’. One prominent proponent of scientific realism, Richard Boyd, defines scientific realism as the view that:

¹⁸ Ibid., pp. 38.

¹⁹ Michael Dummett, *The Logical Basis of Metaphysics* (Cambridge: Harvard University Press, 1991).

(i) “Theoretical terms” in scientific theories (i.e., nonobservational terms) should be thought of as putatively referring expressions; scientific theories should be interpreted “realistically”

(ii) Scientific theories, interpreted realistically, are confirmable *and in fact often confirmed* as approximately true by ordinary scientific evidence interpreted in accordance with ordinary methodological standards.

(iii) The historical progress of mature sciences is largely a matter of successively more accurate approximations to the truth about both observable and unobservable phenomena. Later theories typically build upon the (observational and theoretical) knowledge embodied in previous theories.

(iv) The reality which scientific theories describe is largely independent of our thoughts or theoretical commitments.²⁰

The reader will notice a parallel between the criteria for moral realism discussed above and what Boyd sees as the first two elements of scientific realism. First of all, to count as a form of scientific realism, according to Boyd, a view must interpret scientific theories as having cognitive content and sentences of those theories containing theoretical terms, e.g., ‘electron’, should be construed as genuinely referring to unobservable entities. This is what it is, according to Boyd, to interpret a scientific theory “realistically.” Anyone who interprets a scientific theory in such a way that it does not make genuine reference to unobservable entities will be a scientific antirealist. This is precisely the line taken by the so-called operationalists who held that sentences containing theoretical terms such as ‘electron’ could be cashed out entirely in terms of sentences making reference to only observable entities and states of affairs.²¹ While operationalists are not themselves non-cognitivists about scientific theory (sentences containing theoretical terms do, according

²⁰ “On the Current Status of Scientific Realism,” p. 195.

²¹ See P. W. Bridgman, *The Logic of Modern Physics* (New York: Macmillan, 1927).

to the operationalist, have cognitive content), they do hold that scientific theories realistically construed are cognitively meaningless. But, a scientific realist not only interprets scientific theories realistically, he also holds that when scientific theories are interpreted realistically, at least some of them are true (or at least approximately true). That is, a scientific realist rejects an error-theory with respect to scientific theories realistically construed. So, just as metaphysical realism is not a necessary condition for moral realism, metaphysical realism is not a necessary condition for the acceptance of the first two of the theses which, according to Boyd, constitute scientific realism.

The third of Boyd's central theses of scientific realism concerns the possibility of scientific knowledge of both observable and unobservable entities and phenomena and the extent to which current scientific theories can be said to provide us with such knowledge. Although the existence of various epistemological worries concerning the subject matter of science has served to motivate the development of many forms of scientific antirealism, it seems to me to be a mistake to build a denial of skepticism into the criteria for scientific realism. It seems to me to be perfectly consistent to suppose that the world contains all or many of the entities postulated by scientific theories but that we do not stand to them in any of the relations necessary for epistemic justification. But, in any case, if metaphysical antirealism is a viable thesis, I see no reason why the metaphysical antirealist would be

committed to denying that we can have knowledge of unobservable entities and phenomena.²²

It is the fourth of Boyd's theses of scientific realism that looks as if it entails metaphysical realism.²³ However, when we look more closely at Boyd's discussion of this thesis, it seems to me that what he wants to say concerning the world described by science would be compatible with the adoption of at least some versions of metaphysical antirealism. In his introductory essay to the chapter entitled "Confirmation, Semantics, and the Interpretation of Scientific theories,"²⁴ Boyd writes:

The component of scientific realism that emphasizes the independence of the reality described by scientific theories from our theorizing is significant because it is with respect to the claimed theory-independence of the reality studied by scientists that realist and social constructivists disagree.²⁵

Social constructivists, according to Boyd, hold that the world is "largely defined or constituted by the theoretical tradition which defines [scientific] methodology."²⁶ That is, the social constructivist holds that, in constructing scientific theories, we construct the theoretical entities referred to therein. The existence and features of theoretical entities is,

²² Many of the proponents of metaphysical antirealism including Hilary Putnam and Michael Dummett would acknowledge that scientific knowledge is both possible and actual. The relation between metaphysical realism and various forms of skepticism will be discussed at length in Chapter Two below.

²³ Indeed, it is quite likely that Boyd himself intends this thesis to entail metaphysical realism. See footnote 14 on p. 12 above.

²⁴ In *The Philosophy of Science*, pp. 3-35. For a version of social constructivism see Thomas S. Kuhn, *The Structure of Scientific Revolutions* (Chicago: University of Chicago Press, 1962).

²⁵ *The Philosophy of Science*, p. 11.

²⁶ "On the Current Status of Scientific Realism," in *ibid.*, p. 207.

thus, dependent upon our construction of scientific theories according to the social constructivist.²⁷

According to the social constructivist, this connection between the world described by science and the theories which do the describing is so tight that when scientific revolutions occur and the paradigms of scientific methodology changes, scientists before and after the revolution cannot be said to be talking about the same entities.

[The social constructivist] takes the fundamental laws containing a theoretical term to constitute its conventional definition....What characterizes scientific revolutions is that fundamental laws are revised; but revisions of those laws amount to changes in the definitions of the fundamental theoretical terms in science. Scientists before and after a scientific revolution are not talking about the same theoretical entities even when they use the same terms. Thus they are really talking past each other in their disputes and no rational method can effect a choice between their competing conceptions.²⁸

What makes social constructivism objectionable from a scientific realist's point of view is partially the fact that it rules out our being able to make sense of scientific progress.²⁹ If scientists at an earlier stage of scientific theorizing were not talking about the same entities and phenomena that current scientists are, then it is nonsense to suggest that our current scientific theories are improvements on older theories. This fact again ties in with my claim that the main dispute between Φ -realists and antirealists concerns the existence of

²⁷ While this may sound on the face of it like the central thesis of metaphysical antirealism, it is important to keep in mind that a social constructivist need not be a metaphysical antirealist if he holds that the existence of observable entities is independent of the existence and activities of any mind or minds.

²⁸ "Introductory Essay" in *The Philosophy of Science*, p. 13.

²⁹ Boyd suggests that he has something similar in mind on p. 208 of "On the Current Status of Scientific Realism."

Φs. In being dependent upon our scientific theorizing, unobservable entities and phenomena would seem to be very different from observable ones. If unobservable entities such as electrons do not exist *in the same way* as observable ones such as trees, it is not hard to see why such a view should be categorized as a version of scientific antirealism. If electrons do not exist in the same way as more paradigmatic existents such as trees, there is a genuine worry that they cannot be literally said to exist. For these reasons, it seems to me to be plausible to read Boyd's fourth thesis of scientific realism as the requirement that the world described by scientists be relatively stable, that it not change as revolutions in scientific theorizing and methodology occur, and that the existence of unobservable entities be independent of our thoughts or theoretical commitments *to the extent that* and in the same way as is the existence of observable entities. Construed in this way, there is no reason to hold that a metaphysical antirealist could not accept Boyd's fourth thesis of scientific realism.³⁰

Once again, it seems clear that if we had a decisive *a priori* argument that metaphysical antirealism were true, this would not in and of itself entail that there are no theoretical entities or (approximately) true scientific laws, or that science does not aim at theories which constitute more and more accurate approximations of the truth, or that the theoretical entities described by science have different ontological status from the observable entities referred to in our commonsense view of the world. I submit that the

³⁰ Indeed, it is clear that the version of metaphysical antirealism developed and defended by Hilary Putnam in *Reason, Truth, and History* is intended to be consistent with the thesis so construed and entails the rejection of social constructivism. See pp. 113-119.

truth of metaphysical realism is, therefore, not a necessary condition even for the truth of scientific realism.

So, the point at issue between the metaphysical realist and antirealist is not to be confused with the issue concerning Φ realism, for any particular sortal term Φ . Debates between Φ realists and antirealists are debates concerning (a) whether or not claims containing the sortal term Φ have cognitive content (whether or not they purport to make reference to Φ s), and (b) whether or not any such claims are true (whether or not Φ s can be said to exist at all). That is, debates between Φ realists and antirealists are primarily debates concerning *the existence* of individual Φ s. The debate between the metaphysical realists and metaphysical antirealists, on the other hand, is a debate about *whether the entities that make up the world are mind-independent*. While a particular stance with respect to the metaphysical realism/antirealism issue can influence one's stance with respect to the existence of certain Φ s,³¹ it need not. Metaphysical realists and metaphysical antirealists can largely agree on what kinds of things exist in the world. Their fundamental disagreement is over the fundamental nature of that existence, and what something's existing in the world entails.

³¹ For example, if one believes that the existence and (most of) the features of tables, chairs and other ordinary objects are mind-independent, but that the existence and (most of) the features of moral properties and objects are not, one may be more inclined toward moral antirealism.

4. Now that we have a clearer picture of what is at issue in the debate between the metaphysical realist and metaphysical antirealist (as I am using those terms) and how it differs from what is at issue in other debates with which it might be confused, we can finally turn to an assessment of the current state of the metaphysical realist/antirealist debate. It is commonly assumed that metaphysical realism is the default position in this debate, that it is the metaphysical antirealist who must shoulder the burden of proof. The reason for this is that metaphysical realism, as I indicated above, is thought to more accurately reflect many of our pretheoretic and scientific intuitions about the nature of the world. It is difficult to see how the existence and features of a great many entities—distant stars, mountains, living creatures, just to name a few—could fail to be mind-independent. In what sense could such entities depend on the activity of minds for their existence and features, when it seems so intuitive to suppose that many would have existed even if the world had contained no minded creatures at all?³² Indeed, it is intuitive to suppose that if all of the minded individuals in the world suddenly ceased to exist, even artifacts like tables, chairs and paintings would continue to exist and have the features that they do (at least until worn down by natural forces). Such artifacts may depend on the workings of minds (not to mention hands) for their coming into being with the features

³² This intuition may not be shared, however, by anyone who is convinced that there is a God in the sense of an intelligent being who, in virtue of being the creator of the world, is responsible for the existence and features of all other individuals in the world. The reader will recall that I have exempted the mind of God from the set of minds relevant to the notion of mind-dependence in my sense; see footnote 1, p. 2 above. Even so, a theist may find that he does not share this intuition, or has no intuition at all once God is removed from the picture.

that they have, but once made, they seem to do pretty well on their own. It is a plausible thesis that if a philosophical view is a more accurate reflection of our pretheoretic and scientific intuitions on a given matter than any of its viable competitors, it should be presumed to be true unless a significant reason to doubt it is given. Since metaphysical realism seems to best reflect at least one central aspect of these intuitions, it is the metaphysical antirealist who must show that there is sufficient reason to question the metaphysical realist presumption.

But, although it may seem that the metaphysical realist may rest quite comfortably, as his is the default position in the debate, the balance in the debate is much more delicate than it may at first glance appear. To disrupt the initial balance, the metaphysical antirealist need not, in my judgment, come up with a compelling argument to the conclusion that metaphysical realism is false. For, if the metaphysical antirealist can show that it is plausible that metaphysical realism has various consequences that are in conflict with others of our equally basic commonsense and scientific intuitions, this will suffice to equalize the balance in the debate somewhat.³³ I do not mean to suggest that the metaphysical antirealist would receive the benefit of the doubt, if metaphysical realism were shown to be somewhat problematic. I only mean to suggest that any apparent conflict between metaphysical realism and others of our most basic intuitions is reason for taking a healthy agnosticism to be, in the absence of further argumentation, the default

³³ Provided, of course, that the metaphysical antirealist is able to demonstrate that his own view is philosophically viable.

position in the debate between metaphysical realists and antirealists. At the very least, we should require that the metaphysical realist do a little explaining before we again presume that his view is correct.

Since much rides on the initial assumption that metaphysical realism best reflects many of our most basic intuitions about the world, it is important that we reflect on those central intuitions before going on to see whether or not the antirealist can make a case that metaphysical realism is more problematic than it would at first appear. First, it is important to notice that not all ontologies consistent with metaphysical realism fit comfortably with our intuitions. Although seemingly metaphysically possible, it is highly counterintuitive that *this* is a world which, despite the appearances, contains nothing more than Cartesian souls and their delusions. It is highly counterintuitive that this is a world in which abstract entities like sets and numbers are metaphysically real, but material beings are not. The intuition that ostensibly gives metaphysical realism the edge in this debate is not simply that metaphysical realism is true, but that most of the entities that we ordinarily (or pre-reflectively) take to exist really do exist and have most of the features that they do mind-independently. Since it is an important part of our intuition that we are not radically mistaken about what kinds of entities have genuine ontological status, only a relatively narrow range of metaphysical realist ontologies will accommodate our commonsense intuitions about the world.

Moreover, the intuition is not merely that a vast range of commonsense and scientifically certified entities exist mind-independently, but that an important relation

obtains between our ideas and language, on the one hand, and the mind-independent world, on the other, in virtue of which our ideas and sentences succeed in being *about* the mind-independent world. It is highly counterintuitive, though again *metaphysically* possible, that while there are metaphysically real entities corresponding to many of our concepts and terms, our concepts and terms should nonetheless fail to refer to those entities because they fail to be related to those entities in the way relevant to establishing reference. Consider a world very much like the Cartesian soul world described above in section I, but with one important exception—the souls are not the only metaphysically real entities that exist in this world. While the immaterial souls go on about their business of constructing a stable world-picture in their minds, various other metaphysically real entities randomly flash into existence, enjoy brief careers, and then flash out of existence. It seems quite plausible to suppose that if it were to happen, entirely by chance, that the arrangement of metaphysically real entities matched the world-view of the Cartesian souls for some amount of time, the goings on in the minds of the Cartesian souls *would not thereby be about* the goings on of the metaphysically real entities. If one is to rule this out as a metaphysical possibility, one must make some further assumptions about the nature of the reference relation. Thus, metaphysical realism alone, even when combined with an acceptable ontology of metaphysically real entities, is not sufficient to fully satisfy our intuitions in this regard. To accommodate this aspect of the intuition, metaphysical realism is frequently combined with a thesis known as semantic realism, the view that

most, if not all, of our concepts and terms *do* succeed in making reference to mind-independent entities and (at least many of) their features.³⁴

So, there really seem to be three crucial assumptions that underlie our commonsense intuitions about the nature of the world: (i) that most of the entities that we pre-reflectively take to exist do in fact exist, that is, that we are not radically mistaken about what kinds of things have genuine ontological status; (ii) that their existence and (most of their) features are mind-independent; and (iii) that we can and do succeed in making reference to these entities and these features. If the metaphysical antirealist can show that it will be difficult to hold these three assumptions together, or that when held together they entail other things that conflict with other, equally basic intuitions, he will have provided reason to question the *presumption* of metaphysical realism.

In the next three chapters, I will examine several respects in which the metaphysical realist may be thought to have difficulty reconciling various consequences of his view with other, very strong intuitions that we have about the world and our place in it. In Chapter Two, I will suggest that there is at least one substantial skeptical worry to which metaphysical realism is susceptible and metaphysical antirealism is not. I will argue that it will be difficult for the metaphysical realist to establish that we do in fact, or even can, have knowledge of a mind-independent reality or the entities therein. In Chapter

³⁴ Note that it is only when metaphysical realism is combined with semantic realism that it bears a resemblance to a variety of Φ realism. It will be possible to be a semantic antirealist either by claiming that our concepts and terms do not refer to mind-independent entities even though such entities exist, or by claiming that there are no completely mind-independent entities.

Three, I will give reasons to doubt that the metaphysical realist can rule out various very bizarre ontologies, given anyhow the acceptability of a not implausible view about the ultimate character of the physical world. In this context I will examine several papers by Sydney Shoemaker. In Chapter Four, I will examine an argument offered by Hilary Putnam, the so-called model-theoretic argument, which suggests that semantic realism, reference to a mind-independent reality, is problematic. While the arguments presented in these chapters will not, in all likelihood, cause the committed metaphysical realist to doubt the truth of metaphysical realism—there will always be some bullet that he can bite—they will, I hope, convince the realist that he cannot retain his comfortable spot as the default position in the debate without more ado. In short, the arguments in these chapters will show (a) that the metaphysical realist requires a defense of his position just as much as the metaphysical antirealist does for his, that he cannot blithely assume that the burden of proof is on the shoulders of the metaphysical antirealist; and (b) that anyone who gives any weight at all to the intuitions which appear to conflict with metaphysical realism will have a *prima facie* reason to accept some form of metaphysical antirealism.³⁵ So, it will be clear, I hope, by the end of Chapter Four, that the metaphysical realist will need to say something in defense of his position.

Chapter Five will examine a defense of metaphysical realism offered by Peter van Inwagen in his book, *Metaphysics*.³⁶ This defense takes the form of a series of arguments

³⁵ Again, provided that there is a philosophically viable version of metaphysical realism that does not also conflict with these intuitions.

³⁶ Peter van Inwagen, *Metaphysics* (Boulder: Westview Press, 1993).

against metaphysical antirealism. Van Inwagen's arguments all turn on the assumption that metaphysical antirealism is incompatible with the existence of objective truth. I will argue that van Inwagen has not shown that the metaphysical antirealist needs to deny the existence of objective truth, and thus that his defense of metaphysical realism is ultimately unsuccessful. I conclude Chapter Five by suggesting that the metaphysical realist's best defense may be a good offense. The metaphysical realist would seem to have every right to demand a formulation of metaphysical antirealism that is relatively complete and *prima facie* coherent.

The sixth and final chapter will explore the possibility of meeting this quite legitimate demand. Despite the fact that his formulation of it is sketchy and exasperatingly metaphorical, it seems to me that the view (sometimes called internal realism) expressed by Hilary Putnam in *Reason, Truth and History* is one of the most plausible versions of metaphysical antirealism developed to date. I devote considerable space in Chapter Six to explicating and clarifying the positive theses of this view. However, Putnam's internal realism is not without its problems. Mark Johnston has argued quite convincingly that internal realism faces an objection which it cannot answer without giving the game away.³⁷ If internal realism is not itself a viable alternative to metaphysical realism, are there any other options for the would be metaphysical antirealist? In this context I

³⁷ See Mark Johnston, "Objectivity Refigured: Pragmatism Without Verificationism," in *Reality, Representation, and Projection*, John Haldane ed. (New York: Oxford University Press, 1993). See especially, pp. 87-97.

examine the constructivist metaphysical program developed by Paul Grice.³⁸ This program yields, or is at least consistent with, a form of metaphysical antirealism which, may well be able to avoid the problems faced by Putnam's internal realism. While the metaphysical picture that results from Grice's program is not without its own difficulties (some of them quite severe), it does not seem obviously incapable of being developed into a viable alternative to metaphysical realism. The upshot of the discussion will be that both metaphysical realism and metaphysical antirealism have problems (although different ones) and that it is a matter of philosophical judgment which are the more severe.

³⁸ See Paul Grice, "Reply to Richards," in *Philosophical Grounds of Rationality: Intentions, Categories and Ends*, Richard Grandy ed. (New York: Oxford University Press, 1986), pp. 45-106 and "Metaphysics and Value," TD, presented as the third of his *Carus Lectures*.

CHAPTER TWO: THE PROBLEM OF KNOWLEDGE OF A MIND-INDEPENDENT WORLD

As was mentioned in the previous chapter, it is a fundamental part of our commonsense picture of the world that we are not radically mistaken about what kinds of things have genuine ontological status. Indeed, we not only *believe* that most of the entities we take to exist do in fact exist, we believe that in many cases we *know* that they exist and have many of the features that we think they have. To accommodate our intuitions in this regard, the metaphysical realist must hold that we can and in some cases do have knowledge of the existence and features of mind-independent entities. This seems simple enough: if there are, quite independent of all minds, instances of, e.g., the kind *gold atom*, then, given our reasons for believing that there are gold atoms, it would seem that we know that gold atoms exist in the metaphysical realist's sense, and that we know them to have various features, for example, an atomic weight of 197.2.

Unfortunately, things are not so simple as they might at first appear. Knowledge is commonly taken to require justified true belief.¹ So, to have knowledge of the mind-independent world, one would have to have a justified belief whose cognitive content correctly represents or corresponds to a state of affairs that obtains independently of the existence or workings of any minds. Unfortunately, such justification does not seem easy

¹ This is a necessary condition rather than a complete definition. It is also generally recognized that some condition for the avoidance of Gettier-type problems (situations wherein a justified belief is true, but fails to count as knowledge because it is only accidentally true; the justifying conditions are not properly related to the truth of the belief) is also required. I will ignore this difficulty in my discussion of the various epistemological theories below and the possibility of knowledge of the mind-independent world. All of the theories discussed allow for the possibility of Gettier-type problems. More importantly, it is plausible to suppose that Gettier problems would arise even if any of the theories were combined with metaphysical antirealism instead of realism.

to come by. In the first two sections of this chapter, I will sketch two distinct skeptical arguments and show how metaphysical realism is susceptible to each. In the third section, I will show that several versions of metaphysical antirealism are not susceptible to these skeptical arguments. Finally, I will argue that if there is at least one skeptical worry to which metaphysical realism is susceptible and to which metaphysical antirealism is not, this will constitute a reason to question the presumption that metaphysical realism should be regarded as the default position in the debate.

1. One of the most common forms of skepticism is so-called “Cartesian skepticism.” Cartesian skepticism is the view that we are not justified in holding any (or at least most) of our empirical beliefs. Such a view can be motivated by arguments such as the following. Consider a world like the world of Cartesian souls discussed in Chapter One.² Since there are no gold atoms in this world (and never have been or will be), the Cartesian souls’ belief that gold atoms have genuine ontological status is false. Indeed, if semantic realism is true (which it has to be in order to accommodate other aspects of our commonsense world view) and their words and concepts fail to refer to existent entities, then most of their beliefs about the world will be false.

Epistemic justification is supposed to be truth-conducive; it’s supposed to make a belief significantly more likely to be true than false. Now, the Cartesian souls’ belief in the existence of, e.g. gold atoms, is false. But, this alone does not entail that their beliefs are

² Recall that that was a world inhabited entirely by nonmaterial Cartesian souls who believe that they live in a world very much like our own.

not justified. It is possible that their reasons for believing in the existence of mind-independent gold atoms (or reasons that they could come to have) are of the sort that make the belief significantly more likely to be true than false. The Cartesian souls' beliefs will be truth conducive in this sense only if (i) some principle is true which connects the grounds they have for their beliefs and the likelihood of the truth of the beliefs based on those grounds and (ii) they have good reasons for accepting the principle itself.

Unfortunately, it is far from obvious that there is any principle which has these features. So, it is far from obvious that the Cartesian souls' reasons are truth-conducive. Since, *ex hypothesi*, the world would appear in every respect the same to those deluded Cartesian souls as it does to us, any reason we have (or could come to have) that would *prima facie* justify our beliefs about the mind-independent world is one which the Cartesian souls have (or could come to have).³ Thus, since it is not obvious that their reasons are truth-conducive, it is not obvious that our reasons are truth-conducive. If this is correct, then, for all we know, *we* could be Cartesian souls inhabiting an entirely nonmaterial world. It is thus difficult to see how *we* could ever be justified in believing that there is a mind-independent world roughly corresponding to the appearances, even if such a belief is true. So, if genuine knowledge is knowledge about a mind-independent world, it is difficult to see how genuine knowledge could even be possible for us.

³ To be sure, this premise will be rejected by a proponent of an externalist theory of justification. According to an externalist theory of justification, justification is a matter of the obtaining of certain states of affairs that are external to the agent's doxastic system, e.g., the beliefs being produced by a reliable cognitive mechanism. I present and criticize such a theory of justification in the appendix, "Some Heroic Attempts to Avoid Skepticism."

This argument for Cartesian skepticism obviously turns on the assumption that a world inhabited by nothing more than deluded Cartesian souls is a genuine metaphysical possibility. However, this particular assumption is not essential. If the idea of an immaterial mind is suspect, we can instead consider a world wherein all minded creatures are, always have been, and always will be, brains in a vat of nutrients hooked up to a supercomputer that feeds them information which the brains interpret as sensory information about a material world external to the vat. Or, following Nagel,⁴ one may adopt the more general skeptical worry that it is possible that I cannot even think the truth about what I am or the truth about the way the world is due to the fact that my constitution makes it impossible for me to acquire the relevant concepts. In any case, some assumption asserting the genuine metaphysical possibility of a state of affairs in which there are creatures who have many of the same beliefs that we have (or could have) but (most of) whose beliefs are false is required to make the argument work. One way to avoid the skeptical conclusion is to mount an argument for rejecting all such assumptions. However, this line of defense does not seem available to the metaphysical realist as his view would seem to guarantee the metaphysical possibility that some assumption which could be used in an argument for Cartesian skepticism is true.

The other premise in the argument which might be questioned by the metaphysical realist is the one which states that it far from obvious that any of the Cartesian souls' grounds for holding the beliefs about the world which they hold are truth conducive. The

⁴ Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986), pp. 71 and 73.

worry that motivates this premise is that if the truth about the world *can* be so radically different from the way the world appears to the doxastic agents in it, it is difficult to see how any principle could connect the grounds the doxastic agents have for their beliefs with the likelihood that those beliefs are true, and, even if there were such a principle, it is difficult to see what sort of grounds the agents would have for believing that such a principle held. That is, it seems quite plausible that the metaphysical possibility that we are radically mistaken about the nature of the world guaranteed by the truth of metaphysical realism entails the epistemic possibility that we are radically mistaken.

What the metaphysical realist might do at this point in order to accommodate our intuitions that we do have knowledge concerning (many of) the things that have genuine ontological status is to try to provide a detailed epistemological theory that makes possible the justification of our belief in a mind-independent world that, in at least most respects, matches (or comes closest to matching) our commonsense and scientific picture of the world. That is, he may attempt to provide reasons for accepting a principle whose truth would connect the grounds *we* have for our beliefs⁵ about the world to the likelihood of the truth of these beliefs. To this end, there have been several epistemological theories designed to avoid the skeptical result developed since Descartes first offered his argument for skepticism in the First Meditation. Unfortunately, none has been widely accepted as successful.⁶ Thomas Nagel has aptly used the term ‘heroic theories of knowledge’ to

⁵ Even if no such principle would connect the Cartesian souls’ grounds to the likelihood of the truth of their beliefs.

⁶ See Laurence Bonjour, *The Structure of Empirical Knowledge* (Cambridge: Harvard University Press, 1985), esp. pp. 16-84, and Alvin Plantinga, *Warrant: the Current*

describe metaphysical realist epistemologies that attempt to avoid such skepticism. In *The View from Nowhere*, Nagel writes,

Heroic theories [of knowledge] acknowledge the great gap between the grounds of our beliefs about the world and the contents of those beliefs under a realist interpretation, and they try to leap across the gap without narrowing it. The chasm below is littered with epistemological corpses.⁷

While a fairly complete survey of epistemological theories and their relative success at combating Cartesian skepticism is well beyond the scope of this chapter, I have included an appendix which sketches some of the ways in which a metaphysical realist might try to avoid this skeptical result and discusses some of the problems for each. In particular, the appendix briefly covers the attempt by Laurence Bonjour to avoid Cartesian skepticism by appeal to the relatively low *a priori* probability of the truth of a skeptical hypothesis like Descartes', and the attempt by the so-called "externalist epistemologists" to redefine justification in such a way that cognitive access to the justificatory features of a belief is not required.

2. Another skeptical argument arises from the fact that it does not seem possible for creatures of the sort we believe ourselves to be to have unmediated access to the mind-independent world. Any information that we have about the mind-independent world is filtered through our senses and memory or based on the operation of other cognitive

Debate (New York: Oxford University Press, 1993) for fairly compelling criticisms of many of the epistemological theories designed to combat Cartesian skepticism and account for the justification of our beliefs about a mind-independent world.

⁷ *The View From Nowhere*, p. 69.

mechanisms (such as our ability to develop and evaluate theories). Further, it seems to be an entirely contingent matter that we have the sensory organs and cognitive mechanisms that we in fact do. We might, for example, have relied on sonar rather than on vision to locate distant objects. Since our access to the mind-independent is not unmediated, we must be cautious about which of the properties of our experience of the world can be attributed to the mind-independent world itself and which are determined by contingent facts about our constitutions. Nagel makes much the same point:

Objectivity and skepticism are closely related: both develop from the idea that there is a real [mind-independent] world in which we are contained, and that appearances result from our interaction with the rest of it. We cannot accept those appearances uncritically, but must try to understand what our own constitution contributes to them. To do this we try to develop an idea of the world with ourselves in it, an account of both ourselves and the world that includes an explanation of why it initially appears to us as it does.⁸

As Nagel uses the term, ‘objectivity’ is a method of coming to understand the mind-independent world by forming views that rely less and less on the subjective contributions of our make up.⁹ The idea is that our beliefs about the world are objective (that is, are about the way the world is in itself and not about the way that it appears to creatures with cognitive mechanisms like ours) when they are not ultimately rooted in idiosyncratic facts about our cognitive mechanisms about which we have imperfect knowledge.¹⁰ The worry then is that, if completely objective knowledge of the world is not possible for creatures

⁸ Ibid., pp. 67-8.

⁹ For Nagel’s definition of ‘objectivity’, see *ibid.*, pp. 4-5.

¹⁰ The idea is that if we lack knowledge concerning the idiosyncratic facts about our cognitive mechanisms, we cannot screen out or compensate for any distorting influence they might have.

like us or if we have no reason for thinking that it is possible, we have no grounds for thinking that, in our pursuit of more and more “objective” pictures of the world, we are ever getting any closer to understanding the mind-independent world as it is *an sich*.

This worry would be of no moment if we could somehow verify that the information about the world which we receive through our senses and our memories or acquire through the use of other cognitive mechanisms is accurate and is in no way corrupted by those senses and cognitive mechanisms. Occasionally we are able to discover that our constitution gives rise to beliefs about the world that are false or otherwise distorted. For example, if our current physics is correct, many of the so-called “primary qualities” of the objects that we perceive, such as solidity (where “solidity” is understood in such a way as to rule out that a solid object’s mass is largely concentrated in a tiny portion of the space the object occupies), do not correspond to any properties that mind-independent objects actually have. In such a case we may be able to compensate for the distorting influence of our cognitive apparatus by coming up with a theory which explains how and why our cognitive mechanisms mislead us, and if so, the worry would seem to be of no moment.

Unfortunately, it seems likely that such a verification of the reliability of the cognitive mechanisms which help determine our beliefs or compensation for the distorting influence of these cognitive mechanisms is never entirely possible. For this to be possible, we would need the ability to somehow step outside of the perspective on the world afforded by the cognitive mechanism(s) in question and view the world from a more

objective perspective. This would enable us to see either that the world sufficiently resembles the picture of the world that results from using the cognitive mechanism(s) in question or how and why our cognitive mechanism(s) cause us to misrepresent the world.

But, as Nagel points out,

whatever we use to understand certain interactions between ourselves and the world is not itself the object of that understanding. However often we may try to step outside of ourselves, something will have to stay behind the lens, something in us will determine the resulting picture, and this will give grounds for doubt that we are really getting closer to reality.¹¹

At each cognitive position we come to occupy in the pursuit of more and more objective views, there will be, for all we know, some residual element of subjectivity in our view(s).¹² And, what that residual element of subjectivity is and how significant it is, we won't know until we've been able to move to a still more objective view.

But why should this fact worry us? If, as seems plausible, there is a maximally objective view and there is, at least in principle, no outer limit¹³ on our pursuit of more and more objective views, then we could, in principle, obtain this maximally objective view and have knowledge of the mind-independent world *an sich*. However, there is at present no good reason to think that we anywhere near a maximally objective view, nor is there

¹¹ *The View from Nowhere*, p. 68.

¹² Nagel's point is stronger. He seems to think that there will be *ineluctably* some residual element of subjectivity in our views. I want to make the slightly weaker claim that it is epistemically possible that some element of subjectivity remains. It seems to me that it could be the case (even if we can't know that it is) that whatever cognitive mechanisms we employ in achieving the most objective viewpoint possible for creatures like us will provide us with information about the world *an sich* that is in no way corrupted or distorted.

¹³ As Nagel uses the expression, an outer limit on our pursuit of more and more objective views would be one that would result from our inability to form certain concepts and entertain certain propositions.

reason for denying the existence of outer limits on our pursuit of more and more objective views. Indeed, there seem to be reasons to think that it is very probable that there are such outer limits. Nagel develops this worry in his famous example involving the congenital nine-year-olds. Nagel asks us to imagine a species with a permanent mental age of nine, “able to think and know about the world in certain respects, but not in all.”¹⁴ To alter Nagel’s example slightly, we can imagine that one of these congenital nine-year-olds comes to believe that there are no outer limits on his coming to have more and more objective views of the world, that the world *an sich* does not extend beyond the reach of his potential understanding of it. From our perspective, knowing what we know about the world and about the cognitive limitations of his species, his belief seems not only unwarranted, but arrogant in the extreme. But, just as we can imagine a species of congenital nine-year-olds, we can imagine a species related to us as we are to the congenital nine-year-olds. And, once we do this, any denial that there are outer limits on our ability to achieve more and more objective view points would seem equally unwarranted and arrogant.

In his discussion of Kant’s transcendental idealism, Ralph C. S. Walker provides further reason for thinking that there will likely be outer limits on our pursuit of objectivity.

Our theory goes beyond the evidence, in claiming for example that objects continue to exist unperceived; to do this it relies on the principle that the simplest theory is likely to be correct—only so can we decide against the alternative hypothesis that objects spring into existence whenever a

¹⁴ *The View from Nowhere*, p. 95.

percipient comes near. It is such principles that the skeptic doubts, and they therefore need to be justified. And they cannot be justified by appeal to experience, for it is these very principles that we have to use to assess the legitimacy of any appeal to experience. If we could find some transcendental argument which showed that they would—always or usually—yield the right results, that would serve our turn, but it would be optimistic to suppose that any such transcendental argument can be provided. For it would have to show not simply that these principles are natural or convenient, but that they lead us to the truth about how things really are.¹⁵

The idea, supported by research in cognitive psychology, is that our senses are bombarded by the world and that we come up, at a very early age, with a picture of the world as containing, among other things, enduring material objects. Since more than one picture is consistent with the sensory evidence, the picture of the world at which we ultimately arrive is determined, at least in part, using principles (such as the principle that the world is simple in various respects) which seem right to us and which may, in fact, be innate. As we develop cognitively, we make these principles explicit and they come to form the basis of science and function in the development of our more objective views of the world. However, it is metaphysically possible¹⁶ that there are creatures with constitutions very different from ours who, due to their differing constitutions, have different innate principles which seem right to them. Perhaps they are inclined to accept the most complex and elaborate theory consistent with the empirical evidence. As a result, their picture of the world might be radically different from ours. If there could be such creatures or if we ourselves might have had a radically different picture of the mind-independent world had

¹⁵ Ralph C. S. Walker, *Kant* (London: Routledge & Kegan Paul, 1978), p. 123.

¹⁶ At least if metaphysical realism is true.

our constitutions been different in various ways, then we have no basis for thinking that our current picture, or any picture we could come to have given our actual constitutions, is a more accurate or objective picture of the world *an sich* than the one which would be developed by such creatures.

So, it seems that we do have reason to suspect that there are outer limits on our pursuit of more and more objective views (or at least that we have no reason to deny that there are). Thus, it seems likely that no matter how objective our view becomes, we will find ourselves unable to know whether or how our cognitive mechanisms distort or infect our beliefs and hence we shall be unable to be sure that we can correct them or fully compensate for the distorting influence of our cognitive mechanisms. Whatever we discover about the way our constitutions affect our beliefs, there will always be, for all we know, some way or ways our beliefs are affected by our constitutions which we have not taken into account and corrected for. If this is correct, we will have no reason to believe that we are coming any closer to a completely objective, mind-independent view of the world rather than just restructuring our subjective, mind-dependent perspective. So, it would seem that I can have no knowledge of the mind-independent world as it is in itself; I can only have knowledge of the way such a world appears to creatures whose sensory organs and other cognitive mechanism are similar to mine.

The skepticism which results from an argument of this kind we might call “Kantian skepticism” after the view that Kant develops in the *Critique of Pure Reason*. Kant held that all knowledge is ultimately knowledge only of appearances, and never of the world *an*

sich, the way it is independent of any and all finite minds. Since Kant himself believed that knowledge of the way the world appears to us was genuine knowledge, he did not regard his own view as a form of skepticism. It only becomes a form of skepticism if one also holds that any genuine empirical knowledge (or at least any knowledge that is not explicitly about mental states) would have to be knowledge of the character of the world as it is in itself, knowledge that reflects only the world and not facts about the knowers as well. So, one could avoid this form of skepticism by rejecting the assumption that genuine empirical knowledge is necessarily knowledge of the world *an sich*. Unfortunately, this assumption, which is sometimes called ‘epistemic realism’, is a very natural one for a metaphysical realist to make, particularly if he also accepts semantic realism. If semantic realism is true and our terms that are ostensibly about material objects, such as tables, brown cows and gold atoms, refer to putatively mind-independent entities that have many of the properties that we believe them to have, then when I assert that, e.g., I know that there is a brown cow outside my window, I am claiming to have knowledge of the spatial location of a mind-independent entity that has (at least most of) the properties of a brown cow. I am not claiming to have knowledge about the way some aspect or feature of the world *an sich* appears to creatures like us. Epistemic realism is, thus, a natural assumption for a metaphysical and semantic realist to make.

To be sure, Kantian skepticism is not as thoroughgoing a skepticism as Cartesian skepticism. Unlike the Cartesian skeptic, the Kantian skeptic will be justified in his belief that there is some state of affairs in the mind-independent world that is a cause of things

appearing to him the way that they do.¹⁷ He simply holds himself to be unjustified in believing anything further about the way the mind-independent world is in itself, apart from the appearances. But even so, if genuine knowledge is knowledge of the character of the mind-independent world in itself (and not just the way that world appears to creatures like us), then it would seem that even this more limited skepticism is at odds with our commonsense world-view.

One might try to resist the conclusion of the Kantian skeptical argument by claiming that there is some cognitive mechanism we might use which would provide us with beliefs that are unquestionably about the way the world is in itself and not merely about the way it appears to creatures with cognitive mechanisms like ours, perhaps a mechanism, call it 'rational intuition', which gives us access to necessary truths of a kind that enables us to know about the world *an sich*. Leibniz, for example, held that we can use a form of rational intuition to prove various things about God and His features which, in turn, enable us to know things about the world, e.g., that it will be simple in various respects. However, such a suggestion seems fairly implausible (at least if what is rationally intuitable is to go beyond the set of analytic truths and truths of logic) in that people disagree about what is rationally intuited (or even intuitable) and there seem to be no objective (or even inter-subjective) criteria to which we can appeal to settle such

¹⁷ To be sure, this claim would have made Kant himself a bit uncomfortable. He believed that causality was an aspect or feature of the world of appearances and not of the mind-independent world of things-in-themselves. But in any case, it is clear that he supposed there to be some analogue of causality that held between the things-in-themselves and the appearances of things.

disagreements.¹⁸ Interestingly, Walker cites the seemingly intractable nature of the disagreement concerning what is rationally intuited as one of the things which motivated Kant's development of his transcendental idealism.

Among the metaphysical assertions which [Christian August] Crusius¹⁹ confidently made were such claims as that the world has a purpose, and that every existent thing—including God—is in space, claims which Kant considered far from self-evident and which conflicted with counter-claims that appeared equally obvious to other metaphysicians.²⁰

In the absence of agreement concerning what is rationally intuited (or even rationally intuitable), we need a criterion for determining who gets things right. Not only is there currently no such criterion available, but the prospects for coming up with one at any point in the future seem quite hopeless.

So, we have now seen two fairly powerful forms of skepticism to which the metaphysical realist is susceptible. Both of these forms of skepticism are incompatible with the view that we are justified in accepting our commonsense and scientific world views. But it was precisely because metaphysical realism seemed to fit well with the common view that we justifiably believe our commonsense and scientific world views that it was accorded the default position in the debate. If it turns out that metaphysical realism

¹⁸ Contrast a debate concerning, e.g. whether something is a proof of Fermat's last theorem. In such a debate there is something objective to which we can appeal in an attempt to settle the debate—the alleged proof itself and the theorems of logic along with the axioms of Robinson's arithmetic. In the case of disagreement concerning rational intuition, all of the data that would be relevant to settling the dispute are purely subjective mental states or properties thereof.

¹⁹ A contemporary of Kant's whom Walker credits with distinguishing the class of those necessary truths Kant would go on to call analytic from those which Kant would call synthetic. See *Kant*, p. 3.

²⁰ *Ibid.*, p. 4.

is open to skeptical challenges the cogency of which are incompatible with this common view, then it is far from clear that it should retain its privileged status as the default position. To be sure, I have provided no conclusive reason to think the metaphysical realist cannot avoid skepticism. There may well be arguments that he can mount to meet both of the types of skeptical arguments I have sketched. However, I hope that I have succeeded in showing that the metaphysical realist does have a problem to which there is no obvious solution—as the long history of highly controversial attempts to deal with it makes clear.

3. So, the metaphysical realist needs to handle several kinds of skeptical challenge if his view is not to conflict with our commonsense intuition that we are not radically mistaken about the existence and features of the entities that have genuine ontological status. But what of metaphysical antirealism? Does the metaphysical antirealist have analogous difficulties with Cartesian and Kantian skepticism? If this should turn out to be the case and all else remains equal, the metaphysical realist will retain the default position in the debate as his is the view that accords with more of our deeply held intuitions about the nature of the world. In the sections that follow, I will argue that while metaphysical antirealism is not immune to any and all skeptical worries, many varieties of antirealism do fare significantly better than metaphysical realism. This will not come as a surprise to some. Many have held that skepticism is an inevitable consequence of metaphysical

realism and that metaphysical antirealism is an essentially anti-skeptical view.²¹ Indeed, it may well have been the worry that skepticism was intractable given a realist metaphysics that historically led to the rise of antirealist metaphysics. Even so, I think that it will be worthwhile to examine some varieties of metaphysical antirealism and evaluate the extent to which they provide answers to the skeptical arguments sketched above.

Recall that metaphysical antirealism is the view that while there are entities that exist, neither their existence nor any of their features are independent of the existence or activities of any mind or minds. Generally, the metaphysical antirealist subscribes to a theory of truth or meaning that rules out the possibility that anything can be said to exist and have the features that it does in the metaphysical realist's sense. I will examine two such theories: rational acceptability theories of truth and idealized rational acceptability theories of truth. In this discussion, I shall assume for the sake of the argument that each of the types of theory presented is a philosophically viable view, and shall here consider only the extent to which each is vulnerable to certain skeptical arguments. A more complete evaluation of these forms of metaphysical antirealism will be put off until Chapters Five and Six of the dissertation.

First let us examine a form of metaphysical antirealism which results from an adherence to a rational acceptability theory of truth.²² According to such a theory of

²¹ See Nagel, *The View From Nowhere*, pp. 67-74, and Robert Coburn, *The Strangeness of the Ordinary: Problems and Issues in Contemporary Metaphysics* (Savage, MD: Rowman and Littlefield, 1990), pp. 144-147.

²² Putnam discusses such a theory on p. 55 of *Reason, Truth, and History* (Cambridge: Cambridge University Press, 1981).

truth, a sentence is true if and only if it is currently rationally acceptable, where a sentence is rationally acceptable, e.g., if it is analytic or contributes to the explanatory power of some well-confirmed theory or has been directly verified by observation.²³ These conditions are by no means exhaustive. They are simply intended to give the reader the general flavor of a rational acceptability theory of truth. It is important to keep in mind that a proponent of a rational acceptability of truth is not merely providing a stipulative definition of the term 'truth'. Nor is he providing a theory according to which truth is, as a purely contingent matter, coextensive with what is currently rationally acceptable. Rather, he is engaged in spelling out the considerations that go into the concept of truth, thereby providing a theory about the *nature* of truth itself.²⁴ At its core, the concept of truth involves the notion of agreement with reality, getting the world right. But what is it to agree with reality? What is it to get the world right? According to a proponent of the rational acceptability theory of truth, a proposition agrees with reality (gets the world right) when it is currently rationally acceptable. Such a theory will clearly yield metaphysical antirealism: what things can be truly said to exist will depend on what it is currently rational for certain minds (namely ours) to accept.

Anyone who accepts a rational acceptability theory of truth will be able to reject the conclusion of any typical argument for Cartesian skepticism. *Ex hypothesi*, if our world is, e.g., a world inhabited entirely by nonmaterial Cartesian souls, it is not currently

²³ If a sentence does not fall into one of the categories of currently rationally acceptable sentences, it is not currently rationally acceptable.

²⁴ Compare the various analyses of persistence through time discussed by David Lewis in *On the Plurality of Worlds* (Oxford: Basil Blackwell, 1986), p. 202.

rational to accept that our world is a world inhabited entirely by nonmaterial Cartesian souls.²⁵ What it is rational to accept is our current world-view according to which we are, among other things, material beings in a material world. Hence, on the rational acceptability theory of truth, it will be simply false that our world is a world of the sort the skeptical hypothesis describes. So, to the extent that we have reason to accept the rational acceptability theory of truth, we have reason to reject the epistemic possibility that we are deluded Cartesian souls or in some other way radically mistaken about the world. Cartesian skepticism can, thus, be avoided.

The proponent of a rational acceptability theory of truth will be able to avoid Kantian skepticism as well if he can show that it is not rational to accept the skeptical hypothesis which drives it. That hypothesis was, in essence, that any beliefs we have or could come to have concerning the way the world is *an sich* (and not merely the way that it appears to creatures like us) might, for all we know, be false (or distorted) because infected by idiosyncratic facts about our constitution. Some will have been convinced by the discussion above²⁶ that it is rational to accept the epistemic possibility that our sensory and other cognitive mechanisms infect and corrupt our beliefs about the world. But that argument was based on the assumption that metaphysical realism was true. If the world exists and has the features that it does mind-independently, it does seem rational to accept the Kantian skeptical hypothesis. But, on the version of metaphysical antirealism that we

²⁵ Similarly for any other skeptical hypothesis that could be used to support Cartesian skepticism, since all such hypotheses involve the supposition that our rationally acceptable beliefs about the world are false.

²⁶ See above, pp. 38 and following.

are now considering, if it is rational to accept that, e.g., the world contains electrons and quarks (as it would in fact seem to be), then the world does contain electrons and quarks—on the rational acceptability theory of truth, there won't be a distinction between the way the world is *an sich* and the way it appears. Any belief we have about the way the world is and which is rationally acceptable (such as our belief that the world contains electrons and quarks) will be true *simpliciter* and not merely true of the appearances. So, our rationally acceptable beliefs about how the world is *an sich* can't be false (or distorted) because infected by idiosyncratic facts about our constitutions.

To hold that it is (or could become) rational to accept the Kantian skeptical hypothesis is to assume that there is some sense in which something can be true of the world which goes beyond anything that it would be rational to accept. To put the point another way, it would be rational to accept the Kantian skeptical hypothesis only if the rational acceptability theory of truth were false; any reason we have for accepting the rational acceptability theory of truth is *ipso facto* a reason for rejecting the Kantian skeptical hypothesis. Thus, a defender of the rational acceptability theory of truth can reject the crucial premise that drives Kantian skepticism.

Even if there is a way of reformulating the Kantian skeptical hypothesis in such a way that it is (or could be) rational to accept on a rational acceptability theory of truth, the success of the rational acceptability theory of truth against the Cartesian skeptic should be sufficient for shifting slightly the balance of the debate. It is Cartesian skepticism that is most at odds with our commonsense intuitions about the extent of our knowledge about

the world. Cartesian skepticism entails that we are not justified in believing that entities even roughly corresponding to most of the entities that we believe to exist have genuine ontological status. Kantian skepticism on the other hand need only entail that we are not justified in believing that the entities with genuine ontological status have exactly the properties and features that we believe them to have. Kantian skepticism would seem to be consistent with the claim that there are entities with genuine ontological status that correspond to and are the cause of our beliefs about the entities that make up the world. Whether or not the metaphysical antirealist who accepts a rational acceptability theory of truth is any better off with respect to the Kantian skeptic than the metaphysical realist, it is clear that he can answer the Cartesian skeptic more easily than can the metaphysical realist. Thus, the two metaphysics are on a more even footing with respect to the deeply held intuitions about the world with which they appear to cohere or conflict.

Some philosophers²⁷ have adopted a theory of truth that holds that the concept of truth is not to be analyzed in terms of what it is *now* rational to accept, but rather that truth is to be analyzed in terms of what *would be* rational to accept at some ideal limit of inquiry. More will be said about this view in Chapter Six of the dissertation. For now, this very rough sketch of what might be called the idealized rational acceptability theory of

²⁷ Perhaps most notably, Hilary Putnam. See *Reason, Truth, and History*. "The view which I shall defend holds, to put it very roughly, that there is an extremely close connection between the notions of *truth* and *rationality*; that, to put it even more crudely, the only criterion for what is a fact is what it is *rational* to accept....But the relation between rational acceptability and truth is a relation between two distinct notions. A statement can be rationally acceptable *at a time* but not *true*..."(p. x).

truth should be sufficient for evaluating its relative success against the skeptical arguments.

It is clear that the idealized rational acceptability theory will be able to refute the Kantian skeptic to the extent that, and in much the same way as, the rational acceptability theory is able to. If the idealized rational acceptability theory of truth is correct, then there is no distinction between what it is rational to accept about the way the world is at the ideal limit of inquiry and the way the world is *simpliciter*. While we may not currently have a completely objective view of the world, the idealized rational acceptability theory of truth guarantees that it is in principle possible to achieve a maximally objective view of the world. So, the Kantian skeptical hypothesis, that any beliefs we could come to have concerning the way the world is *an sich* might, for all we know, be false (or distorted) because infected by idiosyncratic facts about our constitution, can be true only if the idealized rational acceptability theory of truth is false. However, the fact that, on such a view, what is true is what it would be *ideally* rational to accept makes it more difficult to provide a completely adequate response to the Cartesian skeptic. Since we certainly have not now reached the ideal limit of theorizing and inquiry, we are not now in a position to know with any certainty that, at the ideal limit of theorizing and inquiry, it would not be rational to believe that we are in fact nothing more than Cartesian souls, for example. It seems possible that at the ideal limit of inquiry, it would be rational to accept one of the Cartesian skeptical hypotheses. Since this theory of truth does not entail that the Cartesian skeptical hypothesis makes no sense, or is known to be false, it cannot simply

deny the conclusion of the skeptical argument that we are not justified in our beliefs about the existence and features of the entities with genuine ontological status without further ado.

So, is the metaphysical antirealist who adopts an idealized rational acceptability theory any better off than the metaphysical realist with respect to Cartesian skepticism? I believe that the answer is 'yes'. A proponent of the idealized rational acceptability view might very well point to the continuity and relative stability of our scientific theories in an effort to develop at least a partial response to the Cartesian skeptic. While there have been significant revolutions in scientific thought over the ages, developing theories seem to be converging on a single, coherent world view. This might give us reason to believe that our current set of rationally acceptable beliefs about the existence and features of entities in the world is a reasonable approximation of the truth about the world. If our rationally acceptable theories seem to be converging on some world view, it is plausible to infer that what it would be ideally rational to accept is that world view on which our theories are converging. It is plausible that ideal rational acceptability is a sort of asymptotic limit which our currently rational acceptable theories approach.

If that is right, it would appear that the metaphysical antirealist who accepts an idealized rational acceptability theory of truth can avoid the highly counterintuitive conclusion of the Cartesian skeptical argument. We may not know with certainty that the Cartesian skeptical hypothesis is false; but, we are justified in believing that it is false since the apparent historical convergence of our rationally acceptable theories would seem to

give us some insight into what it will be rational to accept at the ideal limit of theorizing and inquiry. We might be somewhat off target, but we probably won't be radically mistaken as the Cartesian skeptical hypothesis implies we are.²⁸

The previous remarks may be sufficient to convince some that an idealized rational acceptability theory of truth can provide a response to the Cartesian skeptic; others may still have reservations. It is plausible to suppose that the historical convergence of rationally acceptable views does suggest that the ideally rationally acceptable view is an asymptotic limit which we will continue to approach. But it is certainly not impossible that some scientific revolution should occur which would cause a radical shift in what it is rational to accept about the existence and features of various entities in the world.²⁹ After a sufficiently far-reaching revolution, it might not be possible to view previously rationally acceptable theories as converging on the same world-view as our current rationally acceptable theories. It could well be that after centuries of relatively stable scientific advances, we might become convinced that we had been on the wrong track all along. Indeed, it is logically possible, even given the current apparent convergence on an ideal

²⁸ This line of response may seem at first to be so promising that one may wonder why the metaphysical realist cannot himself respond to the Cartesian skeptic in such a way. Perhaps he can. But doing so will require a much more heroic effort on the part of the metaphysical realist than the proponent of the idealized rational acceptability theory of truth. It is quite plausible (although far from certain as we will see) that a historical convergence of rationally acceptable views justifies us in a belief that what it would be ideally rational to accept is the view on which our current theories are converging. It is far less plausible that the historical convergence of rationally acceptable views justifies us in believing anything about a completely mind-independent world. After all, even an ideally rationally acceptable view might be false, on the metaphysical realist's view.

²⁹ Relativity plausibly marked such a shift. Perhaps the success of Quantum Mechanics represents another such a radical shift. (Although, not all will agree.)

scientific theory, that at the ideal limit of theorizing we would accept that our earlier scientific and commonsense world views had been radically mistaken, that, in fact, we are nothing more than nonmaterial Cartesian souls in a nonmaterial world. As long as this remains a metaphysical possibility, some may find it difficult to see how we could be currently justified in believing that our currently rationally acceptable beliefs are even a reasonable approximation of the truth about the existence and features of the entities with genuine ontological status.

Even so, the proponent of the idealized rational acceptability theory of truth is in a slightly better position relative to Cartesian skepticism than the metaphysical realist. Both views will have difficulty avoiding the counterintuitive result that we just might be radically mistaken about the existence and features of the entities that have genuine ontological status. However, for the metaphysical antirealist who accepts the idealized rational acceptability theory it will be, at least in principle, possible to have knowledge about the entities that genuinely make up the world even if there should turn out to be nothing more than Cartesian souls. Indeed, it is in principle possible that we would have evidence that would decisively refute the skeptical hypothesis. After all, it is possible that what will be rational to accept at the ideal limit of theorizing is something very like our current commonsense and scientific world-views. For the metaphysical realist, on the other hand, it would appear that, apart from the unlikely success of a heroic epistemology, it will not even be in principle possible to have such knowledge. So, the metaphysical

antirealist who adopts an idealized rational acceptability theory will (most likely) be better off in this regard than the metaphysical realist.

4. So, as we have seen, metaphysical realism faces a serious skeptical challenge from both the Cartesian and the Kantian skeptic. Various forms of metaphysical antirealism are able, to greater or lesser extents, to avoid the skeptical challenges that dog metaphysical realism. The rational acceptability theory would seem to be able to adequately refute both Cartesian and Kantian skepticism. The idealized rational acceptability theory can refute Kantian skepticism and, while not completely successful in refuting Cartesian skepticism, is able to avoid the skeptical worry that knowledge about the existence and features of the entities with genuine ontological status is not even in principle possible for us. While these views are by no means the only varieties of metaphysical antirealism, they do constitute a fairly representative sample.

The fact that versions of metaphysical antirealism have an easier time with certain types of skeptical arguments than does metaphysical realism is by no means enough to show that metaphysical realism is false or that our beliefs about the mind-independent world aren't in fact justified. Perhaps some heroic epistemology will be successful after all. Indeed, even in the absence of a successful heroic epistemology, many will still find the skeptical results of the Cartesian and Kantian arguments less counterintuitive than the metaphysical antirealist's nonskeptical alternatives. However, it does reveal at least one respect in which metaphysical realism is not the completely intuitive, commonsensical view

it first appeared to be. Metaphysical realism was accorded the default position in the debate on the grounds that it accorded with more of our commonsense and scientific intuitions about the world than did metaphysical antirealism. But we have now seen that it is the metaphysical antirealist who is better able to accommodate our intuitions about our current and possible states of knowledge with respect to the existence and features of the entities that have genuine ontological status. This, combined with still other difficulties which I will raise for metaphysical realism in the chapters that follow, will provide us with sufficient reason for questioning the presumption of metaphysical realism. A defense of metaphysical realism will, therefore, be called for, if we are to have grounds for preferring it to an antirealist metaphysic.

CHAPTER THREE: THE PROBLEM OF THE POSSIBILITY OF A VERY CROWDED ONTOLOGY

1. As we saw in the previous chapter, the metaphysical realist will have a significantly more difficult time answering the skeptic and preserving our deeply held, commonsense intuition that we are not radically mistaken about the existence and most of the features of the objects with genuine ontological status than will many versions of metaphysical antirealism. If the Cartesian skeptical hypothesis is true, then my belief that there are material objects at all is false. If the Kantian skeptical hypothesis is true, then while I am correct that there is something in the mind-independent world corresponding to the material objects that I ostensibly perceive and believe to exist, my ways of perceiving and thinking about these objects, for all I know, distort my beliefs about their features.

Putting aside these two worries, however, there is still another sense in which we might be mistaken about the things with genuine ontological status. Among the objects that we believe to have genuine ontological status are various kinds of enduring material objects. In fact, we seem to be relatively well equipped for perceiving such enduring material objects and many of their features. Suppose that our belief in this regard is correct, that there are material objects which endure through time (with the result that the Cartesian skeptical hypothesis is false). Suppose further that the Kantian skeptical hypothesis is false and that our perceptual and other cognitive mechanisms do not, for all we know, distort or contaminate the beliefs we form about many of the features of these objects. However, even if all of this is true, it is *prima facie* metaphysically possible, although highly counterintuitive, that there are, in addition to the objects and their features referred to in our true beliefs about the world, a wide variety of objects and properties that

escape our notice. It is not that these putative objects with these putative properties would be too small or too far away to get noticed; many would not be. It is instead that they are too strange, too fishy, as Sydney Shoemaker has put it,¹ to seem genuine to us. It will be incumbent upon any metaphysic which is to accord with our commonsense view of the world to provide criteria for distinguishing between genuine objects and their features and these fishy objects and properties which we do not pre-reflectively believe to exist. Only once such an account is provided can we be confident in believing that our commonsense belief that we are largely correct about what things have genuine ontological status is itself correct.

We must be clear about what the substantive issue is here. The question is not why there are, e.g., tigers, but not mountains made entirely of gold, or why giant ground sloths once existed, but never centaurs. Given a theory of truth, say the correspondence theory, we can say that 'there are tigers' is true because the stereotypical properties associated with the sortal term 'tiger' are instanced by various individuals in the world. 'There is a golden mountain' is false, because there are no individuals in the world that instantiate the stereotypical properties associated with the sortal term 'golden mountain'. But there could have been! There are countless possible worlds wherein there is at least one golden mountain. While most of us are inclined to be antirealists about golden mountains,² when we consider the whole of logical space most of us would grant that

¹ Sydney Shoemaker, "On What There Are," *Philosophical Topics* 16 (Spring 1988), p. 201.

² See above, pp. 11-13, for a discussion of Φ -realism and Φ -antirealism.

‘golden mountain’ is a sortal term that refers to a genuine sort, instances of which have various genuine properties—it is just that the sort *golden mountain* has no instances in the actual world. Any substantive criteria for distinguishing between genuine and fishy objects and properties will have to take into account much more than the merely contingent instantiation (or lack thereof) of individuals of a given sort (or exemplifying a given property) in the actual world.

If various sortal and non-sortal property terms can in some sense count as referring to genuine sorts and non-sortal properties even if there are, in fact, no instances of those sortals or non-sortal properties in the actual world, what is to prevent just *any* putative sortal or non-sortal property term from referring to a genuine kind or non-sortal property in this sense? Clearly terms such as ‘thing which is red all over and green all over’ and ‘married bachelor’ fail to refer to anything in any possible world because instances of either would have to have inconsistent properties. It is impossible for anything which is red all over at a given time to be green all over at that time; it is impossible for anything that is a bachelor to be married at the same time.

But is it only sortal and non-sortal property terms whose definitions include inconsistent properties (let us call these ‘inconsistent sortal and other property terms’) that fail to refer to genuine sorts? Intuitively, that is not sufficient; we don’t want just any set of consistent properties to define a genuine sortal or non-sortal property term. To see why this is so, suppose that all consistent sortal terms refer to genuine sorts, and consider, for example, the sortal term ‘klable’ introduced by Sydney Shoemaker in “Identity,

Properties and Causality.”³ Anyone who has exactly one kitchen table and exactly one living room table will have a klable. A klable is a putative object whose history is identical with the stages of the history of the kitchen table from midnight till noon and the stages of the history of the living room table from noon till midnight. The klable is in the kitchen for every instant after midnight up to and including noon and it is in the living room every instant after noon up to and including midnight.⁴ At no time is the klable in both the living room and the kitchen. Since there is nothing logically impossible about supposing an individual to have such properties, ‘klable’ would have to be a genuine sortal term. And not only are there klables in some possible worlds or other on this hypothesis, but there are klables right here in the actual world. Indeed, I’m writing on one now.

I suppose that someone could say, “So there are klables as well as tables, big deal. They don’t take up any extra room in our houses.” But I imagine that most will find the idea that there are such things as klables quite absurd. For example, the existence of klables should seem absurd to anyone inclined to believe that two distinct material objects cannot occupy, at exactly the same time, exactly the same region of space. After all, the klable and the kitchen table are clearly distinct objects (if the klable is an object at all) since they have distinct persistence conditions; if my living room table is destroyed by fire,

³ Sydney Shoemaker, *Identity, Cause, and Mind* (Cambridge: Cambridge University Press, 1984), p. 258.

⁴ If time is continuous (or even if it is merely dense), there will be no first instant at which the klable is in the kitchen, although there will be a last, noon. Similarly for the living room stages of the klable.

my klable ceases to exist,⁵ but (provided that the fire didn't consume my kitchen as well) my kitchen table is unaffected. However, if there are klables, it would seem that the spatial region occupied by my kitchen table at 10:30 a.m. is simultaneously occupied by a klable!

Despite the counterintuitive nature of the idea that there are klables, it is far from easy to provide sufficient philosophical grounds for denying their existence altogether. The providing of such grounds is one of the central concerns for any view that would lay out criteria for distinguishing between genuine and fishy objects and properties. Any metaphysic that is to cohere well with our basic intuitions concerning the existence and features of objects that comprise the world and our knowledge of those objects and features should be able to provide such criteria. If it should turn out that the metaphysical realist will have a significantly more difficult time ruling out the existence of fishy objects and properties than the metaphysical antirealist, then we will again have reason to suspect that metaphysical realism does not deserve its traditional status as the default position in the debate. Metaphysical antirealism will again be seen to better accord with some of our most basic intuitions about the world.

⁵ This too seems absurd. 'Klable' is a term that refers to a putative object with an arbitrarily gerrymandered history. It should be obvious that we could similarly introduce terms referring to an infinite number of putative objects whose histories overlap in some way or other with the history of my living room table. It seems absurd to suppose that when my living room table is destroyed by fire, an infinite number of objects are simultaneously destroyed. This absurdity was brought to my attention by Robert Coburn.

In two well-known papers,⁶ Shoemaker sets out a view concerning the individuation of properties and objects that attempts to explain what it is about klables and other fishy objects and properties, that prevents them from being included in our commonsense ontology. This view focuses on the causal powers which are possessed by individuals in virtue of their instantiating genuine properties. A genuine property is one which, according to Shoemaker, is a cluster of conditional powers. Thus a thing has the causal powers it has owing to the genuine properties which it instantiates. While Shoemaker's approach is far from being the only way to distinguish between genuine and fishy properties and objects, it is one of the more thoroughly worked-out views on the subject and is, in many ways, fairly plausible.

In the remainder of this chapter, I will, first, briefly sketch Shoemaker's causal account of property and object individuation and discuss how the account seems to rule out the existence of such strange objects as klables. Second, I will explicate and discuss a later turn in Shoemaker's thinking that raises a difficulty for his earlier account, and indeed any causal account of object or property individuation—namely that it fails to rule out a wildly counterintuitive ontology crowded with fishy objects and properties. Third, I will argue that such an ontology is ruled out by the adoption of many forms of metaphysical antirealism. Fourth, I will argue that metaphysical realism, when conjoined with a widely accepted view about properties and the idea that sets of spacetime points that instantiate

⁶ "Causality and Properties," in *Identity, Cause, and Mind*, pp. 206-233, and "Identity, Properties and Causality," in *ibid.*, pp. 234-260.

properties of certain sorts count as genuine objects leads inexorably to this wildly counterintuitive ontology. Finally, I will argue that, while such an ontology is easily ruled out by the adoption of many forms of metaphysical antirealism, the metaphysical realist will have to defend some controversial theses in order to rule it out. Some may view the difficulties that arise when metaphysical realism is conjoined with certain ideas Shoemaker advances as a reason for rejecting these latter ideas rather than something which tells against the plausibility of metaphysical realism. However, given the initial plausibility of Shoemaker's ideas, it would seem that the metaphysical realist has a heavier burden to discharge in order to rule out the fishy objects than does the metaphysical antirealist. And, if this is right, we have further reason to question the idea that the metaphysical realist occupies the default position in the debate.

2. As I mentioned above, a *klable*, if indeed there are such things, may be thought of as an object whose history corresponds, roughly, to the midnight till noon stages of a kitchen table and the noon till midnight stages of a living room table in the same home. Anyone inclined to view tables and other material objects as mereological sums of time slices or thing-stages will likely find the putative existence of *klables* relatively unproblematic. After all, if the objects we view as ordinary continuants are themselves nothing more than mereological sums of thing-stages, what could be lacking in a *klable* that would rob it of ontological status? Spatiotemporal continuity might be a requirement for our taking much notice of or having an interest in objects of a given sort. But it is not

at all clear that such continuity (or the lack thereof) is a metaphysically relevant criterion for the very existence of a mereological sum of thing stages.

Shoemaker and others have argued that the metaphysical view that drives such klable-friendly intuitions is mistaken. It is wrong to view ordinary objects such as tables as nothing more than sums of thing-stages, since sums of thing-stages do not have the modal properties we take ordinary objects to have.⁷ We believe, for example, that my kitchen table, this very table, could have been destroyed by fire on September 17, 1994, although it in fact was not. But if my table is merely a collection of table-stages, there are no possible worlds where this table gets destroyed by fire on September 17, 1994. Any table in any possible world that is destroyed by fire on that date is a different mereological sum of table-stages than the one in my kitchen which was not destroyed on that date,⁸ and hence a different table. Mereological sums cannot be composed of different summands in different possible worlds.⁹ But if my table is not just a collection of table-stages, what does it have that the klable lacks that would be of metaphysical relevance to its ontological status?

⁷ See p. 204 of "On What There Are."

⁸ My table includes table stages (all of the ones after September 17, 1994) not included in the table in the other possible world.

⁹ To be sure, I am not entirely convinced by this reasoning. Many questions remain concerning the cross-world identity of even ordinary objects. Lewis has argued that the notion of cross-world identity should be replaced with a theory that makes reference to otherworldly counterparts. If a counterpart theory should prove necessary as a replacement for cross-world identity, it is not at all clear that mereological sums of thing-stages are any worse off than ordinary continuants. However, it is not clear to me that the metaphysical dispute concerning whether there are ordinary continuants or only mereological sums of thing-stages is ultimately relevant to the issue at hand.

Shoemaker's original answer is that the klables lack the relevant causal relations between their successive stages. When we consider the sort of evidence that is relevant in judgments of identity over time, we find that the fact that an object A and a later object B have similar properties can be used as evidence that A and B are the same object *only* if we also have good reason to believe that the similarity is evidence of a counterfactual dependence holding between B's now having the property (or properties) and A's having had it (or them) earlier.

What this strongly suggests is that those series of thing-stages that are histories of persisting things, genuine continuants, are distinguished from series that are not, that is, series made up of stages from histories of different things, by the fact that what properties are instantiated in later stages of a genuine history is a function of, among other things, what properties were instantiated in earlier stages of it....[Similarity of properties] is often used as evidence of identity; but,...it is evidence of identity only insofar as it is evidence of counterfactual dependence, which in turn I take to indicate a causal relationship. And when things change, their new properties are normally a function of their old; how something is affected by interaction with other things will depend on what properties it had prior to the onset of the interaction, and in processes such as biological growth later stages are causally generated out of earlier ones in a way that yields a fixed pattern of change.¹⁰

On this view, what distinguishes the series of table-stages that make up the history of the table from those that allegedly make up the history of the klable is that the stages of the former but not the latter stand in relations of appropriate causal dependence on earlier stages. In order to understand what the appropriate causal dependence between stages is,

¹⁰ "Identity, Properties, and Causality," pp. 239-240.

according to Shoemaker, we must understand his causal account of the identity of properties.

The notion of a causal dependence between thing-stages that is involved when they count as stages in the history of a genuine object is tied to the notion of genuine change, which in turn is tied to the notion of a genuine property. But, just as there can be said to be fishy objects like klables, there can also be said to be fishy properties like Nelson Goodman's *grue*,¹¹ and relational properties like *having been slept in by George Washington*. So, Shoemaker provides us with criteria for distinguishing genuine properties from both fishy properties and relational properties.

[W]hat makes a property the property it is, what determines its identity [when it is a genuine property],¹² is its potential for contributing to the causal powers of the things that have it. This means, among other things, that if under all possible circumstances properties X and Y make the same contribution to the causal powers of the things that have them, X and Y are the same property.¹³

To say that a property contributes to the causal powers of things that have it is to say that when co-instantiated with various other properties it gives rise to certain causal powers. Shoemaker's favorite illustration of this phenomenon is the property of being knife-shaped. This property, when co-instantiated with the property of being made of steel, gives rise to the power to cut wood. Thus, giving rise to the power to cut wood when co-

¹¹ The property that anything has just in case it is green and examined before 2000 C.E. and otherwise is blue.

¹² Shoemaker is perfectly happy to count relational properties and fishy properties as properties, although not as properties with genuine ontological status. He refers to such properties as "mere Cambridge" properties; see "Causality and Properties," pp. 207-8.

¹³ Ibid., p. 212.

instantiated with the property of being made of steel is part of the essential nature of the property of being knife-shaped. Non-genuine, properties, on the other hand, do not contribute to the causal powers of things that have them. As evidence of this claim, Shoemaker asks us to consider two beds, made by the same craftsman out of the same piece of oak, one of which has the further putative property of having been slept in by George Washington. Let us examine the two beds both with the naked eye and with any fancy scientific equipment that would be appropriate.¹⁴ Examine as we like, we will find that the two beds are identical in their causal powers, and thus, according to Shoemaker's view, have all the same genuine properties.¹⁵ If this is correct, then *having been slept in by George Washington* is not a genuine property.

This causal account of genuine properties allows Shoemaker to say a bit more about the appropriate causal connections between thing-stages required for the existence

¹⁴ Let us further imagine that no one involved was previously aware that the one bed had been slept in by the father of our country.

¹⁵ Some will worry that it is possible, and indeed plausible, that two distinct sets of properties can give rise to the same causal powers in the objects which instantiate them. Consider that, if a set 1 of co-instantiated properties gives rise to causal powers A and B, a set 2 of co-instantiated properties gives rise to causal powers C and D, a set 3 of co-instantiated properties gives rise to causal power A, and a set 4 of co-instantiated properties gives rise to causal powers B, C, D, (where the intersection of any two of these sets is empty), then any object which instantiated the properties of sets 1 and 2 would have the same causal powers as an object which instantiated the properties of sets 3 and 4. So, it should be possible for two objects with distinct properties to share identical causal powers. However, in the example in question, we can presuppose that the two beds are known to share all of the same properties except the putative property *having been slept in by George Washington*. Since the possession of this alleged property by the one bed would seem to have no effect whatever on its causal powers, it is safe to assume that the alleged property *having been slept in by George Washington* is not a genuine property at all in Shoemaker's sense.

of a genuine object. According to Shoemaker, it is both a necessary and sufficient condition for a series of thing-stages' being the history of a genuine object that its stages be causally connected in such a way that its later stages develop from the earlier ones "in accordance with the laws of immanent causality that are, as it were, built into the nature of the properties instantiated in the states of the series (and in the stages of other series with which it interacts)."¹⁶ The earlier stages of a genuine object will have properties which will entail the possession by the object of certain causal powers, and its having those powers ensures that a later stage of the same object will, e.g., bring about various genuine changes given various external circumstances, provided that no other changes occur to prevent their occurrence. For example, if a given object has the property of being hot at *t*, then if a later stage of the same object comes into contact with my finger, and no external influences have caused the object to lose this property, it will give rise in me to a sensation of heat (provided that I am in a normal perceptual state). If it turns out that one thing-stage has the property of being hot and another thing-stage comes into contact with my finger without giving rise to a sensation of heat (and no external influences have caused the "object" to lose the property), then the two thing-stages cannot be stages of the history of a genuine enduring object (provided that I am in a normal perceptual state). This is due to the fact that it is, as Shoemaker puts it, "built-in" to the nature of the property being hot that a thing having that property will continue to have it in the absence of (or change in) external influences (such as my blowing on it) which would cause it to

¹⁶ "Identity, Properties, and Causality," p. 254.

cool off, and further, that anything that has the property of being hot will have (among others) the power to cause a sensation of heat in normal perceivers for as long as it has that property. So, the two thing-stages in the second case are not stages in the history of a genuine continuant since they violate Shoemaker's condition that a series of thing-stages which is the history of a genuine continuant be connected by the laws of immanent causality.

Shoemaker argues, quite convincingly, that the notions of causality, property and identity are all bound up together; there will be no way to give an analysis of any one of these notions that does not involve reference to the other two. Thus, an analysis of causality must make reference to the notion of a continuant and the idea of a property.

[S]uppose that a branch is blown against a window and breaks it. Here the constituent objects [of the causal relation] include the branch and the window, and the causal relationship holds because of, among other things, the massiveness of the one and the fragility of the other. It would appear from this that any account of causality as a relation between events should involve, in a central way, reference to the constituent objects of the events.¹⁷

And, as we have seen, spelling out the nature of a genuine property involves reference to causality and to the idea of a genuine persisting object in which the property is instantiated. Finally, the constitutive criteria for being a genuine persisting object require both the notion of a property and the idea of causality. While the truth of these claims does rule out the possibility of providing a reductive account of any of these notions,

¹⁷"Causality and Properties," p. 207.

Shoemaker argues that such circularity does not render his account uninformative, and is, indeed, unavoidable.¹⁸

Now, exactly how is all this supposed to rule out the existence of klables and other fishy objects? Suppose that my kitchen table is painted red but that my living room table is painted blue. Then, since at 11:59 P.M. my kitchen table had the property of being red, any entity that shares that portion of its history with my kitchen table also has to have had the property of being red at that time. On Shoemaker's account of property individuation, to say that an object is red is to say, among other things, that if viewed in sunlight under otherwise standard viewing conditions, it will have the power to produce a red sensation in me, provided that I am a normal perceiver. So, if I view my table in standard sunlight at noon, it should produce in me a red sensation.¹⁹ Similarly, if the klable is a genuine object and is red at 11:59, when I view it at noon it too should produce a red sensation in me. But when I look at my klable (which is now in the living room) at noon, a blue sensation is produced in me! The appropriate causal conditionals have been falsified. To paraphrase Shoemaker, if there really are klables, then what happens to them at noon and midnight violates the laws that spell out the essential natures of the properties that they would have to have.²⁰

¹⁸ Ibid., p. 222, and "Identity, Properties, and Causality," p. 260.

¹⁹ To be sure, this conditional holds only if the table has not been caused by external forces (by its being painted, for example) to lose the property of being red.

²⁰ "On What There Are," p. 210. Shoemaker goes on to sketch two possible responses to such a worry. The second response, and the one Shoemaker favors when he is feeling klable friendly, "is to deny that klables have the very same properties that tables do." This line of response will be discussed in more detail below.

3. Causal criteria for the genuineness of properties and objects of the sort offered by Shoemaker are intuitively quite plausible. Unfortunately, it is not clear that the causal analysis offered by Shoemaker does anything more than provide an account of why it is that we find certain putative kinds fishy. It is certainly possible that fishy objects like klables exist and are individuated quite independently of the role played by conditional powers in individuating the objects and properties recognized in our common-sense ontology.

In “On What There Are,” Shoemaker decides to take a different approach to the question concerning the role of causality in the individuation of the kinds recognized by science and our common-sense view of the world.

It is, of course, compatible with the [view that there are properties like grue and kinds like klables] that what marks off, from all the rest, the properties and objects we are interested in outside of philosophy, those that are recognized in ordinary and scientific language, is the role causal considerations play in their identity conditions. A way of trying to shed light on these matters is to begin by trying to characterize in some way all of the possible ways of cutting the world up into (putative) things and properties, and then inquiring into what accounts for our finding some of them natural and some fishy. One need not at the outset decide whether fishiness implies unreality; one needn't decide between the view that there are no such things as grueness, klables and incars,²¹ and the more permissive view that their fishiness is just some sort of unsuitability for

²¹ The term ‘incar’ applies to any car which is entirely in a garage or to any segment of any car which inside a garage. Similarly, ‘outcar’ would apply to any car entirely outside of a garage and any segment of any car which is outside of a garage. So, when you drive your car into your garage, an outcar gets smaller and smaller and then ceases to exist and an incar comes into existence and grows until it is the size of your car. The terms ‘incar’ and ‘outcar’ are due to Eli Hirsch. See his book *The Concept of Identity* (Oxford: Oxford University Press, 1982), p. 32.

recognition in our ordinary thought and talk about the world and that they do indeed exist.²²

To this end, Shoemaker considers cutting the world up into (putative) properties and objects in the following way: Let every function from possible worlds to extensions (collections of things in those worlds) “represent” a (putative) property.²³ Let us, following Shoemaker, call these ‘thin properties’.²⁴ Further, let every function from possible worlds to collections of spacetime points “represent” a (putative) object. This is a natural way to go once we recognize that we can represent an object in a possible world by the spacetime points occupied by the object during its career in that world. Let us call these ‘thin objects’.²⁵ Although this way of cutting up the world, making reference as it does to positions in spacetime, rules out there being any non-material kinds such as Cartesian souls, it otherwise allows us to vastly increase the number of putative objects that there are. Shoemaker then distinguishes three different ontological views concerning

²² “On What There Are,” p. 203.

²³ This way of representing properties rules out there being any properties that an object can have at one time but not at another; every object in the extension of a property will have its properties timelessly. Shoemaker states (*ibid.*, p. 205) that the use of slightly more complex functions from possible worlds to sets of ordered pairs consisting of an object and a time (or better an object and a temporal cross-section of its spacetime path) should overcome this difficulty, but uses the simpler functions for the sake of simplicity and ease of explication. I will follow him in this practice.

²⁴ Cf. David Lewis’ distinction between “abundant” and “sparse” properties and relations in *On the Plurality of Worlds* (Oxford: Basil Blackwell, 1986), pp. 59-61.

²⁵ Since thin properties are represented by functions from possible worlds to collections of things in those worlds, once we add the merely thin objects into our ontology, we get even more thin properties.

which of these putative properties and objects may be said to be genuine: the extreme permissive view, the moderate permissive view, and the restrictive view.

According to the extreme permissive view, all of the putative properties and objects represented by the above functions are genuine properties and objects. Although the objects recognized by our ordinary and scientific discourse constitute but a tiny subset of all the things that there are on such a view, it is easy to see why many of the entities do not get recognition in our language. Spacetime is thought to be continuous. There are nondenumerably many functions from possible worlds to collections of spacetime points (and hence nondenumerably many functions from possible worlds to extensions within those worlds which are related in the way the spacetime points we occupy over our careers are related). Many of these functions will yield as value a nondenumerable, randomly scattered collection of spacetime points. It would be impossible for us to refer to such entities (or their properties) by specifying their corresponding function. Further, many of these entities will fail to stand in the kind of causal or explanatory relation to the objects which do get recognized by our ordinary and scientific discourse which would put them within epistemic and semantic reach of creatures like us. Still other objects (and properties) will not get recognized by our ordinary and scientific discourse because of conventions in force in our linguistic community which reflect idiosyncratic facts about our nature or which are freely chosen. It is not hard to imagine, e.g., people very much like ourselves in a world much like our own who, perhaps due to slight differences in their

nature, use terms referring to incars and outcars but have no term referring to cars²⁶ or who use the terms 'grue' and 'bleen' instead of 'green' and 'blue'. It is even easier to imagine that *we* might have picked out just slightly different objects and properties than the ones we in fact pick out, if our freely chosen social and linguistic conventions had been different. For example, we might have decided that mountains begin at the tree line rather than at sea level. In such a case, we would have picked out a different object as, e.g., Mount Everest than the one we in fact pick out.²⁷

So, on this extreme permissive view, there will be a wide range of objects and properties which do not get recognized by our ordinary and scientific discourse. Some of these objects and properties are fishy (such as incars), some are not so fishy (such as the alternate Mount Everest), some (such as the nondenumerable and randomly scattered collections of spacetime points) "hardly get as far as seeming fishy to us."²⁸ According to Shoemaker:

...what determines the position of these entities along this continuum is the extent to which they are governed by causal laws in ways that permit them to stand to us, or to other minds, in the sorts of causal connections involved in knowledge and reference.²⁹

If this is correct, there is still a role for Shoemaker's causal account of the distinction between ordinary and fishy objects and properties. However, it would be wrong of us, on

²⁶ See Hirsch, *The Concept of Identity*, p. 32.

²⁷ For more on the role of conventions in picking out properties and objects, see Chapter Five below.

²⁸ "On What There Are," p. 205.

²⁹ *Ibid.*, p. 207.

this extreme permissive view, to conclude from the fact that an entity fails to be governed by the causal laws necessary for knowledge and reference by us that it is not a genuine existent. On such a view, Shoemaker's causal account of object individuation seems to provide merely criteria for an object's being in reach of a knowing mind, rather than metaphysical criteria for genuine existence.

But perhaps, it may be thought, Shoemaker's causal criteria do point to a genuine metaphysical distinction between the objects recognized by ordinary discourse and the merely thin objects, even if this distinction is deemed irrelevant to the ontological status of the objects. One of the deep insights of "On What There Are" is that even this more modest claim cannot be correct given an extreme permissive ontology and a fairly plausible account of causality. As Shoemaker points out, "[t]hin properties are so plentiful that for every thin property there will be bound to be other thin properties with which it is connected in a quasi-lawlike way."³⁰ If causality is analyzed in terms of constant conjunction of property instantiations that hold in all (or at least some relevant set of) possible worlds,³¹ these "quasi-lawlike" connections between thin properties will be genuine causal connections, and there will be causal connections that hold between the stages of even very fishy objects. If Shoemaker's causal account of property and object

³⁰ Ibid., p. 211.

³¹ To be sure, if this analysis of causality is to be consistent with Shoemaker's causal account of property and object individuation, it cannot be a completely reductive analysis since the notion of a property will itself be analyzed in terms of causality.

individuation is correct, even very fishy objects will turn out to be genuine objects on this modified Humean analysis of causality.

Some will, no doubt, conclude from this unhappy result that the modified Humean analysis of causality that leads to it must be mistaken. A defense of this (or any) analysis of causality is well beyond the scope of this dissertation. However, let me say just a bit to motivate the claim that such a modified Humean analysis of causality is not completely implausible. Whatever the appropriate analysis of causality, it seems eminently plausible that a necessary condition for its being true that A causes B is that in every possible world (or at least certain relevant, nearby worlds) in which A occurs and no preventing conditions obtain, B occurs.³² Such a condition would capture our intuition that causation is not merely a correlation between events in the actual world and would distinguish causal generalizations from merely accidental generalizations. Any account of causation which made reference only to events or property instantiations in the actual world would not be able to capture this intuition. So, a constant conjunction between A and B in all (or some relevant) possible worlds in which A obtains (and no preventing condition obtains) is plausibly a necessary condition for its being true that A causes B. Since it is not at all

³² Here and in the following I take 'A causes B' as entailing that A is (a necessary part of) a sufficient condition for B. Cf. Jonathan Bennett's *Events and their Names* (Indianapolis: Hackett, 1988), pp. 42-46. If causes are ever or only merely necessary conditions for the events that they cause we need to say instead that in all (or some relevant) possible worlds where A does not occur, B does not occur. Similar results follow *mutatis mutandis*.

Those who believe that causality is inherently probabilistic, and that an occurrence of B is only rendered more probable than otherwise given a causal factor A, will be reminded that the analysis of probability will itself likely involve reference to conjunctions of events in various possible worlds.

clear what would need to be added to this necessary condition to make it sufficient, it is not completely implausible to analyze causality entirely in terms of the constant conjunction of property instantiations that hold in all (or some relevant) possible worlds. And if this account of causality is correct, then Shoemaker's causal account of property and object identity would seem to be consistent with an extreme permissive ontology.

Even if we were to restrict the notion of causality to those law-like regularities that are isomorphic to the ones that govern the properties of objects of the sort recognized in ordinary discourse, there will still be merely thin objects that meet Shoemaker's causal criteria for object individuation. In support of this claim, Shoemaker asks us to consider "shadows" of ordinary objects. A shadow of an ordinary object, as defined by Shoemaker, is a merely thin object whose "history" in any possible world has the same spatiotemporal "shape" as the ordinary object in that world, except that its position in spacetime relative to some frame of reference is shifted slightly. For example, there is a shadow of my coffee cup³³ corresponding to the function that differs from the function associated with my coffee cup only in that it takes every possible world in which my coffee cup exists to a collection of spacetime points that has the exact same spatiotemporal shape as the collection of spacetime points corresponding to my coffee cup in the world in question, but that begins five minutes earlier relative to some frame of reference. Indeed, if time and

³³ For ease of explication, I use, e.g., 'coffee cup' here and in the following discussion to refer to a spacetime worm (a four dimensional object occupying three spatial dimensions and a temporal dimension), rather than to a three dimensional object which exists as a whole at a time *t*.

space are continuous (that is, if there are nondenumerably many points in space and time), on the extreme permissive view there are nondenumerably many shadows of my coffee cup corresponding to the nondenumerably many ways the collection of spatiotemporal shapes associated with my coffee cup in each possible world may have been shifted in spatiotemporal position. But now consider the shadows of other ordinary objects arrived at by this way of mapping (occupying in every possible world in which the ordinary object in question exists a position in spacetime five minutes earlier than that of the ordinary object relative to some frame of reference). Call all such objects a 'shadow family'. These objects will have properties determined by functions from possible worlds to the extensions of objects in this shadow family that stand in law-like connections that are exactly isomorphic to the one's that ordinary properties stand in. It would appear that, without a substantially richer notion of causality, the shadow objects postulated to exist on the extreme permissive ontology are just as dependent on causal considerations for their individuation as are ordinary objects, and are thus on a metaphysical par with ordinary objects.

On a causal theory of mind, my shadow has a mind and has exactly the same sorts of mental states as I do....And, of course, I have nondenumerably many such shadows, each belonging to a family of shadows to which it is related just as I am related to the genuine objects in the world. We have no need of Putnam's Twin Earth; each of us has nondenumerably many "dopplegangers" right here!³⁴

³⁴ "On What There Are," p. 212.

Shoemaker points out that thin properties and objects whose lawlike relations resemble the ones that govern the properties of objects recognized by our ordinary discourse would not be limited to shadows of actual entities.

Take any work of fiction you like—say the Sherlock Holmes stories; fill it out so as to be consistent and to be a “complete novel”—i.e., to contain every sentence or its negation—and in such a way that it assigns its fictional events to actual places and times. Then write as many other complete novels as it takes to describe the alternative possible careers of all of the things mentioned in the first complete novel. Then assign these arbitrarily to possible worlds, and consider the set of functions from possible worlds to sets of spacetime points each of whose members takes as its value for each possible world the spacetime path one of these fictional entities would have if the associated novel were a true description of the world.³⁵

Given an extreme permissive ontology, each of these functions from possible worlds to collections of spacetime points represents an object existing in those worlds, one of which we might properly call ‘Sherlock Holmes’.³⁶ Once we have the objects, we can then go on to define functions that will represent the properties of these objects. Since the assignments of complete novels to possible worlds is to be completely arbitrary, the actual world may be assigned any one of these complete novels. Thus, on an extreme permissive view, the actual world will contain shadowy versions of all of the objects to which any of the complete novels refers. All of these objects will be related in various lawlike ways that are, on the modified Humean analysis of causality, in every respect similar to those of the ordinary properties and objects. Further, as Shoemaker points out:

³⁵ Ibid.

³⁶ Shadows of Holmes could then be described relative to this function in much the same way as are shadows of ordinary objects.

...since “complete novels” amount to complete descriptions of possible worlds, each possible world will contain shadows³⁷ of everything that exists in every other possible world—in fact, it will contain as many different shadows of each thing as there are possible worlds in which that thing exists.³⁸

Now the ontology is really getting crowded!

So, despite its initial promise, Shoemaker’s causal account of properties and objects in no way picks out the ordinary properties and objects as a metaphysically privileged subclass of what there is on an extreme permissive ontology. To be sure, there will be some merely thin properties that do differ significantly from the properties picked out by Shoemaker’s causal criteria. The notion of a causally privileged thin property is tied up with the notion of a genuine change on Shoemaker’s account. Thus, what we might call universal and necessary properties, e.g., *being such that $2+2=4$* , and properties which are impossible because contradictory, *being a round square*, will not have causally relevant criteria of individuation. Neither will such relational and historical properties as *being fifty miles south of a burning barn* or *having been slept in by George Washington*. The causal criteria for property individuation can, thus, be seen to pick out a subset of all of the properties that there are on the extreme permissive view. But the only properties that are not a member of this subset are the relational and historical properties and properties corresponding to functions that take every possible world to the universal or

³⁷ Strictly following Shoemaker’s definition, these are not shadows of objects in other possible worlds. They are however, shadow-like counterparts of objects in other possible worlds.

³⁸ “On What There Are,” pp. 212-213.

null set in that world. The subset of properties picked out by the causal criteria of individuation will include, not only more ordinary properties like mass, color, and size, but also the nondenumerably many properties that are the shadow analogue of ordinary properties—hardly a privileged subset at all!

An alternative to the extreme permissive view is what Shoemaker calls the ‘moderate permissive view’. It holds that while the shadow and other funny properties and objects exist, the lawlike uniformities that hold among these properties, even though some are isomorphic to or otherwise resemble the causal laws that hold between ordinary properties, are devoid of genuine causal efficacy.

The suggestion is that just as it is...essential to the ordinary properties that they have certain causal powers, it is essential to causality that it govern *these* properties. The connection between causality and these properties is an internal one. Hilary Putnam has suggested that our natural kind concepts are in a way indexical; to be gold is to have the underlying structure that explains the phenomenal features of *these* things....Perhaps the notion of causality is indexical in a similar way—causality is what underlies the lawlike connections that hold between *these* properties, and any other properties that stand in lawlike connection with them. On this view the introduction or explanation of the notion of causality must involve an ostensive component, a reference to certain paradigms.³⁹

Thus, while the moderate permissive view is still ontologically extremely permissive, it is more moderate than the extreme permissive view in that it does allow that there is a genuine metaphysical distinction between ordinary properties and objects and the shadow-like counterparts thereof. The former have, while the latter, lack genuine causal efficacy. This view requires an explicit rejection of the modified Humean analysis of causality that

³⁹ Ibid., pp. 215-6.

we considered with the extreme permissive view. While there cannot be any way to say in what the relevant causal efficacy consists without making reference to the properties and objects of ordinary discourse, it will be the case on a moderate permissive view that the family of ordinary objects is blessed with causal efficacy while shadow families are not. Although ordinary properties and objects are not ontologically privileged in any way, they do constitute a privileged subclass of thin properties and objects. Before I go on to more fully evaluate the relation between the extreme permissive and the moderate permissive views, let me turn to the third and final view outlined by Shoemaker, “the restrictive view.”

The restrictive view denies that any of the merely thin properties and objects exist. To be sure, even on this restrictive view the set-theoretical entities that correspond to the thin properties and objects postulated by the extreme permissive view exist. It is just that on this view there are no properties or objects that correspond to them; there are just the set theoretical entities, e.g., sets of sets of spacetime points, etc. It is clear that it is this restrictive view that most closely approximates our common-sense ontology, even though Shoemaker believes that, in order to be defensible, the restrictive view must:

...allow enough unrecognized entities to give scope for conventional decision about how to cut up the world, and enough to explain indeterminacy of truth value of identity statements in terms of vagueness and indeterminacy of reference...[and enough to] allow for the entities that would be recognized by possible creatures having quality spaces, and natural ways of “going on,” different from our own....⁴⁰

⁴⁰ Ibid., p. 217. Unfortunately, this would allow klables, incars and other fishy objects back into the ontology if there are possible creatures who would naturally recognize such

Shoemaker claims that all three of the ontological views sketched above require an answer to the question “In virtue of what do objects and properties belong to the same family?” and, more particularly, “In virtue of what do objects and properties belong to the same family as our commonsense properties and objects?”⁴¹ An answer to the second question should allow us to determine which things and properties are admitted into the range of the variables in our ordinary and scientific discourse and provide us a way for distinguishing these more ordinary objects and properties from fishy ones. Shoemaker answers the question this way:

The entities that are candidates for belonging to the range of our variables are, to begin with, we ourselves, the things in our environment to which we have our most direct epistemic access, and the properties of ourselves and those things to which we have direct epistemic access.... That we have direct epistemic and semantic access to these things of course involves there being a mesh between their causal nature and ours. Our “nature” here includes our “quality space” (à la Quine), whatever we bring to the world (à la Kant) in the way of innate conceptual structure, and our propensity to “go on” in certain ways (à la Wittgenstein) in applying language and concepts to the world.⁴²

Such an answer allows much of what Shoemaker wants to say about the role of causality in the individuation of ordinary properties and objects to play an important role in our

objects, given their quality spaces and natural ways of “going on.” It does, however, rule out the shadow properties and objects and many of the other mere thin properties and objects.

⁴¹ “On the extreme permissive view we need an answer to it...in order to say what things are admitted to the range of variables of quantification. On the moderate permissive view we need to answer it in order to say which the things are that have genuine causal efficacy. On the restrictive view we need to answer it in order to say what is required for the existence of genuine substances and properties”(ibid., p. 218).

⁴² Ibid., pp. 218-9.

epistemology and theories of reference, if not in our metaphysics. But such an answer, if correct, seems consistent with and available to each of the three views. Shoemaker wants to be able to claim that the restrictive view is correct, and to deny that there are shadow entities, and the like, but he is unable to come up with a conclusive argument to that conclusion.⁴³ Unfortunately, despite its initial promise, Shoemaker's causal account of property and object individuation can be of no help in providing reasons for favoring the restrictive view, as the causal criteria of property and object individuation are consistent with all three of the ontologies sketched above.

In what follows, I shall first argue that the moderate permissive view is a non-contender in that it will simply collapse either into the extreme permissive view or the restrictive view. Second, I shall argue that the extreme permissive view can be rejected by the proponents of many forms of antirealism. Third, I shall provide reasons that suggest that the metaphysical realist will have a significantly more difficult time rejecting the extreme permissive view. I will then argue that this fact, much like the fact that the metaphysical realist has a more difficult task in rejecting Cartesian skepticism than does the antirealist, will provide us with motivation for questioning the presumption of metaphysical realism.

⁴³ "It appears that if the question whether these things exists has an answer, it ought to be decidable *a priori* what the answer is. I would like to have a decisive *a priori* proof of the restrictive view, but I do not have one"(ibid., p. 220).

4. Are there any reasons to which the metaphysical realist (or anyone) can point that will allow him to reject the extreme and moderate permissive views? Despite Shoemaker's stated skepticism that reasons for preferring or rejecting any of the three ontological views that he outlines can be found,⁴⁴ it seems to me that the moderate permissive view can be ruled out from the very beginning. It is not so much that I have a decisive argument against the moderate view. My reason for ruling out the moderate permissive view is instead that there is not really a distinct view there at all. It seems to me that the moderate permissive view necessarily collapses into either the extreme permissive or the restrictive view. The reader will recall that on the moderate permissive view shadow entities and other merely thin properties and objects exist, but are held to be causally inefficacious. Their lack of causal efficacy is attributed not to a failure to be governed by law-like regularities (they are so governed on this view), but rather to the fact that the notion of causality is in a certain way indexical. Causality is what underlies the law-like regularities that hold between the properties that are capable of producing beliefs in knowing minds like our own. That is, only properties that are within our epistemic and semantic reach can be said to be causally efficacious.

The fact that there is an internal connection between causality and ordinary properties and objects is supposed to entail that such objects do constitute a metaphysically privileged subclass of the nondenumerably many thin properties and objects. This is what is supposed to distinguish the moderate from the extreme permissive

⁴⁴ Ibid., pp. 219-221.

view. But is this really true? Consider one of my many shadows; let us call her 'Kleborah'.⁴⁵ Having read a shadow article by a prominent philosopher in this shadow family, Sydney Kloemaker, entitled "On What There Are," Kleborah decides that, although she much prefers the restrictive view presented by Kloemaker, she has no reason to quarrel with the so-called moderately permissive view that he sketches. After all, her family of entities is still reserved a privileged position in an ontology otherwise crowded with thin properties. Only the properties and objects that are sufficiently related to knowing minds like hers can be said to have causal efficacy. She goes to sleep relieved that, even if there are no good reasons for deciding between the three views explicated by Kloemaker, at least two of them are friendly to her intuition that she and the rest of the objects in her family are in some way metaphysically privileged, and, all things considered, two out of three odds is not so bad in the metaphysics business. It seems that as long as the moderate permissive view asserts that there really are all the shadow properties and objects such a story is perfectly coherent. After all, if we can treat the notion of causality as indexically linked to the properties to which we can have epistemic and semantic access, why can't the shadow entities? However, if this is right, then, contrary to what the moderate permissive view would seem to entail, our ordinary properties and objects are

⁴⁵ Actually, given the way Shoemaker develops the notion of a shadow entity, Kleborah is not really one of my shadows. All of my shadows "do" just exactly what I do. Kleborah's career is, however, identical with one of my possible careers. That is, there is a possible world where I do just exactly what Kleborah does. But as we saw above, each possible world contains shadow-like counterparts of everything that exists in every other possible world. Kleborah is one such shadow-like counterpart in the actual world of one of my possible selves.

not *metaphysically* privileged in any way. Ordinary objects are privileged on this view only relative to us, and our ways of knowing; shadow properties and objects are privileged relative to the various shadow people and their ways of knowing. But this was true even on the extreme permissive view! So, if we take the claim that causality is indexical seriously, the moderate permissive view can be seen to collapse into the extreme permissive view.

Perhaps a proponent of the moderate permissive view will want to reply that we are leaning too heavily on the claim that the notion of causality is in a certain way indexical. It is not that causality is indexical in the sense that utterances of ‘causally efficacious’ make reference to different families of properties depending upon the family to which the utterer belongs. Rather, the point is that there is some genuine metaphysical property that underlies all and only the law-like regularities holding between ordinary properties and objects, but that this property, causal efficacy, cannot be specified beyond stating that it is the property that underlies *these* properties and objects. But if shadow entities exist, why should we not think that there is a genuine metaphysical shadow property that underlies the law-like regularities that hold between the properties in their families? The only reason that I can see for denying that there are genuine metaphysical properties, over and above the mere holding of law-like regularities, that are shadow variants to causality (if not causality itself) and that hold between the properties internal to shadow families is that the existence of the shadow entities is not on an ontological par

with that of ordinary entities.⁴⁶ Perhaps, unlike ordinary objects, shadow objects are not material entities. But such a view seems to me to be no different from the restrictive view. Even the restrictive view grants that shadow properties and objects exist qua set-theoretic entities. If causality is viewed as a genuinely indexical notion, then the difference between the moderate and extreme permissive views is no more than a semantic squabble. Similarly, any reasons we might have to believe that causality is more than a merely indexical notion, will be reasons for believing that, despite the *prima facie* ontological differences, the difference between the moderate permissive and the restrictive views is nothing more than a semantic squabble. I would submit that there are only two ontological views that can really be distinguished given Shoemaker's way of dividing the world into so-called thin properties and objects. The moderate permissive view is a non-starter.

Now that we have found reason to reject the moderate permissive view, let us examine whether there are reasons for rejecting the extreme permissive view (hereafter 'EPV') and accepting the restrictive view by default. The EPV will be incompatible with at least some forms of metaphysical antirealism. The arguments above (Chapter Two) which show that the Cartesian skeptical hypothesis can be rejected as false by the proponent of the rational acceptability theory of truth, and as highly unlikely to be true by the proponent of an idealized rational acceptability theory of truth can be applied *mutatis*

⁴⁶ Cf. Lewis' defense of the claim that 'actual' is an indexical term that refers to the world of utterance in *On the Plurality of Worlds*, pp. 92-94 which bears certain similarities to the point made here and in the previous paragraph.

mutandis to the claim that either of the permissive ontologies is correct. The claim that there are families of shadow objects distinct from our own is similar to the Cartesian skeptical hypothesis that ours is a world inhabited entirely by Cartesian souls in that both entail that the truth concerning what there is is ineluctably beyond our epistemic access. If there are families of shadow objects different from our own, then the individuals that belong to these families are *ex hypothesi* both causally and explanatorily isolated from our family of objects. Since it is a plausible thesis that an object of empirical knowledge must stand in some causal relationship to us or in either a causal or explanatory relation to other properties and objects which do, individuals belonging to families of shadow objects different from our own (if indeed there are any) are ineluctably beyond our epistemic access.⁴⁷ In different ways, the metaphysical antirealist views we have considered so far tie truth to epistemic accessibility. It is for this reason that they can provide reasons for rejecting any thesis that entails that the truth about what there is is ineluctably beyond our epistemic access. Since the EPV asserts that there are numerous shadow families of entities distinct from our own (as well as other very fishy objects), given the truth of a metaphysical antirealism that ties truth to epistemic accessibility, we *do* have reason to prefer the restrictive view to the EPV.

⁴⁷ Even if there are philosophical reasons that seem to favor the EPV so that the shadow families are not completely beyond our *a priori* access (and we will see below that there may be such philosophical reasons), the *individuals* that belong to these shadow families would still be beyond our epistemic access in the sense that it would be impossible for us to know any of the contingent facts about them.

But what of the metaphysical realist? The reasons given in the previous paragraph for rejecting the EPV are not available to him.⁴⁸ Are there other reasons for rejecting the EPV to which he can appeal? Some might find the highly counter-intuitive nature of the EPV to be sufficient reason for rejecting it. Shoemaker sketches such an objection and offers a reply to it on behalf of a proponent of the extreme permissive view.

It may seem that the proponent of the extreme permissive view is committed to denying things that any sensible person would assert. For example, any sensible person would assert that no one has ever done all of the things that Sherlock Holmes is described as doing in the Arthur Conan Doyle stories, yet it would seem from what I said earlier that the extreme permissive view is committed to there being, in this very room, nondenumerably many people who have done those things. But the proponent of the extreme permissive view has a ready reply to this. Following the lead of David Lewis, he can say that in our ordinary and scientific existence claims we often implicitly restrict the range of our variables so that it does not include all of the things there are.⁴⁹

So, it would seem that the extreme permissive view can, in some sense, be reconciled with our commonsense intuitions about the range of the variables in our ordinary and scientific

⁴⁸ Shoemaker himself seems to acknowledge that the EPV is a particular problem for the metaphysical realist. “[I]t seems at least to some extent a matter of convention how we cut up the world into properties and things. Since the adoption of conventions can hardly be thought to *create* the entities to which it gives recognition, it would seem that the role of convention can only be that of selecting from a set of preexisting entities certain ones to give linguistic recognition to. And if we think that there are still other ways of cutting up the world that are not available to us, because of the way we are built, but which would be available to possible creatures whose sensory and cognitive constitutions are different from our own, then it seems that we must allow that there are the entities such creatures would or might recognize—we *must allow this as the cost of rejecting the idealist view that what exists is mind dependent*....[I]t may seem that we have opened the floodgates and that we cannot stop short of adopting the permissive view”(“On What There Are,” p. 209, my italics).

⁴⁹ Ibid., pp. 213.

discourse. However, Shoemaker is quick to point out that while it may not be true to say in our ordinary and scientific discourse that such entities exist, it is nonetheless highly counter-intuitive to suppose that, so long as we do not restrict the range of our variables, they can be truly said to exist.

Despite the availability of these replies, it still seems to me outrageous to maintain that, quantifying over absolutely everything, there are in the actual world nondenumerably many doppelgangers of all of the ordinary objects plus nondenumerably many instantiations of every individual concept, and that *sub specie aeternitatis* the entities we recognize in our ordinary and scientific dealings with the world do not have a special status.⁵⁰

Even if the proponent of the EPV cannot reconcile his view completely with our intuitions, it does not seem to me that this fact can be used all by itself as a *reductio* of the view. After all, several hundred years ago, any reasonable person would have found it outrageous to believe that there were such strangely behaved entities as the quarks and electrons postulated by quantum mechanics. As quantum theory is well confirmed today, it is clear that any supposed *reductio* of an ontology including quarks and electrons based on its counter-intuitive nature would have been fallacious. Similarly, such a *reductio* of the EPV would itself be fallacious. Our finding an ontology outrageous may provide us with a pragmatic reason for rejecting it. But, it does not seem to be the sort of reason that is truth-conducive. This seems to be even more evidently the case if metaphysical realism is true. Why should our finding an ontology outrageous in and of itself make it more likely that the ontology is incorrect, if the existence and features of the objects that make

⁵⁰ Ibid., p. 214.

up the correct ontology exist and have the features that they do independently of any and all minds? Why should our intuitions have any bearing on reality if metaphysical realism is true? The metaphysical realist will require some further reasoning if he is to reject the EPV.

Perhaps the metaphysical realist will want to turn the question on its head. Instead of asking “do we have any cogent reasons for thinking that the EPV is false?,” he will ask “Do we have any cogent reasons for thinking that the EPV is true?” If we ignore the skeptical worries raised in the previous chapter, it would seem that we have good reasons for thinking that the objects belonging to our family and the properties that they have exist. But, why would we even be tempted to think that there were such funny objects as the shadows and their ilk? Shoemaker himself provides some considerations in support of the EPV.⁵¹ *If* any of these reasons are sound, then the metaphysical realist is in a very desperate position indeed. However, the metaphysical realist may find none of these arguments terribly convincing. In every case, it should be possible for the metaphysical realist who is not otherwise committed to the premises of the argument to run the argument for the EPV backward, assuming that there are none of the funny objects postulated by the EPV and rejecting the premises that suggested that there were.

Unfortunately for the metaphysical realist, one of the arguments in support of the EPV has premises which are fairly plausible. It is fairly plausible to view properties, à la

⁵¹ See *ibid.*, pp. 208-9 and pp. 219-20.

Lewis,⁵² as sets of possibilia. On such a view, a property is reduced to a set of objects, the set of objects which instantiate the property in any possible world.⁵³ Such a view is *prima facie* quite plausible in that the rejection of the Lewisian account of properties involves embracing some fairly controversial theses. As Shoemaker puts it,⁵⁴ “if we have possibilia anyhow, it is metaphysical excess to suppose that in addition to sets of possibilia we have properties—and that gives us a reason for reducing properties to sets of possibilia....”

In a footnote on p. 76 of *On The Plurality of Worlds* Lewis provides similar reasons for holding the view that the career of a material object in a possible world⁵⁵ (or if one is a proponent of a perdurance theory, the material object itself) may be reduced to a set of spacetime points, the points occupied by the object during its career.⁵⁶

There are three different conceptions of what the spatiotemporal relations might be. There is the dualist conception: there are the parts of spacetime itself, and there are the pieces of matter or fields or whatnot that occupy some of the parts of spacetime.... There are two simpler monistic conceptions. One of them does away with the occupants [of spacetime] as separate things.... The properties that we usually ascribe to occupants of spacetime... belong in fact to parts of spacetime themselves. When a part of spacetime has a suitable distribution of local properties, then it is a particle, or a piece of a field, or a donkey, or what have you.... The other monistic conception does the opposite: it does away with the parts of spacetime in

⁵² See Lewis, *On The Plurality of Worlds*, pp. 50-69.

⁵³ Appeal to other possible worlds is required to rule out accidentally coextensive properties (the standard examples are the properties of *being a creature with a heart* and *being a creature with a kidney*) counting as one and the same property.

⁵⁴ “On What There Are,” p. 219.

⁵⁵ We need to appeal to other possible worlds here in order to accommodate the intuition that most (if not all) material objects could have had different careers.

⁵⁶ Such a view is accepted by Lewis, at least for the actual world, and discussed sympathetically by Jonathan Bennett.

favor of the occupants (now not properly so called)....I tend to oppose the third conception, at least as applied to our world, for much the reasons given in Nerlich, *The Shape of Space*.⁵⁷ I tend, more weakly, to oppose the dualist conception as uneconomical.

The idea is that, if we have reason to believe in spacetime and parts thereof, then it is metaphysical excess to suppose that, in addition to properties instantiated in spacetime, we have objects which occupy the spacetime and are the genuine bearers of the properties. And to paraphrase Shoemaker, that gives us a reason for reducing objects to zones of properties instantiated in spacetime.

A Lewisian property, then, is a function from possible worlds to sets of sets of spacetime points.⁵⁸ Once we have spacetime points and Lewisian properties instantiated in space and time, we get material objects. The career of a material object (or the object itself on the perdurance view) is a “zone”⁵⁹ of properties instantiated in space and time; any point-sized temporal slice of such a zone is an object (or part of an object). If we believe in spacetime points or minimal spacetime zones, and we believe in sets of other things we believe in (with suitable qualifications for avoiding Russell’s Paradox), then we seem to get all of the thin properties, and with them, all of the thin objects postulated by the EPV. Without something to stop it (such as the objections to the EPV available to the metaphysical antirealist), this way of viewing properties and objects would seem to lead

⁵⁷ Graham Nerlich, *The Shape of Space* (Cambridge: Cambridge University Press, 1976).

⁵⁸ As Lewis points out, on one standard use of the term ‘property’ each such function from possible worlds to sets of sets of spacetime points will be property, “because for any set whatever, there is the property of belonging to that set” (*On the Plurality of Worlds*, p. 60).

⁵⁹ See Bennett, *Events and Their Names*, pp. 12-13.

inexorably to the EPV. So, the metaphysical realist, it seems, can reject the EPV only by developing adequate reason to give up a not obviously false account of properties and objects and to embrace (one or the other of) the metaphysical excesses seemingly entailed by the rejection of such an account.

Some will, no doubt, want to argue that, while (the careers of) ordinary objects can be viewed as represented by functions from possible worlds to sets of spacetime points, not every such function represents an object. Only those zones of spacetime whose point-sized temporal slices (or stages) are causally connected in the way the stages of (the careers of) ordinary objects are connected count as (the careers of) genuine objects. Thus (the objection goes), we can accept the Lewisian view of properties and the reductive account of objects (thereby avoiding the controversial dualisms) but reject the EPV; there may be all of the thin properties, but there aren't all of the thin objects.

However, we have already seen enough to say why it is very plausible that such a response won't work. While it certainly isn't the only way to provide an account of the necessary causal connections between the temporal stages of a putative object required for its counting as a genuine object, we have already seen a very plausible account:

Shoemaker's account in "Identity, Properties, and Causality." And unfortunately, we have already seen why Shoemaker's causal account of the individuation of properties and objects will not provide the metaphysical realist (or anyone else for that matter) with a justification for denying the EPV.⁶⁰ Once we see why Shoemaker's account fails, it is

⁶⁰ Above, pp. 75-81.

hard to see how the adoption of any other causal account of continuant individuation would provide the metaphysical realist with any better justification for rejecting the EPV.

As we saw above, it is a fundamental part of our commonsense picture of the world that we are justified in believing that we are not radically mistaken about what kinds of things may be said to have genuine ontological status, and hence, that we are justified in rejecting the EPV. We have seen that, if metaphysical antirealism is true, then we are so justified. If the metaphysical realist is to be accorded the default position in the debate in virtue of the fact that his view seems to accord with more of our commonsense intuitions about the nature of the world and our knowledge of it, then he too must show that we are justified in rejecting the EPV and denying the existence of these funny objects. However, it would appear that the metaphysical realist must embrace some fairly controversial theses if he is to do so, and again we see a respect in which metaphysical realism is not the completely intuitive view it first appeared to be.

In Chapter One, I suggested that there were at least three crucial assumptions that underlie our commonsense intuitions about the world: (i) that we are not radically mistaken about what kinds of things have genuine ontological status, (ii) that their existence and at least some of their features are mind-independent, and (iii) that we can and do succeed in making reference to these entities and their features. We saw that it is because metaphysical realism accords so well the second intuition that it is commonly thought to be the default position in the debate. However, we have now seen that there is

reason to suspect that the metaphysical realist will have, while the metaphysical antirealist will not have, considerable difficulty showing that the first of our intuitions that we are not radically mistaken about the kinds of entities which have genuine ontological status is justifiably held. In Chapter Two, the threat came from the Cartesian and Kantian skeptical arguments, each of which suggests that we might be mistaken that the objects recognized by our commonsense and scientific world-views exist and have most of the features that we take them to have, and hence, are not justified in believing them to exist and have most of the features we take them to have. In this chapter, we have seen that even if we are not mistaken about the existence and features of the entities recognized by our commonsense and scientific discourse, there might be, for all we know, many more entities in our world. Indeed, our ontology might be so crowded that the entities which we recognize in our commonsense and scientific discourse make up an insignificant subset of all of the things that there are. In this case too, our intuition that we are justified in believing that we are not radically mistaken about what kinds of things there are would be false. In the next chapter, I will examine an argument offered by Hilary Putnam which suggests that semantic realism, reference to a mind-independent reality, is problematic and, hence, that the metaphysical realist will have difficulty reconciling his view with the third of our basic intuitions about the world.

CHAPTER FOUR: A PROBLEM FOR METAPHYSICAL REALISM: PUTNAM'S ARGUMENT AGAINST SEMANTIC REALISM

We pre-reflectively believe that we are not radically mistaken about what objects and properties have genuine ontological status. But it is an equally basic belief that we can and sometimes do succeed in referring to objects with genuine ontological status. (Indeed, if none of our terms refer to entities with genuine ontological status and we are still capable of having beliefs about what does and what does not have genuine ontological status, it is hard to see how these beliefs could be anything other than radically mistaken.) Suppose that our commonsense and scientific ontologies are largely correct; that is, suppose that tables (but not klables!), cats, black holes, electrons, etc. have genuine ontological status. Suppose further that all of these objects exist and have (many of) the features that they do independently of the existence and activities of our or any other minds. As we saw in Chapter One,¹ it would be possible in a situation such as this for our terms to fail to be related to these objects in the way relevant for establishing reference. Metaphysical realism alone, even when combined with an acceptable ontology of metaphysically real entities, is not sufficient to fully satisfy our basic intuitions about the world and our semantic and epistemic relations to it. Somehow or other, our term 'cat' must refer to mind-independent cats, our term 'electron' to mind-independent electrons, etc. In order to satisfy this intuition, the metaphysical realist must adopt some form of semantic realism.

¹ See above, p. 25.

Semantic realism is the view that (i) (many of) our terms refer to mind-independent objects and their mind-independent features and (ii) (many of) our sentences assert that mind-independent states of affairs obtain. On such a view, a sentence will (generally) be true if and only if the mind-independent state of affairs which the sentence asserts to obtain actually does obtain. While semantic realism may at first blush seem relatively unproblematic, Hilary Putnam has argued that it is far from clear that it is possible to refer to mind-independent objects. In this part of the chapter, I will present and examine Putnam's argument against semantic realism. I will also raise and respond to several semantic realist rejoinders to the argument. This discussion will reveal, once again, that the metaphysical realist will have difficulty (and at least some varieties of metaphysical antirealism will not) accommodating some of our most basic intuitions about the world and our place in it. Again, we will have reason to question the presumption of metaphysical realism.

1. Putnam has presented two distinct versions of the argument against semantic realism. The first version of the argument first appeared in "Realism and Reason"² and later in "Models and Reality."³ The second version of the argument appears in chapter 2

²*Proceedings and Addresses of the American Philosophical Association* 50 (1977), pp. 483-498, reprinted in *Meaning and the Moral Sciences* (London: Routledge and Kegan Paul, 1978), pp. 123-140. Page references to this work refer to *Meaning and the Moral Sciences*.

³ *Journal of Symbolic Logic* 45 (1985), pp. 464-482.

and the appendix of *Reason, Truth, and History*.⁴ However, both involve premises supported by various model-theoretic tricks and both are susceptible to roughly the same objections. The formulation of the argument against semantic realism which I will present and examine will more closely resemble the argument from “Realism and Reason.” While the outline of the argument is my own, my thinking on the structure of the argument has been greatly influenced by David Anderson’s “What is the Model-Theoretic Argument?”⁵

As I see it, Putnam’s argument against semantic realism is as follows:

1. If semantic realism is true, then it is possible for an epistemically ideal theory T to be false.
2. If theory T is consistent, then (given certain restrictions on the size of the world) there will be a model M of T on which it is true of the world.
3. There is nothing which could fix a determinate relation of reference between terms of T and parts of the world and so rule out M’s counting as the correct or intended interpretation of T.
4. So, it is not possible for an epistemically ideal theory to be false on its intended interpretation.
5. Thus, semantic realism is false.

While this argument in its entirety is frequently referred to in the literature as “the model-theoretic argument,” the model-theoretic argument proper, as I see it and as I will use the term, is actually the sub-argument in support of premise 2. In the remainder of this

⁴ (Cambridge: Cambridge University Press, 1981), pp. 22-48; 217-18.

⁵ *The Journal of Philosophy* 90 (June 1993), pp. 311-322.

section, I will present what I take to be the basic thrust of the model-theoretic argument in support of premise 2 of the argument above.⁶

Consider an epistemically ideal⁷ theory T, thought of as a mere string of symbols in the notation of first order logic. Since T will be syntactically consistent, by the Skolem Löwenheim theorem it will have a model (an interpretation on which it comes out true) in which its variables range over the domain consisting of the natural numbers. Call this model A. On model A, theory T is true of the natural numbers (or a subset thereof). We can then (at least in principle) map the natural numbers onto objects in the mind-independent world.⁸ We can use this mapping to define a new model, call this model W, whose domain is the set of objects in the world and which assigns various sets of these objects to the monadic predicates of T and sets of ordered n-tuples of these objects to the

⁶ In Putnam's presentation and in much of the literature, the reasoning in support of premise 2 is included along with the main argument against semantic realism and is what gives it its common name in the literature. However, premise 2 is by no means the most controversial premise in Putnam's argument. As we will see below, it is premise 3 which the semantic realist will want to question and whose acceptance Putnam must motivate. Putnam's argument in favor of the crucial premise 3 takes the form of an argument by elimination and is not as easily presented along with the main argument. The remaining sections of this chapter are largely devoted to an examination of the support for premise 3.

⁷ Actually, for the sub-argument to go through, it is not necessary for T to be epistemically ideal; it is sufficient that T be syntactically consistent.

⁸ At least we can do this if, as seems plausible, the world contains an infinite number of objects or can be broken into an infinite number of pieces. If the world is finite, T will have no model on the world (and hence will not be true of the world) if T has only infinite models or if the smallest finite model of T is a model whose domain is larger than the number of objects in the world. A proponent of semantic realism who is willing to claim that the world can only be broken into a finite number of pieces and that T will have no finite model whose domain is the same size or smaller than the set of objects in the world can reject the conclusion of the model-theoretic argument.

n-ary predicates of T in such a way that a sentence Φ is true in W if and only if Φ is true in

A. Model W will be an interpretation on which theory T is true of the mind-independent world.

In order to see more clearly what is going on in the argument above, it will be useful to consider a very simple application of the basic idea of the argument. This example has been borrowed from Alan Musgrave in an unpublished paper, “What’s Wrong with the Model-Theoretic Argument Against Realism?” Suppose that the actual world contains exactly ten individuals:

i1 is a white swan	i6 is a table
i2 is an armchair	i7 is (yet another) cat
i3 is a cat	i8 is a fireplace
i4 is a lampshade	i9 is a black raven
i5 is (another) cat	i10 is a phonograph

Now, consider the theory that all ravens are white. This theory can be formalized in the notation of first order logic as $(x)(Rx \rightarrow Wx)$. It will then be true just in case the extension of predicate R is included in the extension of predicate W . On the standard interpretation I of the predicates of the theory (where ‘ R ’ stands for ravens, and ‘ W ’ stands for white things), the extension of R will be $\{i9\}$ and the extension of W will be $\{i1\}$ (provided of course that none of the other objects is white). Thus, it is false, on interpretation I , that the extension of R is included in the extension of W . This is not surprising. After all, $i9$ is a black raven and would seem to serve as a counterexample to our theory.

However, since the theory, now thought of simply as a string of symbols in the notation of first order logic, is consistent, it will have a model A in which it is true of the first ten natural numbers (1, 2, 3,...,10). Let A be the interpretation that assigns the prime numbers greater than 2 to predicate R, and the set of odd numbers to predicate W. So, the extension of R in A will be {3, 5, 7} and the extension of W in A will be {1, 3, 5, 7, 9}. Thus, it will be true on interpretation A that the extension of R is included in the extension of W. Thus, the theory will be true under this interpretation. This alone is not a threat to the realist, since the theory on this interpretation is now true of the natural numbers (or an initial segment of them) and not of the objects of the world.

However, once we have obtained an interpretation on the natural numbers which makes the theory true, we can map this back on to the objects of the world. Let us map the first ten natural numbers onto the objects in the world in the following, quite natural way:

$$\begin{array}{lll} 1 & \Leftrightarrow & i1 \\ 2 & \Leftrightarrow & i2 \\ & \vdots & \\ 10 & \Leftrightarrow & i10 \end{array}$$

This allows us to define an interpretation M on the objects of the world wherein the extension of R is {i3, i5, i7} and the extension of W is {i1, i3, i5, i7, i9}. Now the extension of R in M is included in the extension of W in M, so the theory is true under this interpretation. Since the domain of this interpretation is the objects of the world, the

theory under M is now true of the world.⁹ Since it will always be possible to apply such model theoretic tricks to a theory, there will always be an interpretation under which the theory is true of the world.¹⁰ Premise 2 of the argument for semantic realism is true.

However, it would appear that the semantic realist can cheerfully acknowledge the truth of premise 2 of the argument against his view. The semantic realist will want to claim that while there will be (or will likely be) an interpretation M whose domain is objects in the world and which satisfies any theory thought of as a mere string of symbols, since this interpretation is not the *intended* interpretation of the theory, this fact is irrelevant. Premise 2 does not, all by itself, entail that an epistemically ideal theory must be true on its intended interpretation. So, the obvious place for a semantic realist to push on the argument against his view is to question premise 3, the claim that there is nothing which can single out a determinate relation of reference and thereby rule out M's being the intended interpretation. Unless Putnam can provide some motivation for accepting premise 3, it open to the semantic realist to simply deny it. In the sections that follow, I will examine Putnam's arguments in support of this crucial premise. But, before I do, I

⁹ It is interesting to note that Musgrave has set up the original example in such a way that this new interpretation assigns all the cats to R and all the animals to W. So, in effect, the theory under this new interpretation says all cats are animals rather than all ravens are white. However, the new interpretation that results from the mapping need not yield a theory that makes any sense at all to us. All that is important is that it makes the theory true—true of the world.

¹⁰ Provided, of course, that the world is infinite or that there is a finite arithmetical model which satisfies the theory whose domain is no larger than the set of objects in the world.

will first briefly explore the possibility that a semantic realist could deny premise 1 of the argument.

2. Why should a semantic realist accept premise 1, the claim that if semantic realism is true, then it is possible for an epistemically ideal theory T to be false? Recall that semantic realism is the view that (i) (many of) our terms refer to mind-independent objects and their features and (ii) (many of) our sentences assert that mind-independent states of affairs obtain. A sentence under its intended interpretation will be true if and only if the mind-independent state of affairs which the sentence asserts to obtain actually does obtain. An empirical theory under its intended interpretation can be thought of as a conjunctive sentence that asserts that the world contains various objects (and perhaps kinds of objects) which have various (basic) properties and stand in various (basic) relations to each other. The theory will not merely say that there are some objects or other with some properties or other and which stand in some relations or others to one another; it will specify *which* objects there are, *what* (basic) properties they have, and *what* (basic) relations they stand in to one another. What makes an empirical theory true or false, according to the semantic realist, is the obtaining or failing to obtain of a (conjunctive) state of affairs in the mind-independent world. While the fact that a theory is epistemically ideal will justify a belief that it is true, this fact is not what makes the theory true (if it is true). This point is sometimes made by saying that on a semantic realist view, truth is non-epistemic. So, unless ideal epistemic justification for empirical theories

is a logically sufficient condition for truth, it will be possible for an epistemically ideal theory to be false.¹¹ Since the existence of epistemically ideal conditions is not a logically sufficient condition for the truth of a theory, according to the semantic realist, premise 1 of the argument against semantic realism has to be true.

As I mentioned above, what a semantic realist should say at this point is that, while Putnam's model-theoretic argument shows that we can artificially produce an interpretation M of the language of T under which T comes out true of the world, since M is (very likely) not the correct or intended interpretation of T, the mere existence of interpretation M does not entail that T is true. It is under its correct or intended interpretation that an empirical theory (epistemically ideal or otherwise) must correspond to the world if it is to be true rather than false. And, or so the response would go, on its intended interpretation it is possible for T to be false. This response relies on there being a correct or intended interpretation of T that is (often) distinct from the interpretation generated by the model-theoretic tricks, and thus, amounts to a denial of premise 3 of the argument against semantic realism.

Putnam realizes that he must say something in support of the crucial premise 3 if his argument is to have any force. To this end, he examines various ways the semantic realist might go about attempting to provide a specification of the intended interpretation of

¹¹ Of course, some false theories are better than others. It only takes one false conjunct to make a theory false. If any false conjuncts are fairly peripheral claims and sufficiently many of the conjuncts which comprise the theory's central tenants are true, a false empirical theory may be very useful. Indeed, such a theory may even be of more use than a weaker, but entirely true theory.

a sentence. In each case, he argues that the semantic realist will not be able to provide a substantive and plausible theory of reference which would somehow determine the intended interpretations of our sentences and fix the extensions of our terms. In the remainder of this section and in the two sections that follow, I will examine and elaborate upon Putnam's arguments in support of premise 3.

At first, it may seem that all that is necessary for specifying the intended interpretation of a sentence is our very intention to specify *that* interpretation rather than some other one. However, as Putnam points out,¹² if I am able to intend that my terms (or even my mental states) pick out some object, I must antecedently be able to think about that object. Since I can intend that a term refers to *x* only if I can already think about, and so refer to *x*, I can't fix the interpretation of a theory by intending that its terms have such and such reference.¹³ So, how do our terms get hooked up with mind-independent objects and properties in a way sufficient for determinate reference to those objects?

One way a semantic realist might answer this question is to adopt what Putnam calls a "magical theory of reference." A magical theory of reference, according to Putnam, is one which holds that certain mental representations are intrinsically intentional; they intrinsically refer to objects in the mind-independent world. The intrinsic intentionality of such a mental representation, on this view, may be independent of how

¹² *Reason, Truth, and History*, p. 2.

¹³ Unless, of course, certain mental states are intrinsically intentional. As we will see below, Putnam considers such a response unacceptable.

the representation was caused and how the thinker is disposed to use it.¹⁴ On such a theory of reference (which I will hereafter refer to as the ‘intrinsic intentionality theory of reference’), the question of how our terms get hooked up with mind-independent reality is a misguided question. If one assumes that the relation between a mental representation and what it represents is contingent, one may very well be puzzled as to how mental representations could hook up to reality. But this assumption is false on such a view. The relation between a mental representation and what it represents is a necessary connection. There is no puzzle. The appearance of a problem stems from the fact that the terms of a natural language (thought of purely syntactically) are only contingently related to their referents—we might have used the term ‘cats’ to refer to cherries. But, what goes for natural languages in this case does not go for thought. Mental representations are not mere syntactic symbols which may be conventionally used to refer to one thing rather than another; they represent or refer intrinsically. If this is the case, we can identify the correct or intended interpretation of the language of an epistemically ideal theory T (the interpretation under which it is possible for T to be false) by reference to T’s translation in the language of thought, thereby denying premise 3 of the argument against semantic realism.

¹⁴ Laurence Bonjour has pointed out that while the intrinsic intentionality of a mental representation on this view doesn’t result from a thinker’s dispositions to use the representation in a certain ways, it would presumably affect his dispositions to do so, and thus not be entirely independent of them. Even so, the two are independent in the sense that it is possible that a thinker never makes use of a mental representation in the presence of or in connection with the thing it represents even though there are such things and he does make use of the representation.

Putnam acknowledges that adopting an intrinsic intentionality theory of reference can allow the semantic realist to deny premise 3 of his argument. However, he is inclined to think that this advantage of an intrinsic intentionality theory of reference is not worth the cost.

Suppose we assume a ‘magical theory of reference’. For example, we might assume that some occult rays—call them ‘noetic rays’—connect words and thought-signs to their referents. Then there is no problem. The Brain in a Vat [or any other language user] can think the *words* ‘I am a brain in a vat’, and when he does the word ‘vat’ corresponds (with the aid of noetic rays) to real external vats and the word ‘in’ corresponds (with the aid of noetic rays) to the relation of real spatial containment. But such a view is obviously untenable. No present day philosopher would espouse such a view. It is because the modern realist wishes to have a correspondence truth *without* believing in ‘noetic rays’ (or believing in Self-Identifying Objects—objects that intrinsically correspond to one word or thought-sign rather than another) that the Brain in a Vat case is a puzzler for him.¹⁵

But why should we agree with Putnam that this view is untenable? The mind has many remarkable powers, the ability to experience various sensations, to remember the past, to make reasoned guesses about what the future will hold. Why should we find its intrinsic intentionality so bizarre? Why should we find “noetic rays” any stranger than the light waves, electro-chemical impulses, etc., which (if our theories are largely correct) connect certain mental representations with physical objects? Putnam doesn’t really say.

However, I believe that his reference to brains in a vat in the passage above gives us insight into why Putnam finds such theories “magical,” and perhaps why we should as well.

¹⁵ *Reason, Truth, and History*, p. 51.

Putnam uses the story of the brains in a vat for many purposes in *Reason, Truth, and History*. In perhaps its main role, the story functions in an argument to the conclusion that the proposition that *we* are all brains in a vat is either meaningless or false. However, it seems to me that Putnam also uses the story independently of that argument to illuminate the shortcomings of various semantic realist theories of reference.¹⁶ The story is as follows: Consider a metaphysically (although perhaps not physically) possible world which contains only a super-computer and a very large vat filled with nutrients and containing many human-like brains. The brains are alive and their nerve endings are connected to the super-computer which causes them to have the collective experience that they live in a world very much like the actual world.¹⁷ The question then is, could the brains in the vat think or say truthfully that they were brains in vats? The proponent of an intrinsic intentionality theory of reference will say 'yes, they could'. Noetic rays, or whatever, connect their mental representations to mind-independent objects and properties in such a way that the mental analogue of the sentence 'we are all brains in a vat' refers to a mind-independent state of affairs which actually obtains. So far so good for the proponent of a magical theory.

However, one begins to see a problem for an intrinsic intentionality theory of reference when one asks a slightly different question concerning the brains in a vat: Can a brain in a vat think or say truthfully that there really are no trees? Imagine that, as things

¹⁶ It plays a similar role in "Realism and Reason," p. 127.

¹⁷ This is essentially a materialistic version of the deluded Cartesian soul story discussed in Chapter Two.

seem to him, he stands in front of a great redwood tree. For some reason, he briefly entertains the proposition that he is and always has been, a brain in a vat. He begins to speculate that, if he is a brain in a vat, then it is likely that there really are no trees. He begins to entertain the proposition that there really are no trees. He walks up and touches the rough bark of the redwood tree and, looking up the length of its trunk, dismisses for good the idea that he is a brain in a vat. Now, if the noetic rays have been doing their job through this whole episode, they should hook his mental representations up with sets or ordered n-tuples of mind-independent objects. His mental analogue of the term 'vat' gets hooked up with vats, his mental analogue of the term 'brain' gets hooked up with brains, and his mental analogue of the term 'trees' gets hooked up with...what? By hypothesis, there are no mind-independent trees for his term to hook up with. All there are are the images of trees produced by the super-computer. Perhaps he can't really refer to trees or think that there are no trees. However, it would certainly seem that, if a brain in a vat can truthfully think that he is a brain in a vat, he should be able to truthfully think that (perhaps as a consequence of this) there really are no trees. It is hard to see why he should be able to do the one but not the other. But the problem is that there are no trees for the noetic rays to connect his mental analogue of the term 'tree' up with.

One might reply by saying that, while there are no actual trees in the world in question, there are possible trees for a brain in a vat's noetic rays to hook onto. But for such a reply to work, noetic rays must somehow be able to hook onto mind-independent trees which exist in other possible worlds. If this doesn't violate outright the individuation

principles for possible worlds, it at least makes it clear how truly bizarre noetic rays would have to be to accomplish this feat. In the absence of much more information about how these noetic rays could possibly work, such a theory of reference would be magical.

But, perhaps, Putnam makes too much of the metaphor of “noetic rays.” This is, after all only a picture of intrinsic intentionality, and pictures can sometimes be misleading. The proponent of the intrinsic intentionality theory may simply want to say that ‘tree’ in the brain in a vat’s language of thought expresses a *concept*, one that would apply to mind-independent trees if there were any. There is no need for such admittedly bizarre things as noetic rays, and so, no problem if, in a particular case, there is nothing for the rays to hook up to. After all, we are familiar with the idea that, e.g. our concept *intelligent, extraterrestrial being* is one that would apply to beings from other planets with whom we could learn to communicate, if there are any such beings. We have many concepts which may, as a contingent matter of fact, fail to refer to anything in the world. Why should we find it strange that brains in a vat would too?¹⁸

This commonsense reply to the worry raised above is initially quite tempting. However, it seems to me that, when we look at the cases involved more closely, it becomes apparent that there is a substantial disanalogy between the case in which the brain in a vat’s concept *tree* fails to apply to actual mind-independent trees (because there are none) and the more common case in which one of our concepts contingently fails to apply

¹⁸ This response on behalf of the proponent of the intrinsic intentionality theory was pointed out to me by Robert Coburn.

to anything in the world. When we imagine that, e.g., our concept *intelligent extraterrestrial* fails to apply to anything in the world, we are generally imagining a case in which most of our other concepts which we think apply to the world actually do apply and in which most of our commonsense and scientific pictures of the world are (at least approximately) correct. We imagine ourselves and our relation to the universe as described by science; we imagine the vast reaches of empty space and the blazing intensity of distant stars. We think of the delicate balance of cosmic conditions required to support life at all, let alone intelligent life. What makes the supposition that our concept *intelligent extraterrestrial* does not apply to anything in the universe seemingly unproblematic is, in large measure, that it is made against the background supposition that we do not have any completely compelling reasons to think that there are any intelligent extraterrestrials. This seems to be a concept which we invented when it became plausible to imagine that it might have application to things in the world, but which (at least so far) we have never found reason to actually apply to anything in the world.

It is different, however, when one imagines how truly strange it would be to entertain the supposition that our concept *tree* is a concept that fails to apply to the world.¹⁹ We would have to imagine that we are all suffering from some collective hallucination, or that *we* are all brains in a vat, or that some other skeptical hypothesis is

¹⁹ Here, I don't mean to suggest that it is difficult to imagine that our concept does not apply because some terrible disease has killed all of the trees or because trees never evolved on Earth. Rather, I mean to suggest that it is difficult *as things stand now* to imagine that our concept *tree* has no application.

true. And, once we imagine that, it begins to seem much more problematic that our concept *tree* is one that would apply to mind-independent trees, if there were any, rather than to the tree-like images with which we are familiar. In virtue of what does our concept *tree* apply to something in a possible world only if it is mind-independent rather than to the more familiar kind of thing, an instance of which I seem to be seeing out my window even now?²⁰ It is not in virtue of the holding of some relation of similarity (there is, on this supposition, no mind-independent object for my concept *tree* to be similar to). It is not in virtue of the way we use the term ‘tree’ in our natural language (this we use to refer to things of the tall, leafy sort I see out my window). This does begin to seem a bit mysterious.

The proponent of the intrinsic intentionality theory of reference who takes the “commonsense” line sketched above will likely remind us at this point that the concept *tree*, on his view, just *intrinsically* applies to something in a possible world only if it is a mind-independent object. The worry, however, is that we no longer have any clear idea what this means or how it could possibly be true. The metaphor of “noetic rays” may be a misleading picture, but at least it has the virtue of being a fairly coherent picture. It is the noetic rays which form the link between our concepts and the mind-independent things to which they apply. Perhaps the proponent of an intrinsic intentionality theory of reference

²⁰ Of course, that tall leafy thing out my window may very well be a mind-independent object. However, we are here considering that *my* concept *tree* fails to actually apply to anything (because there are no trees in the mind-independent world), in which case, whatever that is out my window, it isn’t a mind-independent object.

would do well after all to develop the noetic ray model of intrinsic intentionality and find a suitable solution to the problem that arises when, e.g., there are no trees for a brain in a vat's noetic rays to hook onto.

A more traditional response to this worry would be to postulate the existence of the mind-independent but uninstantiated property, *treehood*.²¹ Such a property is an abstract, necessarily existent object, very much like a Platonic form. Properties (and relations) of this sort combine to form abstract states of affairs such as a cat's being on a mat or there being a tree. Every state of affairs that could possibly obtain exists as an abstract object in every possible world.²² Worlds vary in that different states of affairs may obtain in different worlds. Once he can make use of such a lush ontology, the semantic realist who adopts an intrinsic intentionality theory of reference can explain how it is that a brain in a vat can truthfully think or say that there are no trees. The problem had been that there were no trees for his noetic rays to hook onto. On this view, his noetic rays needn't somehow hook his term 'trees' onto non-existent (or merely possible) trees. His noetic rays hook his term 'trees' onto the abstract property *treehood*, which

²¹ Two contemporary philosophers who develop this kind of view are Alvin Plantinga in "Actualism and Possible Worlds," *Theoria* 42 (1976), pp. 139-160 and Peter van Inwagen in "On Always Being Wrong," *Midwest Studies in Philosophy*, vol. 12, *Realism and Antirealism*, Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, eds. (Minneapolis: University of Minnesota Press, 1988), pp. 95-111. Laurence Bonjour has expressed the belief that the kind of view favored by these philosophers is not ultimately distinct from the concept view just discussed.

²² It is not necessary for the semantic realist's purposes that every possible property and every possible state of affairs exist necessarily. What is necessary is that every property and state of affairs which is a possible referent of the terms and sentences of a mental language in a given world exist in that world.

does exist.²³ When he wonders whether or not there really are trees, he is wondering whether or not this property is instantiated as part of an obtaining state of affairs. His sentence, ‘there are no trees,’ will be true if (as is the case) there are no obtaining states of affairs which involve the instantiation of *treehood*, otherwise it will be false. Problem solved! We have mental representations that intrinsically refer to Platonic properties, relations, and states of affairs, and these abstract objects exist and have the features that they do mind-independently. Determinate reference to mind-independent objects and properties is achieved.

To the extent that this story makes sense and is plausible, semantic realism is relatively unproblematic. Unfortunately, it is far from clear that such a story makes sense and, in addition, is plausible. A theory of reference of this sort will inherit all of the traditional problems associated with transcendental Platonic forms.²⁴ Even if these problems can be resolved, the view’s ontological commitments may be a bit too rich for some semantic realists’ tastes. There is also the further worry, expressed by Putnam,²⁵ that any theory of reference which trades on the idea that mental representations, unlike

²³ Notice that the noetic rays are playing a role analogous to the traditional role of Plato’s relation of participation.

²⁴ The problems associated with transcendent Platonic forms can be divided into roughly two categories: (i) problems concerning the obscure relation between forms and the particulars which “participate in” them, and (ii) problems concerning the ontological status and nature of non-particular entities. Plato himself raised many of these worries in his dialogue, *Parmenides*. For a more recent discussion of these problems and suggestions for further reading, see *The Encyclopedia of Philosophy*, edited by Paul Edwards, (New York: Macmillan and the Free Press, 1967), pp. 195-197, 206.

²⁵ *Reason, Truth, and History*, p. 2.

physical signs, are intrinsically intentional is incompatible with materialism about the mind. If no physical object can intrinsically refer to one thing rather than to another, but mental representations are intrinsically intentional,²⁶ then it would seem to be impossible for mental states to be identified with any purely physical states.

Many semantic realists will share Putnam's intuition that this is too high a price to pay for an easy reply to the argument against semantic realism. Some, however, will find nothing problematic in the notion that the mind is immaterial, that Platonic forms (or entities very much like them) exist and exist necessarily, and that mental states intrinsically refer to such Platonic forms, or, alternatively, that some brain states possess the required intrinsic intentionality. Those who find nothing problematic with such a theory of reference will be able to deny premise 3 of the argument against semantic realism. Putnam has no response to such a move other than to reiterate its apparent problems. For this reason, we should perhaps regard the conclusion of the argument against semantic realism as conditional: *if* an intrinsic intentionality theory of reference is to be rejected, *then* one has reason to believe that semantic realism is false. In the sections that follow, I will discuss several ways in which a semantic realist may attempt to fix a determinate relation

²⁶ Some, perhaps most notably John Searle (see "Minds, Brains, and Programs," *The Behavioral and Brain Sciences* 3 (1980), pp. 417-424), will find it unproblematic to suppose that a brain state could possess intrinsic intentionality. They will not share this last worry. However, there will be many others who will want to hold that, whatever may be the case with mental states, brain states are not the kinds of things which could possess intrinsic intentionality. This seems especially true when we are reminded that the brain state must somehow be about, say, mind-independent trees, even when the brain (as in the case of the brains in a vat) has never stood in any causal relation whatever to actual mind-independent trees.

of reference between our terms and mind-independent objects and properties (and thus deny premise 3 of the argument against his view) that do not involve the adoption of an intrinsic intentionality theory of reference.

3. If a semantic realist does not want to adopt an intrinsic intentionality theory of reference, what are his other options? According to Putnam,²⁷ the “received view” of how reference is fixed by us is associated with the notion of operational and theoretical constraints. Operational and theoretical constraints serve to restrict admissible interpretations of a theory by stipulating the truth conditions of the sentences of the language of the theory. Operational constraints on a theory will be sentences of the form: ‘an admissible interpretation of theory T is such that the sentence S is (probably) true when (most of) the set C of conditions is fulfilled’. Putnam’s example of such a constraint is that an admissible interpretation of the sentence ‘electricity is flowing through this wire’ will be such that most of the time when ‘electricity is flowing through the wire’ is true, (among other things) the needle of a properly functioning voltmeter connected to the wire will be deflected.²⁸

David Lewis calls the view that operational constraints alone can be used to fix the reference of the entire vocabulary of a language (as opposed to merely fixing the reference of a term newly added to an existing language) “global descriptivism” and provides a fairly

²⁷ *Reason, Truth, and History*, p. 29.

²⁸ *Ibid.*, pp. 29-30.

clear picture of how such a theory would work. He begins by describing a more local descriptivist theory of reference:

[W]e may add new language to the old, a little at a time, by introducing undefined terms in our theorizing. Thereby we associate clusters of old-language descriptions with our new terms; and thereby, if the world cooperates, we bestow reference on the new terms. 'Jack the Ripper did this, that, and the other' says the detective; his point is in part to hypothesise that there is someone who did this, that, and the other, and in part to stipulate that the one who did, if such there be, is to become the referent of 'Jack the Ripper.' The new term 'Jack the Ripper' is to acquire the referent, if any, of the old-language description 'the one that did this, that, and the other'. The intended interpretation of the augmented language is to be an extension of the old interpretation of the old language.²⁹

He notes that this local descriptivism only works if there already was a language with determinate reference and says we may ask how this previous language got its reference.

It is therefore tempting to try the same method on a grander scale. We can introduce terms in little families. How about bigger families? How about the biggest family of all—the entire vocabulary of a language. Then we needn't worry how the old vocabulary got its reference....We just go on as before. The intended interpretation will be the one, if such there be, that makes the term-introducing theory come true. (Or:...come near enough to true. Or: the intended interpretations will be the ones, if such there be...with indeterminacy if there are more than one.)³⁰

So, global descriptivism is the view that operational constraints alone can fix the set of acceptable interpretations for an entire language. But, operational constraints needn't be made to perform the job alone.

²⁹ "Putnam's Paradox," *Australasian Journal of Philosophy* 62 (September, 1984), p. 222.

³⁰ *Ibid.*, p. 224.

According to Putnam, theoretical constraints, which are usually stated as constraints on the acceptability of a theory, may be used in addition to operational constraints to further restrict the class of admissible interpretations of a theory or language.³¹ So, to borrow Putnam's example, the constraint that a theory be conservative can be reformulated as the constraint on interpretation that "an admissible interpretation is such that it renders true sentences which have been accepted for a long time, except where this would require undue complication in the theory consisting of the set of sentences true under the interpretation, or too great a revision in the operational constraints."³² Together, operational and theoretical constraints are supposed to fix the reference of a language from *inside*, as it were. The set of admissible interpretations for the language is to be completely specifiable from within the language. If such a theory can be made to work to fix reference between our terms and objects and properties in the mind-independent world, the semantic realist could deny premise 3 of his view without having to adopt an intrinsic intentionality theory of reference.

Unfortunately, Putnam has a fairly powerful objection to such a view. If it is to be possible for an epistemically ideal theory to be false, then there must be a determinate relation of reference that pairs terms of our language with sets and sequences of items in the mind-independent world and that is, in some sense, the correct one, with the result that the fact that a theory comes out true on some other pairing is irrelevant to the question of

³¹ *Reason, Truth, and History*, pp. 30-31.

³² *Ibid.*, p. 31.

whether or not the theory is true. The problem for this view is that a variant of Putnam's model-theoretic argument can be used to show that no view which only fixes the truth-values of whole sentences by imposing operational and theoretical constraints is sufficient to fix a determinate relation of reference between the terms of our language and parts of the mind-independent world.³³ It is a theorem of model theory that given any interpretation I of a language L, one can construct an interpretation J of L which is isomorphic to I. An interpretation I is isomorphic to an interpretation J if and only if (i) they are interpretations of the same language (thought of syntactically), (ii) they assign the same sentences the same truth values, and (iii) there is a one-to-one function H whose domain is the domain of I and whose range is the domain of J such that if I assigns a name the designation d, then J assigns it H(d) (and similarly for the function symbols and n-ary predicates of I and J).³⁴ Theoretical and operational constraints will serve to limit the range of admissible interpretations up to the point of isomorphism. However, as there will likely be an infinite number of isomorphic interpretations of a given theory, reference is still radically indeterminate. Further, the indeterminacy is by no means of the relatively harmless Quinean variety.³⁵ "The argument...shows that the truth conditions for 'A cat is on a mat' don't even exclude the possibility that 'cat' refers to *cherries*."³⁶ If we can fix the reference of the terms of a theory only by appealing to the truth values of its sentences,

³³ See especially *ibid.*, pp. 217-218.

³⁴ For more details, see George Boolos and Richard Jeffrey, *Computability and Logic* (Cambridge: Cambridge University Press, 1989), p. 191.

³⁵ See *Word and Object* (Cambridge: MIT Press, 1960), pp. 68-73.

³⁶ *Reason, Truth and History*, p. 35.

then we can't really fix the references of the terms of the theory at all. The semantic realist will need to try some other approach to specifying the intended interpretation of the theory.

4. One thing that a semantic realist might try to do at this time, if he wants to avoid both the referential indeterminacy of the view sketched above and the intrinsic intentionality theory of reference, is to claim that there must be some naturalistic constraint on the relation D of reference that singles out the intended interpretation. That is, what is needed is some naturalistic definition of reference such that:

x refers to y iff x and y stand in relation D

(where 'x' is a term of our language and 'y' is an object, property, or relation in the mind-independent world). Perhaps because of his role in developing causal theories of reference, Putnam seems to think that the most obvious candidate to play the role of reference fixing relation D is a causal or more complex functionalistic relation.³⁷ But there are other candidates. G. H. Merrill³⁸ and David Lewis³⁹ have suggested that the naturalistic constraint on reference is a constraint on the *kind* of objects, properties and relations to which relation D may connect our terms.

Among all the countless things and classes that there are, most are miscellaneous, gerrymandered, ill-demarcated. Only an elite minority are carved at the joints, so that their boundaries are established by objective

³⁷ "Realism and Reason," pp. 126-7.

³⁸ See "The Model-Theoretic Argument Against Realism," *Philosophy of Science* 47 (1980), pp. 69-81.

³⁹ See "Putnam's Paradox," esp. pp. 227-229.

sameness and difference in nature. Only these elite things and classes are eligible to serve as referents....When we limit ourselves to the eligible interpretations, the ones that respect the objective joints in nature, there is no longer any guarantee that (almost) any world can satisfy (almost) any theory.⁴⁰

If I am looking in the right place for a saving constraint, then realism needs realism. That is: the realism that recognizes a nontrivial enterprise of discovering the truth about the world needs the traditional realism that recognizes objective sameness and difference, joints in the world, discriminatory classifications not of our own making.⁴¹

On the Merrill/Lewis view, any candidate for relation D must pair our terms up only with objects, properties, and relations that are eligible, ones that carve nature at the joints.⁴²

But, perhaps the semantic realist need not settle on a single account of the relation D which holds between terms and objects and classes in the world. Perhaps all that is required is that there be some naturalistic relation D such that D is the relation of reference. If such a relation obtains between our terms and objects, properties, and relations in the mind-independent world, then a semantic realist would seem to be able to deny premise 3 of the argument against his view whether or not he is in a position to give a detailed analysis of relation D.

Putnam's reply to any semantic realist who would deny premise 3 by appealing to naturalistic constraints on reference is known in the literature as the "just more theory"

⁴⁰ Ibid., p. 227.

⁴¹ Ibid., p. 228.

⁴² Notice that the Merrill/Lewis constraint on reference works only if there are elite classes of objects, only if the extreme permissive view (see above, pp. 72-74) is false. So, this line of response to Putnam's argument against semantic realism will be able to get off the ground only if the metaphysical realist is able to solve the problem of a crowded ontology discussed in Chapter Three above.

reply. This reply can be (and has been) interpreted in such a way that it has little or no force. Although I believe this interpretation of the just more theory reply is nothing more than a strawman, it is so prevalent in the literature that I think it warrants discussion. In the remainder of this section, I will present this weak version of the just more theory reply and briefly discuss its weakness. In the next section, I will present another interpretation of Putnam's just more theory reply, due largely to David Anderson, that has much more force. It is this stronger reply that constitutes a genuine difficulty for the semantic realist.

On a common reading of the just more theory reply,⁴³ Putnam's response to the positing of a naturalistic constraint on the reference of an epistemically ideal theory is to say, "Fine. We'll just add this condition into our total theory and see what happens." The claim is now that the same model-theoretic moves can be made with respect to the theory enriched with the naturalistic condition for stipulating the correct interpretation. To relate this back to our simple example above,⁴⁴ we can enrich our theory that all ravens are white with the following two conditions: (a) 'ravens' refers to *i9* and (b) 'white' refers to *i1*. We will now add the individual constants 'r' and 'w' to stand for "the term 'ravens'" and "the term 'white'" respectively, the individual constants '*i9*' and '*i1*' to stand for

⁴³ See among others, Lewis, "Putnam's Paradox," pp. 225; Mark Heller, "Putnam, Reference and Realism," *Midwest Studies in Philosophy*, vol. 12, *Realism and Antirealism*, Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, eds. (Minneapolis: University of Minnesota Press, 1988), pp. 123-4; Alan Musgrave, "What's Wrong With the Model-Theoretic Argument Against Realism?," pp. 6-7; Michael Resnik, "You Can't Trust an Ideal Theory to Tell the Truth," *Philosophical Studies* 52 (1987), esp. p. 152; Michael Devitt, "Review of Putnam's *Reason, Truth, and History*," *Philosophical Review* 92 (1984), pp. 274-7.

⁴⁴ See above, pp. 102ff.

individuals i_9 and i_1 respectively, and the two place predicate D to stand for the relation of referring. Our enriched theory can now be formalized:

$$\begin{aligned} & (x)(Rx \rightarrow Wx) \\ & D(r, i_9) \\ & D(w, i_1) \end{aligned}$$

Unfortunately, we can again give an interpretation A on the first ten natural numbers in which this theory comes out true:

Domain:	{1,2,3,4,5,6,7,8,9,10}	
$I(R)$:	{3,5,7}	$I(W)$: {1,3,5,7,9}
$I(r)$:	2	$I(w)$: 6
$I(i_1)$:	8	$I(i_9)$: 4
$I(D)$:	{<2,4>, <6,8>}	

And again, this interpretation can be mapped onto an isomorphic interpretation M of the objects of the world. Again, the theory will be true by correspondence with the world despite the fact that there is a black raven. In formulating the naturalistic constraint on reference, the semantic realist has just added more theory to T and has in no way alleviated the referential indeterminacy that threatened the original theory—the same game can simply be played again. Apparently, attempting to fix reference by positing naturalistic constraints will not work.

When Putnam's "just more theory" reply is interpreted in this way, the obvious response is for the semantic realist to say that to propose a causal or other naturalistic constraint on reference is to posit a constraint that is external to the theory being interpreted. The constraint is supposed to work from outside the theory and need not be expressible in the theory. It is not necessary on such a view (which might be called

“semantic externalism”) that we be able to determinately state the appropriate relation D. It is sufficient for reference that there be a determinate and appropriate causal (or other naturalistic) relation. So, the addition of more theory is not to be identified with the holding of the naturalistic constraint. Michael Resnik⁴⁵ takes a slightly different line and, with an appeal to Skolem’s Paradox,⁴⁶ points out that the mere fact that a theory “asserts” that a condition holds does not entail that the condition in fact holds. Again, this gives us a reason to think that the addition of a sentence that asserts that the naturalistic constraint on reference holds is in no way the same as the actual holding of the naturalistic constraint. With respect to our simple example, even though it seems to “assert” that ‘ravens’ refers to ravens—the formula ‘D(r, i9)’ comes out true in the arithmetic interpretation—the actual constraint that ‘ravens’ refer to ravens is not met. So, it does look as if the realist might be able to avoid the “just more theory” reply by saying that the interpretation M of the ‘all ravens are white’ theory does not meet the two conditions added to pick out the intended interpretation even though it “asserts” that it does.

Since only interpretations which *actually* meet the naturalistic constraints on reference (whether or not they assert that they do) will count as intended interpretations on this reply, the intended interpretation I of our simple theory does not meet the conditions on reference simply because it “asserts” that it does. Putnam’s just more

⁴⁵ In “You Can’t Trust an Ideal Theory to Tell the Truth”.

⁴⁶ A theory may “assert” that it has only nondenumerable models. But the Skolem Löwenheim theorem ensures that the theory will have denumerable models in which it is satisfied even though it seems to assert that this is not the case.

theory reply, as interpreted above, will not work. If this were the only way to interpret the just more theory reply, the semantic realist would have a response to Putnam's argument against his view. However, there is another interpretation of the just more theory reply which makes things much more difficult for the semantic realist.

In "What is the Model-Theoretic Argument?," David Anderson suggests that it is better to read the just more theory response as showing that it is far from clear that the semantic realist actually has a naturalistic theory of reference that is substantial and plausible.⁴⁷ Merely producing the schema 'x refers to y if and only if x and y stand in relation D' will not be satisfactory. One problem is that the schema is consistent with D's being a magical relation rather than a naturalistic relation. If we are to be certain that D is a naturalistic relation, we must have a more substantial account of the nature of D on the table. The just more theory response, according to Anderson, should then be construed as offering a trilemma for a semantic realist who would try to provide a naturalistic theory of reference that is both long on substance and fairly plausible. Although Anderson, following Putnam, focuses on causal constraints, similar points apply to the other constraints discussed above.

[T]he causal realist offers a name for the reference fixer, 'causality'. But, a name may or may not be a substantive advance over a placeholder for "something we know not what." ...The first lemma is to assume that 'causality' is a mere placeholder....Since the causal realist purports to give a substantive account of the nature of the reference fixer, this alternative is unacceptable....The second lemma...is to assume that causality is of the empirical kind. While empirical causality is a plausible thing for naturalists to believe in, it seems to be incapable of fixing reference to mind-

⁴⁷ "What is the Model-Theoretic Argument?," p. 316.

independent objects....To assume that causality is a genuinely metaphysical notion constitutes the third lemma....The problem of course, is to make sense of such a notion.⁴⁸

Anderson's take on the reply is suggestive, but it is just a sketch. In the section that follows, I will develop this trilemma for the semantic realist in more detail. Although my development of the trilemma will have much in common with Anderson's brief sketch, there will be some points of departure. Whether or not the reply, as I develop it, can be correctly attributed to Putnam, it is, I think, strong enough to show that the semantic realist (who does not want to adopt an intrinsic intentionality theory of reference) will have considerable difficulty responding to Putnam's argument against his view.

5. Suppose that the semantic realist has given us a name for (or perhaps a description of) the naturalistically referential relation D. Perhaps the name he has given it is 'causality'. Perhaps he has described it (à la Merrill/Lewis) by using the term 'the relation which holds between our terms and certain elite objects or classes'. In any case, it would seem that at least one of the following must be true: (i) the name or description provided is a mere placeholder for a more substantial theory of the nature of D, (ii) the named or described relation D is an empirically discoverable relation, or (iii) the named or described relation D is a genuinely metaphysical relation, one that need not be empirically discoverable. The thrust of the "just more theory reply" is that none of these is an entirely acceptable option for the semantic realist. To see whether or not this is so, we need to

⁴⁸ Ibid., pp. 316-317.

examine the options one by one. The first case is easily dealt with. If the provided name or description is a mere placeholder for a more substantial theory of the nature of D, then we do not yet have a substantial and plausible naturalistic theory of reference to mind-independent objects.

Let us now examine the second option. In discussing this case, for the sake of clarity of exposition, I will follow Putnam and Anderson in supposing that the theory on the table is a causal theory of reference.⁴⁹ So, let us suppose that we do have a substantial causal theory of reference on the table. Suppose further that the causal relations which connect our terms to their referents are of the empirically discoverable kind. The problem with this option is not that the resulting theory would fail to be substantial or plausible. Rather, the problem is that the truth of such a theory does not guarantee that our terms ever succeed in referring to mind-independent objects and properties. While such a theory is consistent with semantic realism, it is not in and of itself a semantic realist theory; it does not guarantee reference to the mind-independent world.

To see why this is the case, we can again appeal to the story of the brains in a vat hooked up to the super-computer. This time, let us imagine that the brains in the vat and the super-computer to which they are connected exist in a world which is very much like our own except the only people who exist in this world are the brains in the vat themselves. Although the world that the brains believe themselves to inhabit in many ways

⁴⁹ I believe that analogous results will follow *mutatis mutandis* if we take this to be a theory restricting reference to “elite” objects and classes whose boundaries are established by empirically observable sameness and difference.

resembles the world which they actually inhabit, we may suppose that this is a cosmic coincidence—the super-computer and its “program” have been produced at random. Now suppose that one of these brains utters a sentence of the type ‘There is a tree in front of me.’ To what does his word ‘tree’ refer? In *Reason, Truth, and History*, Putnam asks a similar question:

[D]o their verbalizations containing, say, the word ‘tree’ actually refer to *trees*? More generally: can they [the brains in the vat] refer to *external* objects at all? (As opposed to, for example, objects in the image produced by the automatic machinery.)⁵⁰

One of the results of the arguments contained in the first chapter of *Reason, Truth, and History*,⁵¹ is that the best that a brain in a vat could hope for, given a theory which makes the relation of reference a causal relation of the empirical variety, is reference to the tree-like images (call these “trees-in-the-image”) or perhaps the parts of the computer program which produce such images in him.⁵² None of the relevant type of causal connections hold between the brain’s use of ‘tree’ and actual, mind-independent trees (even though there are such things in this world).

Perhaps this point will be easier to see if we start by imagining that it is we who are the brains in the vat. (Although, as we will see below, Putnam argues that it would

⁵⁰ *Reason, Truth, and History*, p. 12.

⁵¹ See *ibid.*, esp. pp. 5-14.

⁵² Putnam spends much of his time in the first chapter of *Reason, Truth, and History* arguing that something very like the causal theory must be (at least a part of) the correct theory of reference. For our purposes, these arguments are not necessary since we are assuming at this point that a causal theory of reference is true and that it is empirical causality that is at issue.

not really be possible to succeed in thinking this if it were true.) I look outside my window and I see what appears to be a stand of alderwood trees. Since I am imagined to be a brain in a vat, what I actually see (if, indeed, I can be said to *see* anything)⁵³ is a stand of alderwood-trees-in-the-image. I think to myself *there is a tree in front of me*. The referent of my term 'tree' on the view we are considering is the set of objects connected to my use of the term by the relevant causal relations, ones that could be discovered by empirical means. Perhaps even now our best scientists (aided, of course, by our best philosophers) are finishing up a theory which explains to the last detail how it is that a tree is, among other things, able to reflect light in such a way as to affect my nervous system and allow me to perceive it and give it a name. Perhaps these scientists and philosophers are even now working on (and very close to) a theory which would explain to the last detail my ability to recognize the properties that make an object fall into a given kind and give the kind a name. Unfortunately, in this case, our best scientists and philosophers too are brains in a vat. The "trees" which are able to affect my nervous system are really trees-in-the-image or perhaps the parts of the computer program which produce such images; the "lightwaves" are merely lightwaves-in-the-image, etc. No causal regularities

⁵³ I will largely ignore this complication. It may well be that the verb 'to see' is used most properly as a success verb in such a way that, in a case like this, merely "seeing" a tree-in-the-image does not count as a genuine case of seeing. However, it seems to me that even brains in a vat would have reason to make a theoretical distinction between "appearance" and "reality" (although this would really be a distinction between appearance-in-the-image and reality-in-the-image, both of which are mere appearances) in order to explain what is going on in cases in which one of the brains "sees" a tree-in-the-image which none of the other brains who are present-in-the-image are able to "see." So, brains in a vat too can distinguish between cases of successful "seeing" and cases of merely seeming to "see."

exist which connect my uses of the term 'tree' to mind-independent trees. The only causal regularities which obtain between my use of the term 'tree' and any part of the world are ones which hold between my term and trees-in-the-image (or the part of the computer program which produces them). If these relations are to be the relations that fix reference, brains in a vat cannot refer to mind-independent trees. At best, the causal theory of reference developed by the envatted scientists and philosophers explains the nature of a relation that holds between my term 'tree' and trees-in-the-image (or between my term and the parts of the computer program that produce trees-in-the-image); it does not explain how my term 'tree' hooks up (if indeed it does) with mind-independent trees. If we are the brains in a vat, a completed causal theory of reference developed by our best scientists and guided by our best philosophers will either be false or will not be a version of semantic realism.⁵⁴

It is now easy to see why Putnam would be inclined to say that brains in a vat couldn't truly think or say that they were brains in a vat, if the causal theory of reference were correct. Their word 'brain' would refer to brains-in-the-image, their word 'vat' to vats-in-the-image. So, the brain's utterances of 'we are all brains in a vat' would mean *we are all brains-in-vats-in-the-image*. But, it is false that they are brains-in-vats-in-the-image. By hypothesis, they are regularly-embodied people-in-the-image. If one is a brain in a vat, one lacks the appropriate causal connections to real brains and real vats necessary

⁵⁴ Or at least not a semantic realism that allows our terms to refer to trees or any other mind-independent objects and structures other than (parts of) the computer's program.

for believing that one really is a brain in a vat. So, on such a theory, a brain in a vat cannot really entertain the proposition that he is a brain in a vat. At best, he can entertain the (false) proposition that he is a brain-in-a-vat-in-the-image. The proposition that everyone is a brain in a vat can only be entertained if it is false.

A semantic realist, no doubt, will remind us at this point that all that the preceding discussion shows is that causal connections of the relevant sort won't enable brains in a vat to refer to the mind-independent world. None of it shows that causal connections of the relevant sort don't enable *us* to refer to the mind-independent world, if we aren't brains in vats. If we are regularly embodied people, then the complete causal theory of reference developed by our scientists and philosophers *will* explain how it is that our terms hook on to pieces of the mind-independent world. The causal theory of reference is perfectly consistent with semantic realism.

The problem for a semantic realist who would attempt to deny premise 3 of Putnam's argument against his view by adopting this kind of causal theory of reference is not that the theory is inconsistent with his view. Rather, Putnam's claim (or at least my claim) is that a causal theory of reference of this sort will be a substantial and plausible explanation of how it is that *we* are able to refer to mind-independent objects and classes *only if we can be fairly certain that we are not all brains in a vat*. That is, if we are to have reason to believe that a causal theory of this type *is* a version of semantic realism, we

must have antecedent reason to believe that we really are not brains in a vat.⁵⁵ And, unfortunately for the semantic realist, *he* cannot appeal to Putnam's argument that the supposition that we are all brains in vats will be false on a causal theory of reference. If metaphysical realism is true, it is metaphysically possible that we are all brains in a vat.⁵⁶ Further, if metaphysical realism is true, the mere fact that we couldn't say or think truly that we are all brains in a vat does not entail that we really aren't. All it entails is that, in such a case, we would not be able to think or say what we really are.⁵⁷ The semantic realist must find his own argument for the claim that we are not brains in a vat.

But things are even worse than they might appear. In order to show that an empirical version of a causal theory of reference will serve his purposes, the semantic realist needs to be able to show that we are not brains in a vat; that is, he needs to be able to show that the proposition that we are all brains in a vat is false. But, since it is metaphysically possible that we are brains in a vat (whether or not we can coherently say or think that we are), he needs to be able to understand this proposition in such a way that it just might be true. However, as Putnam's argument explicated above seems to show, if a causal theory of reference is correct, it is not possible to understand the proposition that we are all brains in a vat in such a way that it might be true. If we were brains in a vat, we

⁵⁵ Or, alternatively, that "our world" is not merely a world of misleading appearances. Notice that this would essentially amount to a response to the skeptical arguments discussed in Chapter Two.

⁵⁶ Or at least it's metaphysically possible that the common sense and scientific world is a world of misleading appearances.

⁵⁷ A similar point is made by Peter van Inwagen in "On Always Being Wrong."

would not even be able to say that we were; and, necessarily, if we can say that we are, it is false. So, there is a very real sense in which such a causal theory does not sit well with semantic realism. We will have reason to believe that the causal theory counts as a version of semantic realism only if we have antecedent reason to believe that the proposition that we are all brains in a vat is false (where this proposition is understood in such a way that it could have been true). But this entails that there is at least one proposition that makes reference to the mind-independent world independently of any causal connections of the empirical variety between its terms and the world. In other words, it would seem that in order to know that the causal theory allows us to refer to the mind-independent world, we would have to know that some other theory of reference to mind-independent objects is true. And, since we cannot simply take it for granted that the causal theory gets us reference to the mind-independent world (after all, the causal theory does not entail semantic realism), semantic realism and the causal theory of reference would seem to form an unhappy union.⁵⁸

⁵⁸ In this discussion, I have deviated fairly substantially from Anderson's version of this horn of the causality trilemma. Anderson writes, "Since it functions entirely at the empirical level, empirical causality must ultimately be characterizable in terms of theoretical and operational constraints...While empirical causality is a plausible thing for naturalists to believe in, it seems to be incapable of fixing reference to mind-independent objects. If causality is to reach beyond the empirical all the way to things in themselves, then causality must be something more than a creature of theoretical and operational constraints, it must be something metaphysical" ("What is the Model-Theoretic Argument?," p. 317). In my development of this horn, I have focused on Putnam's use of the brains in a vat example in chapter 1 of *Reason, Truth, and History* rather than his discussion of the role of theoretical and operational constraints in chapter 2. But, perhaps the two are related.

The adoption of an empirical version of a causal theory of reference creates problems for the semantic realist since it does not guarantee that our terms refer to mind-independent objects and properties, and further, if such a theory is true, it seems almost impossible to show that we do succeed in referring to mind-independent objects and properties. Let us now examine the third possibility. Perhaps the semantic realist's reference relation *D* should be taken to be a genuinely metaphysical relation. Here, '*D*' is not to be taken merely as a placeholder. Rather, it refers (due to the holding of *D*) to a specific metaphysical relation of such a kind that necessarily, if it holds at all, it holds between our terms and mind-independent objects. More specifically, if it holds at all, it relates 'cats' to mind-independent cats, 'trees' to mind-independent trees, etc. According to Putnam, there are at least two major difficulties for a semantic realist who offers such a theory of reference:

[B]elieving that some correspondence intrinsically just *is* reference (not as a result of operational and theoretical constraints, or our intentions, but as an *ultimate* metaphysical fact) amounts to a magical theory of reference. Reference becomes what Locke called a 'substantial form' (an entity which *intrinsically* belongs with a certain name) on such a view. Even if one is willing to contemplate such unexplainable metaphysical facts, the epistemological problems that accompany such a metaphysical view seem insuperable.⁵⁹

In the remainder of this section, I will spell out and discuss these difficulties in more detail.

The first difficulty for this view is that it is far from clear that it is not a version of the intrinsic intentionality theory. Consider again the case of brains in a vat. If this is the

⁵⁹ *Reason, Truth, and History*, p. 47.

right way to think about reference, then it would appear that if D holds and a brain in a vat's words refer at all, his use of the term 'brain' refers to mind-independent brains and not to brains-in-the-image, his use of 'vats' refers to mind-independent vats and not to vats-in-the-image. So, if D holds, it is possible for a brain in a vat to entertain the true proposition that he is a brain in a vat (although he will likely have no reason to believe that this proposition is true). If the brains are able to refer, the reference relation D connects their terms up with parts of the mind-independent world completely independently of the way they use their terms. After all, they usually intend to use the term 'vat' to refer to vats-in-the-image, as when a doctor-in-the-image utters the sentence 'Put that tissue into the vat on the table there so we can send it off to the laboratory' while conducting surgery-in-the-image.⁶⁰ However, if D holds, 'vat' refers to mind-independent vats (or so one would suppose) even in these cases. The way the brains use the term 'vat' in most cases has nothing to do with what the term refers to. It would seem that D holds between their use of 'vats' and mind-independent vats in virtue of some intrinsic property of the term 'vat' in the brains' language. And, it is hard to see how this is much different than saying, as the magical theorist would, that their terms intrinsically refer. The semantic realist needs to say much more about this relation D before he has provided a substantial and plausible *non-magical* theory of reference.

⁶⁰ Can a brain in a vat even *intend* to use 'vat' to refer to vats-in-the-image if D holds and is the relation of reference? Not if intending presupposes reference. In many cases, the brains are simply mistaken about what they intend to be referring to.

The other difficulty for a view such as this is that, even once the semantic realist has said enough to convince us that he is not merely offering an intrinsic intentionality theory of reference, the view will face a number of epistemological worries. If a term refers to a mind-independent object or property, not in virtue of anything intrinsic to the term itself or in virtue of the way it is used, but in virtue of the holding of what Lewis calls an external relation⁶¹ D between a particular term, e.g., 'cats' and mind-independent cats, one might well worry how we could ever know that D held. Suppose that a proponent of this theory says that we can know that D holds because we can develop a theory according to which D holds and attempt to confirm it.⁶² The problem is that no matter how well confirmed the D-theory, I can always ask how I know that the term 'D' in the theory refers to the metaphysical relation D. The answer can only be, if D holds, 'D' refers to D. A proponent of such a theory of reference will never be able to get rid of this 'if'. He can only say, *if* D holds, we can know that it does, and hence, know that 'cats' refers to mind-independent cats. He has no choice but to adopt a version of epistemic reliabilism⁶³ at least with respect to our knowledge that our terms refer. While those who generally favor reliabilist theories of knowledge may not find this troubling,

⁶¹ See *The Plurality of Worlds*, p. 62. An external relation, according to Lewis is one, such as the relations of spatial or temporal distance, that does not supervene on the natures of the relata. Actually, D need not be a completely external relation since it may supervene in part on the natures of the referents of our terms as it does on the Merrill/Lewis view that restricts reference to elite classes and properties.

⁶² Notice that the brains can do this too.

⁶³ For more on epistemic reliabilism, see the Appendix.

those who do not may find this a fairly high price to pay for determinate reference to mind-independent objects and properties.

Another worry is that if D is a genuinely metaphysical relation, one that holds independently of the way we use our terms, then, even if we can be sure that it holds between our terms and mind-independent objects, we may well wonder which mind-independent objects or properties D hooks a given term up to. We began by saying that if D held, it related 'cat' to mind-independent cats, 'tree' to mind-independent trees, etc. So, it may at first appear as if any skeptical worries here are unwarranted. However, it is far from clear that this is the case. If D holds, sentences of the form "'x' refers to x" will be trivially true; they will convey no real information. So, if I am genuinely worried about what the term 'cat' refers to, telling me that 'cat' refers to cats will not help me in the slightest. After all, I can know that the sentence "'cats' refers to cats" is true provided that I know what 'refers to' means and understand that the term mentioned on the left is being used on the right, without knowing what a cat is.⁶⁴ So, if we know that D holds, we do know that, e.g., 'cats' refers to mind-independent cats. However, we are no closer to knowing what mind-independent cats *are like*.

Here it will again be helpful to think of the brains in a vat. Suppose that one of these brains utters the sentence 'a cat is on a mat.' His word 'cat' refers to a set of mind-independent objects in virtue of the holding of metaphysical relation D. But, to which

⁶⁴ A similar point is made by Michael Dummett in "What is a Theory of Meaning," *Mind and Language* (Oxford: Clarendon Press, 1976), pp. 105-108.

ones? It won't help to say "why to *cats* of course!"⁶⁵ What we want to know is which of the mind-independent objects are the cats. Well, the cats are the set of objects related by D to his term 'cat'. But, does this get us any further? There is the brain in a vat using his language; there are all those mind-independent objects. There are also, as the model-theoretic argument shows, countless external relations between those objects and terms of the brain's language. One of these is D, the relation of reference. If we could tell which one was D, then we could identify the mind-independent cats. However, we know no more about the referent of 'D' than we do about the referent of 'cat'. We know that 'D' refers to D, but we have no idea which relation D is. No amount of theorizing about D on the part of the brains in the vat will help. For, as we have seen, stating that relation D is the relation that makes D-theory come true on the intended interpretation picked out by D itself will not serve to pick out a unique relation D. There are still countless external relations which could satisfy such a sentence. One might make D-theory come true by connecting the brains' term 'cats' with what we would take to be mind-independent ravens; another by connecting it with what we would take to be mind-independent cats.⁶⁶ Only one of these relations is D and determines what the brains' terms 'D', 'cat', etc. refer to. And, there is no reason for us to think that D is the relation that pairs the brains' term

⁶⁵ The appearance that this will work is generated by the assumption that the brains speak English (rather than some homonymous language). Such an assumption is not legitimate at this point. We can tell that the brains speak English only once we know that each of their words (or at least most of them) refer to the same thing as each of the words in English. And, what their words refer to is precisely what is in question here.

⁶⁶ Here I am presupposing that there is no worry about what *our* terms refer to.

‘cat’ with cats rather than with ravens or some set of very gruesome objects. Not only can’t we tell which relation is D, and hence, which objects the brains’ term ‘cat’ refers to, neither can the brains themselves.

Even if we are not brains in a vat ourselves, we are in no better position to find out, except in the completely trivial sense, what our term ‘cat’ refers to. There is a real sense in which, if the relation of reference is a genuinely metaphysical external relation, one that is not empirically discoverable (as it isn’t in the case of brains in a vat), we can’t really know what we are talking about. We believe that (in most cases) we know with certainty to what our words refer; “When I [Putnam] think or say ‘the cat just went out’, the thought is usually *about* our cat Mitty; the word ‘cat’ in the sentence I think or say *refers* to a set of entities of which Mitty is a member.”⁶⁷ However, if the semantic realist’s response based on the holding of a genuine, metaphysical relation of reference between terms and mind-independent objects is correct, then Putnam might well be mistaken about what his thought ‘the cat just went out’ is really about. Putnam will be mistaken if he is a brain in a vat, since his word ‘cat’ will refer to a set of mind-independent objects and not to a set of which Mitty (a mere cat-in-the-image) is a member. But even if Putnam is, as we believe, a regularly embodied philosopher, his word ‘cat’ just might stand in the appropriate relation of reference D to cherries rather than to cats.⁶⁸ And, if he is not mistaken, this would seem to be the most amazing of

⁶⁷ *Reason, Truth, and History*, p. 43.

⁶⁸ Of course, if it does and Putnam and I speak the same language, then this sentence is false.

metaphysical coincidences, that the metaphysical relation of reference just happens to be the one that maps 'cats' on to cats, 'cherries' on to cherries, etc.⁶⁹ So, it would seem that this third option allows the semantic realist to reject premise 3 of the argument against his view, but entails a sort of referential skepticism. We succeed in referring to mind-independent objects and properties in such a way that an epistemically ideal theory could be false. But to which properties and objects do we refer? Successful communication requires not only that our words refer, but also that we know to what they refer. I view this referential skepticism as a *reductio* of this proposed attempt on the part of the semantic realist to show how reference to mind-independent objects and properties is possible. Any theory that guarantees reference but only at the price that we cannot know with certainty to what our words refer is not a plausible theory of reference.

6. To deny premise 3 of Putnam's argument, the semantic realist needs to find something which can fix a determinate relation of reference between terms of the ideal theory T and parts of the world. Otherwise, he will not be able to rule out model M's counting as the correct or intended interpretation of T (where M is the model generated by performing various model-theoretic tricks on T thought of as a string of symbols and on which T is true of the world). If he is to avoid an intrinsic intentionality theory of

⁶⁹ There is always the trivial sense in which the relation D is guaranteed to map 'cats' onto cats, 'cherries' onto cherries, etc. However, I hope that I have said enough to suggest that we can raise a genuine worry about whether D maps 'cats' onto cats in a non-trivial sense.

reference, we have seen that he has three basic choices. He can either simply give us a promissory note for a theory that will do the job, he can provide us with a theory according to which the relation of reference is an empirically discoverable relation such as empirical causality, or he can provide us with a theory according to which reference is a genuinely metaphysical relation. As we have seen, none of these alternatives is a particularly happy one. Since a mere promissory note is not a replacement for a genuine candidate reference fixer, the first can be rejected. The second does not guarantee reference to mind-independent objects and properties at all, even if there are such things. The third will allow us to know (or even believe) that there is a metaphysical relation *D* which holds between our terms and pieces of the world only if such a relation holds and entails a somewhat counter-intuitive form of reliablism. Worse, the third option seems to entail that even if we can somehow know that *D* holds, we cannot in any non-trivial sense know which objects *D* hooks our terms up to. So, it is far from obvious that the semantic realist can deny premise 3 of the argument against his view by choosing any of these options.

This result makes it seem likely that the semantic realist will have to resort to the adoption of an intrinsic intentionality theory of reference after all, if he is to deny premise 3 of the argument against his view. But, as we saw above, an intrinsic intentionality theory of reference seems to come with a fairly high price tag. Either it is difficult to know what the theory even amounts to, or, if we help ourselves to the noetic ray model of intentionality, it seems to entail that there are such necessarily existent Platonic entities as

cathood or blueness. Anyone who shares Putnam's intuition that such a "magical" theory of reference is to be avoided will have reason to accept the conclusion of the argument against semantic realism.

Recall that the adoption of semantic realism was supposed to help the metaphysical realist reconcile his view with our basic intuition that we are able to refer to (many of) the objects and properties that have genuine ontological status. It will seem to many, that the metaphysical realist is able to avoid a conflict with one, very basic intuition, only by coming into conflict with others. But what of the metaphysical antirealist? Is his view any better off in this respect? A proponent of metaphysical antirealism will not need to worry about Putnam's argument against semantic realism. Since the metaphysical antirealist does not think that there are any mind-independent objects or properties, he rejects semantic realism. He can perfectly well adopt a theory of reference, perhaps a causal theory, without worrying about whether such a theory allows him to refer to mind-independent objects and properties.

While there will still be many questions concerning just how to fill in the precise details of the theory of reference,⁷⁰ if either of the versions of metaphysical antirealism

⁷⁰ It is important to keep in mind that the difficulty being raised for the metaphysical realist in this chapter does not concern how reference is possible at all, but rather concerns how it is possible to refer to a particular range of objects and properties—the mind-independent ones. I can see no reason to think that the metaphysical antirealist is in any better position than is the metaphysical realist to provide a viable account of reference in general. The claim instead is that the metaphysical antirealist is in a better position to establish that, given that we can refer at all, we can refer to the objects and properties that have genuine ontological status.

sketched in Chapter Two is correct, reference to the objects and properties with genuine ontological status will be possible. For example, on the rational acceptability theory of truth, it will be true that there are rabbits if and only if it is rational to accept that there are rabbits. However, it will not be rational to accept that there are rabbits unless 'rabbit' refers to rabbits. So, the rational acceptability theory of truth, unlike metaphysical realism, guarantees that we can refer to the objects and properties that have genuine ontological status. Similarly, the idealized rational acceptability theory of truth guarantees that, at the ideal limit of rational inquiry, we will be able to refer to the objects and properties with genuine ontological status. So, it would seem that the metaphysical realist will have difficulty, while the metaphysical antirealist (or at least some versions of metaphysical antirealism) will not, accommodating our intuition that we can succeed in referring to the objects and properties that have genuine ontological status.

In the last three chapters, we have seen that despite the initial plausibility of his view, the metaphysical realist will have difficulty reconciling his view with other, very strong intuitions that we have about the world and our place in it. In Chapter Two, we saw that the metaphysical realist will have difficulty while the metaphysical realist will not showing that we are justified in believing that various skeptical hypotheses are false. In Chapter Three, I argued that if a very plausible way of dividing the world up into objects and properties is correct, the metaphysical realist will have difficulty while the metaphysical antirealist will not ruling out the possibility that our world is literally

crowded with very fishy but metaphysically genuine objects and properties. Here, I have argued that Putnam's model theoretical argument, while not showing conclusively that semantic realism is false, does present a great challenge to the metaphysical and semantic realist to show that determinate reference to the mind-independent world is even possible. None of these arguments show conclusively that metaphysical realism is false, or even that the metaphysical realist cannot reconcile his view with all of our basic intuitions about the world, our place in it, and our knowledge of it. However, they do suggest that metaphysical realism should not be so cheerfully awarded the default position in the debate. The metaphysical realist requires a defense of his position just as much as the metaphysical antirealist does for his.

CHAPTER FIVE: METAPHYSICAL ANTIREALISM AND OBJECTIVE TRUTH: IS METAPHYSICAL ANTIREALISM SELF-REFUTING?

In the last three chapters, we saw that there is reason to believe that metaphysical realism is not the completely unproblematic view that it at first appeared to be. The metaphysical realist requires a defense of his view just as much as the metaphysical antirealist does of his. This chapter will focus on one contemporary attempt to defend metaphysical realism offered by Peter van Inwagen. This defense takes the form of a series of arguments against metaphysical antirealism. In each, van Inwagen argues that metaphysical antirealism is not consistent with the existence of objective truth, and hence, that it is ultimately incoherent. I will argue that van Inwagen has not shown that a metaphysical antirealist needs to deny the existence of objective truth. His defense of metaphysical realism is, therefore, unsuccessful. While arguments such as these are by no means the only arguments to which a metaphysical realist can appeal in defending his view, their failure does suggest that the metaphysical realist's best defense may be a good offense. The metaphysical realist can and should demand of the metaphysical antirealist a formulation of his view that is not obviously hopeless.

1. In his book *Metaphysics*,¹ van Inwagen defends a view that he calls the Common Western Metaphysic.² According to the Common Western Metaphysic the

¹ (Boulder: Westview Press, 1993).

² In *Material Beings* (Ithaca: Cornell University Press, 1990), van Inwagen defends a metaphysic that entails that the only composite material objects that exist are living organisms. I will largely ignore his view in this regard as it is in conflict with the Common Western Metaphysic. According to the Common Western Metaphysic, the moon exists;

world contains many individual things: people, chairs, mountains, etc. These things are held to be external to us; that is, they don't merely exist in our minds, but exist out in the world. Statements about them are objectively true or false; that is, the truth or falsity of such statements is conferred on them by the way the world is, and not merely by our linguistic conventions.

Van Inwagen defines the thesis that there is such a thing as objective truth, which he calls 'Realism', as combining two assumptions that can be briefly stated as follows:

(A) A belief or assertion is true or false depending upon whether or not it correctly represents some state of affairs that obtains in the world.

(B) The world exists and has the features it does in large part³ independently of our beliefs and assertions.⁴

Assumption (A) amounts to the claim that some form of correspondence theory of truth is correct. On such a theory of truth, a proposition is true if and only if the state of affairs that it asserts to obtain in the world does in fact obtain. "It is, as one might put it, up to our beliefs and assertions to get the World right; if they don't, they're not doing their job, and that's their fault and no fault of the World's."⁵ Assumption (B) looks very much like a statement of the view that I have called metaphysical realism, the thesis that the various

according to van Inwagen's official view, although it is true to say that the Earth has a moon, the moon does not literally speaking exist.

³ This is a very minor caveat. "...our beliefs and assertions are themselves parts—very minor parts, it would seem—of the World. And, of course, our beliefs and assertions may *affect* other parts of the World, as when my false belief that the traffic light is green causes an accident" (*Metaphysics*, p. 56).

⁴ Ibid.

⁵ Ibid.

things that exist do so and have (at least some of) the features that they have independently of the workings of minds.⁶ If this resemblance is more than superficial, it would appear that what van Inwagen calls “Realism” (capital ‘R’), the thesis that there is such a thing as objective truth, entails what I have been calling metaphysical realism. If his arguments in defense of this Realism are then cogent, we will have reason to reject metaphysical antirealism. In this chapter I will examine van Inwagen’s arguments in favor of the thesis that there is such a thing as objective truth with an eye toward determining the extent to which they constitute a successful defense of metaphysical realism. I will argue that, while van Inwagen is largely successful in defending (A), the thesis that truth involves correspondence with the world, nothing that he says supports (B), the thesis that the world exists and has the features that it does independently of our beliefs and assertions.

One might worry that there is reason to think that van Inwagen’s Realism and metaphysical realism in my sense are not as closely related as they would at first appear. For, as we will see, his best argument for Realism only supports assumption (A), the claim that truth involves correspondence with the world, and fails to support metaphysical realism. Also, it seems *prima facie*, that it would be possible to hold an antirealist

⁶ Van Inwagen only specifically states that the world exists independently of *our* minds. Although van Inwagen makes it clear earlier in the book that he believes that the world and its features are dependent on the mind of God (pp. 4-5; 15-16), I am fairly certain that he would want to hold that, excluding the divine mind, the world exists and has the features that it does independently of the existence or activities of, not only our minds, but *any* mind or set of minds.

metaphysic without holding that it was the *beliefs and assertions* of cognizing agents that shaped reality, as required by a literal reading of thesis (B) of van Inwagen Realism.⁷ However, when contrasting anti-Realism in his sense and idealism, van Inwagen does indicate that he intends his Realism to be very much in line with metaphysical realism. “The antirealist who says that nothing is independent of the mind, however, really does mean something very much like this: the collective activity of all minds is somehow determinative of the general nature of reality.”⁸

Van Inwagen offers two arguments in favor of Realism. The first is a positive argument, the so-called “geological argument.” The second is a negative argument to the conclusion that anti-Realism (in his sense) is ultimately incoherent. Although it is the second of these two arguments that constitutes van Inwagen’s main defense of Realism, van Inwagen believes that it is something very like the geological argument that represents our commonsense intuition that the world and its features are independent of all human mental activity.⁹ According to the Realist, the fact *F* that *Mount Everest is 8,847.7 meters high* is completely independent of “the existence of the human mind or any activity of or fact about the human mind.”¹⁰ The geological argument is then an argument that articulates the reasons that underlie this conviction. As I see it, the geological argument is

⁷ Kant is arguably such a person. For Kant, it is not our beliefs and assertions that shape the phenomenal world. It is rather our more basic cognitive mechanisms that do the work.

⁸ *Metaphysics*, pp. 59-60.

⁹ *Ibid.*, p. 60.

¹⁰ See *ibid.*, pp. 60-61.

best interpreted as an inference to van Inwagen's Realism as the best explanation of various accepted geological facts:

1. It was the collision of the Indian subcontinent with the continent of Asia which caused and continues to cause the rise of the Himalayan mountains, and in particular Mount Everest. It is the continuing effects of this collision, along with the forces of erosion, that is responsible for Mount Everest having the size and shape that it does. This collision took place millions of years before the evolution of human kind and would have taken place even if humans had not later evolved. The same is true for the continuing effects of this collision and the operation of the forces of erosion.
2. The best explanation of the above facts is that fact F is entirely independent of all human mental activity.
3. Therefore, fact F is entirely independent of all human mental activity.

Although as formulated the argument is not deductively valid,¹¹ and thus, even if cogent, will not constitute a knock down argument for Realism, this is of little moment for van Inwagen. He believes that there is no such thing as a knock down argument in philosophy,¹² and instead holds that it is reasonable to believe what seems to us to be true

¹¹ Here is an alternative way of formulating the argument that would not only make it valid, but also dispel any worries that the mind-independence of fact F is better understood as *following from* rather than explaining the geological facts cited in premise 1. Premise 1 remains the same. Premise 2 would be: If premise 1 is true, then fact F is entirely independent of all human mental activity. The conclusion then follows by modus ponens. While this formulation of the argument would constitute a knock down argument against anti-Realism if sound, it is not at all clear why we should accept this formulation of the second premise. Any plausible version of anti-Realism will want to accept the geological facts contained in premise 1. Although it may be plausible to claim that the mind-independence of fact F in some sense best explains the geological facts cited, to claim that the geological facts logically *entail* the mind-independence of fact F would (in the absence of any further argumentation) beg the question against the anti-Realist, and if premise 2 is not a necessary truth, it will need to be supported by appeal to something like an inference to the best explanation.

¹² *Metaphysics*, p. 67.

unless we are given good reason to reject it.¹³ The geological argument is one way of representing what, according to van Inwagen, seems to us to be true. Of course we want to say that the continental collision that thrust up Mount Everest would have occurred even if other geological conditions had prevented humankind from evolving. Even if a large meteorite or volcanic eruption had thrown up a sufficient quantity of particulate matter into the atmosphere to prevent the evolution of life entirely, the tectonic motions responsible for the rise of the Himalayan Mountains might well have occurred (or so we believe). So, according to van Inwagen, we should accept the conclusion of the geological argument unless we are given good reason not to.

What the anti-Realist needs in order to show that we have reason to reject the intuitive pull of the geological argument is a plausible explanation for how the truth of the sentence expressing fact F could be conferred upon it in such a way as to be dependent upon the collective activity of minds. Van Inwagen presents such an anti-Realist argument, but then claims that the argument does not in fact establish anti-Realism. As an analysis of van Inwagen's reasons for thinking that the anti-Realist's response to the geological argument must fail will shed some light on van Inwagen's thinking about anti-Realism in general, it is to a discussion of this anti-Realist argument and van Inwagen's criticism of it that I will now turn.

The anti-Realist argument considered by van Inwagen goes as follows¹⁴:

¹³ See *ibid.*, pp. 55, 61 and *passim*.

¹⁴ *Ibid.*, pp. 61-63.

1. What we call, e.g., a mountain is a matter of convention. (We might have decided that mountains begin at the tree line rather than at sea level).
2. The same is true for the features of mountains, including their height. (We might not have decided to call the result that following a certain procedure involving a theodolite yields with respect to mountains and the result that measuring a certain weighted rope yields with respect to towers as giving the height of the thing in question).
3. Thus, mountains and height are human social constructs; all facts about the heights of mountains (including what heights certain mountains would have had if there had never been any people) are conventional.
4. Therefore, fact F is not independent of all human mental activity.

Van Inwagen seems to take the above argument as a fairly typical anti-Realist response to the geological argument. "This is, or so I maintain, a fair sample of the way in which anti-Realists argue."¹⁵ Van Inwagen thinks that the argument for the more general thesis of anti-Realism would then be an extension of this sort of argument to all other supposedly mind independent facts. While I agree that there have been anti-Realists who have attempted to defend their view with such arguments, an acquaintance with the literature makes it obvious that this sort of argument is not the only option open to the anti-Realist.¹⁶ However, for the moment, I want to focus exclusively on the above anti-Realist response, which I will call 'the social construct argument', and van Inwagen's reasons for thinking that such a response will fail.

¹⁵ Ibid., p. 63.

¹⁶ To be sure, van Inwagen is perfectly aware of this fact (see *ibid.*, p. 95). However, he does tend to talk in places as if this sort of argument were the only option open to the anti-Realist; "If this is the extent of the anti-Realists' case," he says at one point, "then I do not find it very impressive..." (p. 63).

Van Inwagen does not question the first two premises concerning the conventional nature of what we call a mountain and what we call its height. He believes that even if you grant the antirealist his premises, the conclusion that fact F is not independent of all human mental activity does not follow.

Let us grant that we might have drawn these boundaries differently if we had had different interests. Still, we *have* drawn these boundaries in a certain way, and—or so it would seem—in drawing them this way we have picked out certain objects as the objects designated by names like ‘Mount Everest’, ‘Pikes Peak’, and ‘the Matterhorn’, and so on, and there are certain properties that *these* objects will turn out to have when we get round to examining them. They will turn out to have these properties because they already have them, for these properties belong to these objects independently of the human mind and human conventions and human interests and human social activity.¹⁷

Van Inwagen goes on to say that if we had had different conventions, then we would have picked out different properties and objects than the ones we in fact pick out with our terms. But that these objects and properties have the features that they do will also be independent of the workings of minds.¹⁸ To see why this must be so, van Inwagen asks us to imagine that we have a term, ‘schmicht’, that picks out a property that the word

¹⁷ Ibid.

¹⁸ Actually this is too strong in that it essentially begs the question against the anti-Realist. What van Inwagen should say here is that if we had had different conventions, then we would have picked out different properties and objects than the ones we in fact pick out. And, the fact that these objects have just *these* properties will be independent of the linguistic conventions appealed to in the social construct argument. However, it does not follow without further argumentation that these objects will have their properties completely independently of the workings of minds. After all, the creating of linguistic conventions is but one of the many activities in which minds engage.

'height' would have picked out if our conventions had been different. Now, consider the following two statements:

Mount Everest is 8,847.7 meters high.

Mount Everest is 8,773.12 meters schmigh.

According to van Inwagen both of these statements can be simultaneously, objectively true. "All that the impressive-sounding thesis that 'height is a social construct' really comes to, therefore, is this: if we had adopted a certain different set of conventions for using the words 'height' and 'high', then the first sentence...would mean what the second means...."¹⁹ All that follows from this is that we might have spoken a different language than the one we, in fact, speak. Van Inwagen's objection to the anti-Realist's argument is that, far from supporting anti-Realism, the argument only supports a relatively trivial claim concerning the conventional nature of language.

Indeed, the "social construct" argument fails so obviously that one may begin to suspect that it cannot accurately reflect a "fair sample of the way in which anti-Realists argue," if the van Inwagenian anti-Realist, that is, the person who denies that there is objective truth in van Inwagen's sense, is to have much in common with the metaphysical antirealist. Indeed, it is perhaps natural to wonder at this point whether he is simply arguing against a strawman.²⁰ But, as I will show below, there are even more significant

¹⁹ *Metaphysics*, p.64.

²⁰ It is true, however, that Robert Schwartz, in "I'm Going to Make You a Star," *Midwest Studies in Philosophy*, vol. 11, *Studies in Essentialism*, Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, eds. (Minneapolis: University of Minnesota Press, 1986), pp. 427-440, does seem to be arguing for a view of this kind.

reasons for doubting that the social construct argument reflects a fair sample of the way in which *metaphysical* antirealists argue. Let us assume for the moment that the proponent of the social construct argument is supposed to be not only a van Inwagen anti-Realist, but also a metaphysical antirealist (perhaps because van Inwagen anti-Realism entails metaphysical antirealism²¹). Van Inwagen's reasoning is sufficient to show that the social construct argument, even if cogent, will not rule out the possibility of metaphysical realism's being true—the truth of premises 1-3 is perfectly consistent with there being individuals whose existence and features are mind-independent. In what immediately follows, I will argue that, even supposing that we had independent, cogent reasons for believing that metaphysical realism is false, the line of reasoning contained in the social construct argument would not suffice as a complete explanation of the mind-dependent nature of the world. That is, I will argue that if metaphysical antirealism is true, it must be because, at some level, the world is dependent on the workings of minds in a way not captured by the reasoning in the social construct argument.

Let us continue to assume that the social construct argument was originally offered by van Inwagen on behalf of the *metaphysical* antirealist (as well as the van Inwagen anti-Realist) as an explanation of how facts like F could be dependent on the workings of minds. Let us also suppose that what van Inwagen says with respect to the anti-Realist (in his sense), “[t]heir argument for the general thesis of anti-Realism would simply be an application of what has been said in [the social construct argument] to all supposedly

²¹ However, I will argue below that this is not the case.

mind-independent facts,”²² holds true of the metaphysical antirealist as well. If this is the case, then successive applications of the social construct argument should yield a complete explanation of the mind-dependent nature of the world. Let us see if this is so. The gist of the original social construct argument was that mountains (and their height) are mind-dependent in that they are the results of our social and linguistic conventions which impose boundaries upon or “carve up” various rock masses.²³ But of course, if we suppose metaphysical antirealism to be true, then the existence and features of rock masses must also be mind-dependent. So, we will want to explain facts concerning the existence and features of rock masses in terms of our social and linguistic conventions. Now, how does this go? It seems that the proponent of the social construct argument will need to say that all facts about rock masses are the result of our social and linguistic conventions drawing boundaries onto or carving up some more basic stuff, perhaps arrangements of quarks and electrons. But arrangements of quarks and electrons must themselves be mind-dependent if metaphysical antirealism is true. Clearly a regress threatens, if, at every level of explanation, a more basic substance must be postulated for our social and linguistic conventions to carve up. It does not seem possible that the sorts of linguistic conventions appealed to in the social construct argument could be so infinitely complex that there would be conventionally individuated entities all of the way down, so to speak. Thus, if metaphysical antirealism is true, then, if our social and linguistic

²² *Metaphysics*, p. 63.

²³ Here I follow van Inwagen in the use of the term ‘rock masses’ (see *ibid.*, p. 64). But of course, these entities will contain sand, soil, and organic material as well.

conventions can be said to carve up the world according to our interests at all, at some level the world that is thus carved up must be *in some other way mind-dependent*. That is, it cannot be *solely* in virtue of such linguistic conventions that the features of the world are dependent on minds. Successive applications of the social construct argument cannot, therefore, be all there is to the *metaphysical* antirealist's story, if indeed he has a coherent story to tell.

So, not only does the social construct argument fail in the way suggested by van Inwagen to establish metaphysical antirealism, but even if we had independent reasons for thinking that metaphysical realism was false, the social construct argument could not serve as a complete explanation of the way in which the nature of the world is dependent on the mind. I would submit that, even if van Inwagen is correct that the social construct argument represents a fair sample of the way in which anti-Realists (in his sense) argue, there is reason to think that the social construct argument offered by van Inwagen cannot reflect the sense in which the *metaphysical* antirealist wants to say that facts like F are dependent on the workings of minds. Perhaps van Inwagen's anti-Realism is not as related to metaphysical antirealism as it would at first appear.²⁴ Or, perhaps the views are more closely related and van Inwagen has somewhat misconstrued the central claim of the anti-Realist (in his sense). After all, van Inwagen admits in several places that he may not completely understand the view of the anti-Realist.²⁵ These worries arise again in

²⁴ I will suggest below that, while all metaphysical antirealist views will be versions of van Inwagen's anti-Realism, the reverse is not the case.

²⁵ See *Metaphysics*, pp. 59 and 65.

connection with van Inwagen's main argument that his anti-Realism is ultimately incoherent. I now turn to an examination of this argument.

2. As van Inwagen realizes, merely showing that the social construct argument will not work does not in and of itself constitute a defense of Realism. If the social construct argument will not work, perhaps there is some other move that a van Inwagenian anti-Realist (or a metaphysical antirealist, if these are distinct) might make in answering the geological argument, one that would make van Inwagenian anti-Realism (or metaphysical antirealism) seem at least as plausible as van Inwagenian Realism given the problems for assumption (B) of that view (i.e., metaphysical realism) examined in the previous three chapters.²⁶ To avoid such a move, van Inwagen launches a more direct attack against the anti-Realist (in his sense):

I am of the opinion that we can do more than simply show that a certain argument for anti-Realism fails to establish that thesis....We can present a very strong argument against anti-Realism. Now one might wonder how I could promise a strong argument against a thesis when, by my own testimony, I [do] not really know what that thesis is. But nothing mysterious is being proposed. I do not fully understand anti-Realism, but I do understand some of the features that anti-Realism is supposed to have. The anti-Realists have ascribed various features to anti-Realism, and many

²⁶ While providing such an explanation on behalf of the metaphysical antirealist is beyond the scope of this chapter, in the next chapter I will present and discuss an antirealist metaphysic largely based on the metaphysical program of Paul Grice. While the view has its problems, it is by no means obvious that it cannot be developed into a viable explanation of the mind-dependence of the world. Whether or not such a view, if it could be made to work, would be a better explanation of, e.g., the truth of the sentence expressing fact F than the one available to the metaphysical realist will depend in part upon the metaphysical realist's success in addressing the problems raised in the previous three chapters.

of these features are clearly taken by the anti-Realists to be essential to anti-Realism: any thesis that did not have those features would not be anti-Realism. I shall argue that any thesis that combines these features must be incoherent.²⁷

The argument that van Inwagen uses against the anti-Realist (in his sense) is very similar to arguments leveled against the verificationists. Van Inwagen asks us to consider a brief statement of anti-Realism, call it AR. The one he chooses is the following:

AR: Objective truth and falsity do not exist.

Since van Inwagen has defined 'Realism' as the thesis that there is such a thing as objective truth, an acceptance of AR would seem to be an essential feature of any version of van Inwagenian anti-Realism. He then asks us to question the status of AR itself.

Presumably, the anti-Realist will want to say that AR is true, but deny that it is true in virtue of corresponding to a mind-independent reality (since that would entail that there is at least one objectively true sentence, and hence that AR is false). But if that is the case, in virtue of what is AR true? The van Inwagenian anti-Realist needs to be able to provide some account of truth that does not require a correspondence of any kind between our true sentences and the mind-independent world. Van Inwagen thinks that he can show that this challenge cannot be met by the anti-Realist. He believes that the anti-Realist has "conspicuously failed to find a substitute for truth that satisfies the following two conditions: (a) all the uncontroversially "good" statements have it and all the uncontroversially "bad" statements lack it and (b) anti-Realism²⁸ has it."²⁹ If he is correct,

²⁷ *Metaphysics*, p. 65.

²⁸ Or better, a statement of the thesis of anti-Realism such as AR.

this will give us reason to suspect that there is no such substitute for objective truth and that van Inwagen anti-Realism is a doctrine that is ultimately incoherent. If, on any version of van Inwagen anti-Realism, AR will lack its own truth-surrogate, then van Inwagen anti-Realism will be incoherent precisely because the truth of AR implies its own falsehood.

As evidence that the anti-Realist will not be able to find a substitute for truth that will meet his two conditions, van Inwagen examines what he considers to be a typical antirealist theory of truth. "Anti-Realists have offered other substitutes for truth than this, but I am convinced that my general criticism holds: it always seems that anti-Realism itself lacks the anti-Realist's proposed substitute for truth."³⁰ According to van Inwagen, many anti-Realists will want to say that what makes true statements true is that these statements fit in with our experience, while their denials go against our experience.³¹ Van Inwagen, speaking for the antirealist has this to say about such a theory of truth:

...these statements fit in with our experience, and their denials go against our experience. For example, I have seen lions eating meat, and I have never seen any eating vegetables, their teeth are obviously fitted for meat and not for vegetables, all the lion experts *say* that lions are carnivorous, and so on. You Realists admit that there is such a status as this. It's just the status that leads you to accept or believe certain statements....we anti-Realists simply don't see the need for these two additional statuses [sic] that you call 'objectively true' and 'objectively false'.³²

²⁹ *Metaphysics*, p. 68.

³⁰ *Ibid.*, p.68.

³¹ This proposed anti-Realist truth-surrogate has much in common with the rational acceptability theory of truth discussed above, pp. 46-47.

³² *Metaphysics*, pp. 65-66.

He then argues that, while we can make some sense of a sentence such as ‘Lions are carnivorous’ fitting in with our experience more so than its denial (after all we have yet to see any herbivorous lions), it is not clear in what sense a statement of the thesis of anti-Realism such as AR fits in with our experience more than its denial. AR is, after all, a philosophical thesis. And, as van Inwagen points out,³³ philosophical theses are not confirmed or disconfirmed by sense experience, since, unlike scientific theses, they do not make any predictions concerning how our sense experiences will go. The theory that a sentence is true in virtue of fitting in with our experience (if this is taken to be restricted to sensory experience) may satisfy van Inwagen’s condition (a) for a theory of non-objective truth in that all of the uncontroversially “good” statements fit in with our experience while the “bad” ones do not, but it would appear that it does not satisfy condition (b)—AR does not itself seem to fit in with our sense experience.

Van Inwagen claims that even if we broaden our notion of experience to include our experiences of finding various philosophical arguments compelling, AR will still not satisfy condition (b). There are no knock-down arguments in philosophy, according to van Inwagen. So, there will be no argument for AR which “all competent philosophers regard as compelling.”³⁴ But perhaps requiring a knock-down argument for a philosophical thesis in order for it to fit in with our experience is too strong. Perhaps all that is required for AR to fit in with our experience in this broader sense, and thus have its

³³ Ibid., p. 66.

³⁴ Ibid., p. 67.

own truth-surrogate, is for a majority of philosophers to find arguments for AR more compelling than arguments against it. But van Inwagen is quick to retort, “whether or not this would do the trick, it is not something we have, for as matters stand, this is not how things seem to a majority of the philosophical community.”³⁵ If, as van Inwagen suggests, AR does not fit in with our experience in this sense more than its denial, the theory will be false by its own standards. So, this supposedly typical theory of non-objective truth will be incoherent. Van Inwagen concludes from this that anti-Realism (in his sense) is incoherent in general. “Anti-Realists have offered other substitutes for truth than this, but I am convinced that my general criticism holds: it always seems that anti-Realism itself lacks the anti-Realist’s proposed substitute for truth.”³⁶

Unfortunately for van Inwagen, it is far from clear that his proposed anti-Realist truth-surrogate is a truly typical truth-surrogate for the *metaphysical* antirealist. In Chapter Two, I presented two different theories of truth each of which yields metaphysical antirealism: a rational acceptability theory of truth, and an idealized rational acceptability theory of truth. Since van Inwagen’s Realism seems to entail metaphysical realism, any version of metaphysical antirealism will be a version of van Inwagenian anti-Realism. If either of these varieties of metaphysical antirealism can meet van Inwagen’s challenge (i.e., to provide a truth-surrogate that (a) all uncontroversially “good” statements have and all

³⁵ Ibid.

³⁶ Ibid., p. 68.

“bad” ones lack, and (b) that AR has), then we will have shown that van Inwagen’s defense of Realism (and hence, metaphysical realism) is inadequate.

The rational acceptability theory of truth has much in common with the “fits in with our experience” truth-surrogate discussed by van Inwagen. The rational acceptability theory of truth is clearly vulnerable to van Inwagen’s charge that it will lack its own truth-surrogate. It does not seem that it is *now* rational to accept that truth is to be equated with rational acceptability or, more generally, that objective truth and falsity do not exist.³⁷ And, in any case, the rational acceptability theory of truth has the extremely counterintuitive (if not outright absurd) implication that, for example, Ptolemaic astronomy was true, but is no longer. But what of the idealized rational acceptability theory of truth? Has van Inwagen given us any reason to think that this version of metaphysical antirealism, and hence of van Inwagenian anti-Realism, will lack its own truth-surrogate?

The answer is ‘no’. Recall that according to the idealized rational acceptability theory, truth is coextensive with what it would be rational to accept at some ideal limit of inquiry. Van Inwagen may be correct in asserting that as matters stand it is not *currently* rational to accept that truth is to be equated with idealized rational acceptability or, more generally, that there is no such thing as objective truth and falsity, as he understands these terms. However, nothing that he has said supports the claim that at the ideal limit of

³⁷ But see Michael Dummett, *The Logical Basis of Metaphysics* (Cambridge: Harvard University Press, 1991) and Jay David Atlas, *Philosophy Without Ambiguity* (Oxford: Clarendon Press, 1989), esp. pp. 125-6 for arguments to the contrary.

rational inquiry it will not be rational to accept that truth is to be equated with idealized rational acceptability or, more generally, that objective truth and falsity, as he understands these terms, do not exist. To be sure, it *may* be the case that at the ideal limit of theorizing, it will not be not rational to accept AR. It *may* be the case that the idealized rational acceptability theory lacks its own truth-surrogate. But this in itself does not show that the view is incoherent; it does not show that the view *must* lack its own truth-surrogate. After all, it is just possible that a statement of the Realist's theory of objective truth would fail to correspond to any fact in the mind-independent world, and thus that, by its own lights, Realism might be false. But this does not show that Realism is incoherent. Similarly, nothing that van Inwagen says suggests that the idealized rational acceptability theory of truth is ultimately incoherent.

But perhaps someone will want to object that while an idealized rational acceptability theory might be able to meet the second criterion (b) of van Inwagen's challenge, it does not meet (a). Since we do not now know what it will be rational to accept at the ideal limit of inquiry, we do not know that, e.g. the sentence 'Lions are carnivorous' will be rational to accept at the ideal limit of inquiry. Thus, we do not know that all uncontroversially "good" statements have this truth-surrogate while uncontroversially "bad" ones lack it. However, if this is a problem for this variety of metaphysical antirealism, it is also a problem for metaphysical realism. If truth is correspondence with the mind-independent world, then, as was argued in Chapter Two, it is hard to see how we could know that 'Lions are carnivorous' corresponds to any part of

the mind-independent world. It is by no means clear that an idealized rational acceptability theory would have more difficulty meeting the first criterion of van Inwagen's challenge than would metaphysical realism, and hence van Inwagen Realism, itself.

Since we have found one version of metaphysical antirealism, and hence of van Inwagen anti-Realism, that is not obviously vulnerable to van Inwagen's objection, we have some reason to believe that van Inwagen's proposed anti-Realist truth surrogate is not typical in the sense required to support a universal generalization. However, it might be thought that finding this one version of metaphysical antirealism that is seemingly invulnerable to van Inwagen's objection constitutes a fairly weak reply on the part of the metaphysical antirealist. Although he has not actually done so, perhaps van Inwagen, or someone else, could provide adequate reasons for thinking that at the ideal limit of theorizing, it will not be rational to accept that truth is idealized rational acceptability, that objective truth and falsity (in van Inwagen's sense) do not exist. Perhaps if van Inwagen has not himself provided an adequate defense of van Inwagen Realism (and hence metaphysical realism), he has at least shown us the way to go about doing so, by providing arguments for each proposed version of anti-Realism (in his sense) that show that it will not be able to adopt a surrogate for objective truth such that (a) all uncontroversially "good" statements have it and uncontroversially "bad" ones lack it, and (b) AR has it.

However, it seems to me that even this more modest claim is questionable. In what follows, I will argue that there may well be varieties of metaphysical antirealism, which if otherwise coherent, will be immune to van Inwagen's challenge because they

constitute a different sort of threat to van Inwagen Realism than the one considered by van Inwagen.

3. Recall that van Inwagen initially stated the view that he calls ‘Realism’ (and which he describes as the view that there is such a thing as objective truth) as the conjunction of two theses: the thesis (A) that some version of a correspondence theory of truth is correct, and the thesis (B) that we have been interpreting as a statement of metaphysical realism. Since van Inwagen’s Realism is the conjunction of two theses, it is possible to be a van Inwagen anti-Realist by denying either of those two theses. As originally formulated by van Inwagen, AR, the claim that objective truth and falsity do not exist, is general enough to capture any version of van Inwagen anti-Realism. However, when he moves to consider whether AR come out true according to the anti-Realist (a move upon which his argument against AR depends), he drastically limits the versions of van Inwagen anti-Realism to which his criticism applies. Consider, for example, the van Inwagen anti-Realist who claims that our beliefs and assertions are true or false depending upon whether or not they fit in with our experience. Such an anti-Realist is only explicitly committed to denying thesis (A) of van Inwagen Realism. That is, such a person is explicitly giving up a correspondence theory of truth in favor of what might be called an epistemic theory of truth,³⁸ while making no explicit claims about the mind-

³⁸ An epistemic theory of truth is, quite generally, a theory that defines the notion of truth in terms of what can be known or experienced. Both the rational acceptability and idealized rational acceptability theories of truth are epistemic theories of truth.

independent nature of the existence and features of the world. After all, couldn't the statement that there is a mind-independent world at least *prima facie* fit in with our experience better than its denial? Thus, it would seem that if van Inwagen's argument were cogent, it would only explicitly entail that it is incoherent to deny thesis (A) of Realism. To be sure, van Inwagen seems to assume that thesis (A) will entail (B).³⁹ So, if it were incoherent to deny (A), that is, if (A) were necessarily true, (B) would be necessarily true. However, nothing that van Inwagen explicitly says shows that it is impossible that (A) should be true and (B) false. Since it is the person who denies assumption (B) of van Inwagen Realism who is a metaphysical antirealist in my sense, without some further argumentation, it is not at all clear that van Inwagen's argument against the anti-Realist who denies (A) has any bearing on the coherence of metaphysical antirealism in general. And thus it is far from clear that van Inwagen's defense of thesis (A) of Realism, even if cogent, will constitute a defense of metaphysical realism.

But, perhaps van Inwagen has cleared the way for the necessary argumentation to show that his defense of thesis (A) of Realism (in his sense) has some bearing on the coherence of metaphysical antirealism. What we will need to consider is how the two theses of van Inwagen Realism are related. There are really two questions here: (i) Does

³⁹ In "On Always Being Wrong," *Midwest Studies in Philosophy*, vol. 12, *Realism and Antirealism*, Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, eds. (Minneapolis: University of Minnesota Press, 1988), pp. 95-111, van Inwagen identifies the thesis that truth is radically nonepistemic with the doctrine of metaphysical realism. See especially p. 97. Thus, it would seem that van Inwagen believes that the antirealist must be committed to some form of epistemic theory of truth.

the van Inwagen Realist's thesis (B) entail (A)? That is, does metaphysical realism entail a correspondence theory of truth, i.e., is the falsity of the correspondence theory of truth *sufficient* for the truth of metaphysical antirealism? And, (ii) does (A) entail (B)? That is, does the correspondence theory of truth entail metaphysical realism, i.e., is the falsity of the correspondence theory of truth a *necessary* condition for the truth of metaphysical antirealism? If the falsity of the correspondence theory of truth is a sufficient condition for the truth of metaphysical antirealism, then van Inwagen's argument against the coherence of anti-Realisms which deny (A) will have some bearing on the coherence of at least some versions of metaphysical antirealism after all. But, only if the second question can be answered in the affirmative will van Inwagen's argument serve as an objection to all forms of metaphysical antirealism; only if the falsity of the correspondence theory of truth is a necessary condition for the truth of metaphysical antirealism will van Inwagen's argument in defense of (A) count as a defense of metaphysical realism.⁴⁰

In what follows, I will argue that the answer to the first question is 'yes': some non-epistemic theory of truth *is* a necessary condition for the coherence of metaphysical realism; (B) does entail (A). If this is correct, a person who denies (A) by adopting an epistemic theory of truth *will* be a metaphysical antirealist in my sense as well as a van Inwagen anti-Realist. So, van Inwagen's argument against the coherence of an anti-Realism that denies (A) will serve as an argument against some versions of metaphysical

⁴⁰ Provided, of course, that we ignore any threat posed to the metaphysical realist by the view I have called 'nihilism'. See Chapter One above, pp. 8-9.

antirealism as well.⁴¹ However, I will argue that the answer to the second question may well be 'no'. I will argue that nothing that van Inwagen says entails that a metaphysical antirealist *must* be committed to the rejection of a correspondence theory of truth, and hence, that (A) may not entail (B). It will follow that van Inwagen's argument, even if it were cogent, would only entail that it is incoherent to deny thesis (A) of Realism and not that it is incoherent to deny (B). Van Inwagen's argument against the coherence of anti-Realism would, even if cogent, therefore, fail to constitute a successful defense of metaphysical realism.

To address the first question first, does van Inwagen's thesis (B) entail (A)? Or to put it another way, does a rejection of the correspondence theory of truth (or more importantly an adherence to an epistemic theory of truth) commit one to a denial of metaphysical realism? Michael Devitt⁴² argues that it does not. Devitt claims that metaphysical realism does not entail any semantic theory concerning truth or reference.⁴³ It is interesting to note that although Devitt's claim if correct will actually provide an escape for the metaphysical antirealist from van Inwagen's criticisms (since if cogent, it will show that van Inwagen's argument has little or no bearing on either metaphysical

⁴¹ Actually, we already knew this since the rational acceptability theory of truth, which is a version of metaphysical antirealism, is nearly identical with van Inwagen's proposed anti-Realist truth-surrogate. However, I think that it will still be useful to show *why* van Inwagen's challenge is relevant to some versions of metaphysical antirealism.

⁴² *Realism and Truth* (Cambridge: Basil Blackwell, 1991), esp. pp. 39-59.

⁴³ He also argues that metaphysical realism is not entailed by any semantic theory. But I will ignore this point here.

realism or metaphysical antirealism), Devitt's own motive for making the claim is to defend metaphysical realism.

Devitt notes that many arguments against metaphysical realism are semantic in nature, and offers the following argument as an example of a common type of argument against realism.⁴⁴

1. If the realist's mind-independent world exists, then our theories, thoughts, etc., must represent or refer to that reality.
2. Our theories, thoughts, etc. cannot represent or refer to a mind-independent world.
3. Therefore, the realist's mind-independent world does not exist.

Devitt also notes that many of the arguments in the metaphysical realism/antirealism debate have centered around the truth or falsity of premise 2.⁴⁵ However, Devitt believes that it is premise 1 that is false. Premise 1 commits the metaphysical realist to a correspondence theory of truth in demanding that our theories, thoughts, etc., refer to a mind-independent reality. Devitt believes that he can defend metaphysical realism from many of its critics by severing the metaphysical position from any particular semantic theory. Metaphysical realism, Devitt claims, neither entails nor is entailed by any particular semantic theory, and hence *a fortiori* need not entail the correspondence theory of truth. I will not primarily concern myself here with whether or not Devitt's defense of

⁴⁴ *Realism and Truth*, p. 49.

⁴⁵ See especially, Hilary Putnam's so called brains-in-a-vat argument, pp. 1-21 of *Reason, Truth, and History*, and his argument against semantic realism, discussed above in Chapter Four.

realism from semantic arguments for antirealism is ultimately successful.⁴⁶ My primary concern is instead whether or not he succeeds in showing that the denial of a correspondence theory of truth in general, and an acceptance of an epistemic theory of truth in particular, is compatible with metaphysical realism.

Supposing that Devitt is correct that the metaphysical realist is not committed to the correspondence theory of truth, we can now ask, 'what are his other options?' Besides correspondence truth, Devitt distinguishes two other type of truth theory: deflationary theories and epistemic theories, each of which he believes to be metaphysically neutral. On a deflationary theory of truth, there is nothing to the truth of a sentence over and above the holding of the relevant T-schema, a biconditional of the form '['s' is true iff p]' where p expresses the same proposition as s.

...to refer to a sentence and say that it is true is just to assert the sentence, while acknowledging one's lack of originality....On this theory, the truth term is not really a predicate, and does not describe or say anything about the sentence....⁴⁷

According to Devitt, a deflationary theory of truth differs markedly from a correspondence theory in that it does not attempt to explain truth in terms of correspondence to a way that the world is. On a deflationary theory, the concept of truth

⁴⁶ Actually, I think that it is premise 2 of the semantic argument for antirealism that the realist should be concerned to deny. Whether or not the metaphysical realist is logically committed to correspondence truth, it is clear that many realists have held a correspondence theory of truth. And, as we saw above (Chapters One and Four), the correspondence theory of truth is an important part of the doctrine of semantic realism which the metaphysical realist needs in order to accommodate our intuition that we can refer to mind-independent objects.

⁴⁷ *Realism and Truth*, p. 31.

is held to be simple and unanalyzable, and indeed, disappears into the T-schema. Since deflationary theories of truth make no metaphysical claims whatever, such theories would seem to be compatible with both metaphysical realism and antirealism, to the extent to which each is coherent.⁴⁸ If Devitt is correct that a deflationary theory of truth is compatible with metaphysical antirealism, then there will be an obvious way for the antirealist to escape van Inwagen's criticism: by adopting a deflationary theory of truth rather than the epistemic theory van Inwagen anticipates. However, I know of no metaphysical antirealist who has adopted such a theory⁴⁹ and will not discuss it further. Let us turn now to his claim that epistemic theories of truth are compatible with metaphysical realism as well as antirealism.

According to Devitt, an epistemic theory of truth ties truth to what is, at least in principle, knowable. On such a view, the idea of truth may be equated with such notions as warranted assertability, verifiability, rational acceptability at the ideal limit of theorizing, etc. It should be clear that the theory of truth considered by van Inwagen's anti-Realist, which ties truth to what "fits in with our experience," is a (vague and simplistic) version of an epistemic theory of truth. Devitt claims that epistemic theories of truth are not necessarily incompatible with metaphysical realism. *"An argument is needed to get from a*

⁴⁸ Devitt concludes from this that a metaphysical realist will be able to deny premise 1 of the semantic argument for antirealism, and thus avoid the conclusion, by adopting a deflationary theory of truth.

⁴⁹ Unless, of course, the Wittgenstein of the *Philosophical Investigations* (New York: Macmillan, 1958), who appears to accept a deflationary theory of truth (see sections 22 and 136), is to be interpreted as a metaphysical antirealist. I, however, am not convinced that this is the best way to interpret the later Wittgenstein.

theory of truth to anti-realism."⁵⁰ If, as Devitt suggests, the metaphysical realist too can adopt an epistemic theory of truth, van Inwagen's argument against such theories of truth will have no clear bearing on the coherence of metaphysical antirealism unless an antirealist must be committed to an epistemic theory of truth.⁵¹ Devitt concedes, however, that such theories of truth do raise difficulties for metaphysical realism. He believes that metaphysical antirealism follows by abduction if an epistemic theory of truth is correct, as it is difficult for a metaphysical realist to explain the connection between, e.g., our being warranted in asserting that *p* is true and *p*'s being true in a mind-independent way. But since the metaphysical realist might be able to come up with a plausible explanation, or so it seems to Devitt, it would appear that there is no logical incompatibility between realism and epistemic theories of truth.

However, it is not at all clear that any meaningful metaphysical realism is consistent with an epistemic theory of truth. Consider the claim that 'water is H₂O' is true because rationally acceptable. Suppose that such an analysis of the truth of 'water is H₂O' is correct *and* that metaphysical realism is also true on this analysis. But given this theory of truth, to assert that it is true that, independent of all minds, there is water that is H₂O, is just to assert that the sentence 'independent of all minds, there is water that is H₂O' is rationally acceptable. And, or so it would seem, such a sentence could be rationally acceptable, and hence true on this theory, even if there is in fact nothing in the mind-

⁵⁰ *Realism and Truth*, p. 45, (emphasis in the text).

⁵¹ However, the metaphysical realist will still gain some advantage since he will be able to avoid at least some of the skeptical worries raised in Chapter Two.

independent world that has any of the properties water is thought to have. Similarly, on this analysis, metaphysical realism will be true if and only if it is rational to accept metaphysical realism. But, it could be rational to accept metaphysical realism even though there is in fact nothing that exists and has the features that it does independently of any and all minds. If this gives us metaphysical realism, it would seem to have been robbed of all of its metaphysical punch. It would seem that an epistemic theory of truth does not lend itself to the expression of realism as a genuinely metaphysical doctrine. The only way to make the expression of the truth of *metaphysical* realism consistent with an epistemic theory of truth would be to exempt it from falling under the epistemic analysis of truth. But this is just to give up the view that an epistemic theory of truth provides a complete and accurate analysis of truth. I would submit that an epistemic theory of truth is not consistent with any meaningful assertion of metaphysical realism. And thus, despite Devitt's protestations to the contrary, metaphysical realism would seem to entail a correspondence theory of truth (at least if we provisionally ignore the deflationary theory of truth). So, if van Inwagen's argument against epistemic theories of truth were cogent,⁵² it would suffice to show that at least some varieties of metaphysical antirealism are untenable (the ones that deny the correspondence theory of truth).

But would it constitute an adequate defense of metaphysical realism? Only if the metaphysical antirealist *must* be committed to some epistemic theory of truth or other and cannot himself adopt a correspondence theory of truth, that is, only if van Inwagen's thesis

⁵² And I have already provided some reasons to think that it is not.

(A) entails thesis (B). Although van Inwagen is correct that many versions of metaphysical antirealism entail the denial of a correspondence theory of truth, it seems to me false that all forms of metaphysical antirealism would necessarily entail such a denial.⁵³ This possibility seems to be commonly ignored by many metaphysical realists, in part because of a tendency to build metaphysical realism into the very criteria for correspondence truth. While van Inwagen is not himself guilty of such a move at the outset—he is careful to distinguish the assumption that a correspondence theory of truth is correct from the assumption that the world has the features that it does independently of minds and the activity of minds—he does talk as if denying (A), the thesis of correspondence truth, were the only avenue open to the anti-Realist (in his sense). Despite the fact that he wants to claim that correspondence truth is neither necessary nor sufficient for metaphysical realism, Devitt explicitly builds metaphysical realism into his definition of correspondence truth as a third criterion:

Sentences of type *x* are true or false in virtue of: (I) their structure; (II) the referential relations between their parts and reality; (III) the objective and mind-independent nature of that reality.⁵⁴

Given this definition of correspondence truth, it looks as if correspondence truth à la Devitt does entail metaphysical realism, as if the adoption of such a theory of truth is not an option for the antirealist.

⁵³ Indeed, the metaphysical program advocated by Paul Grice which I will discuss in Chapter Six is plausibly interpreted as a version of metaphysical antirealism which is consistent with the claim that truth (at least generally) involves correspondence with the world.

⁵⁴ *Realism and Truth*, p. 29.

However, Devitt acknowledges that we could drop criterion (III) from the definition of correspondence truth and retain all of the merely semantic import of the theory.⁵⁵ Without criterion (III), correspondence truth is, I believe, compatible with metaphysical antirealism.⁵⁶ One way of making sense of what correspondence truth would look like on an antirealist metaphysic is by slightly adapting Kant's distinction between the noumenal and the phenomenal worlds. The noumenal world is the world of things *an sich*, in themselves, independent of the existence of and workings of any mind or minds. Although Kant held that the noumenal world was essentially unknowable and the metaphysical realist holds that it is at least possible to have genuine knowledge of the mind-independent world, it is plausible to equate Kant's noumenal world with the realist's mind-independent world. When a metaphysical realist like van Inwagen holds a correspondence theory of truth, he holds that a sentence is literally true if and only if it corresponds with a way that the world, the now knowable noumenal world, really is. The phenomenal world is, on the other hand, the world of things as known, or as perceived. It includes not only the entities directly perceived by the senses, but also those referred to by our best justified scientific theories. The phenomenal world, unlike the noumenal world, is dependent on the existence and workings of minds.⁵⁷ While the metaphysical antirealist

⁵⁵ Ibid., p. 42.

⁵⁶ At least provided that the version of metaphysical antirealism is otherwise coherent.

⁵⁷ Even if our ways of perceiving and knowing do not in any way corrupt the information that they receive from the noumenal world and the noumenal world and phenomenal worlds are qualitatively identical, they will not be identical in the strict sense. The existence and features of the noumenal world are completely mind-independent while the existence and features of the phenomenal world are, at least partially, dependent on the

will deny that our beliefs and sentences can be said to correspond in any meaningful sense with the noumenal world (largely because such a world cannot coherently be said to exist according to the metaphysical antirealist), he can consistently maintain that what makes our beliefs and sentences true or false is whether or not they correspond to, represent, or refer to some state of affairs in the *phenomenal* world (at least he can if the notion of a phenomenal world makes sense).⁵⁸

While such an account of truth obviously needs to be developed in much more detail to be fully evaluated, it should be clear from this brief account that it will not suffer from the problems that van Inwagen sees for epistemic, anti-Realist theories of truth. Van Inwagen has challenged the antirealist to provide criteria for truth that all uncontroversially “good” statements have and all “bad” ones lack, criteria that are satisfied by a statement AR of antirealism (such as the earlier claim that there is no such thing as objective truth) as well. Earlier, we saw that van Inwagen at least has not shown that *no* epistemic theory of truth can meet this challenge. For all that he has said, an idealized

existence and working of minds. If there had been no noumenal minds, there would have been no phenomenal world.

⁵⁸ Laurence Bonjour has expressed to me the worry that the phenomenal world, as I have introduced it here, is not literally a world at all. Rather, it is the way the noumenal world appears to us. Others will, no doubt, share this worry. However, it seems to me that this is one way of expressing the worry that the metaphysical antirealist will not be able to produce a coherent, fairly complete metaphysical picture. And, as development of a coherent, fairly complete antirealist metaphysic will be the primary topic of Chapter Six, I will defer discussion of this worry until later. For now, I ask those who share Bonjour’s worry to grant me the provisional assumption that correspondence with a phenomenal world makes sense. If the metaphysical antirealist cannot provide a coherent positive view, he has bigger problems than van Inwagen’s arguments against his view.

rational acceptability theory of truth can meet the challenge. The worry was that some further argument might show that even the idealized rational acceptability theory of truth lacks its own truth-surrogate. However, we can now formulate a possible response to van Inwagen's challenge for the metaphysical antirealist that is not anticipated by van Inwagen. The metaphysical antirealist can respond by saying that good statements correspond with the world while bad statements fail to so correspond, but that, contra the metaphysical realist and proponent of van Inwagenian objective truth, the world cannot meaningfully be said to be independent of the mind and the workings of minds.⁵⁹ Unless we have independent reasons for thinking that the notion of a world that is in any way dependent upon the workings of minds is incoherent, there can be no more reason to think that it would be impossible for AR to correspond to such a world, than it would be for a statement of metaphysical realism to correspond to a mind-independent world. Such a version of metaphysical antirealism will be immune to van Inwagen's challenge. While I

⁵⁹ Robert Schwartz at first looks to be expressing a similar view in "I'm Going to Make You a Star" when he writes, "Acknowledgment of our role in constructing the world does not entail a subjective account of truth and certainly does not require denying that 'P' is true if and only if P. Indeed, when it comes to accounts of truth, the whole question of making really does not make a difference"(pp. 428-9). However, it is far from clear to me that Schwartz is a metaphysical antirealist at all. While he clearly believes that there are no mind-independent *kinds* of individuals, his view seems to presuppose metaphysical realism with respect to individuals. "Each of the objects that falls under the label 'star' exhibits an unlimited number of features, but what makes them each a star has to do with the commonalties among the set. Yet, commonalties are all too easy to find. Any set of items can be understood to be a grouping based on some commonality or other. What is important for being a star is sharing the right features. But which features are the appropriate ones, and what makes this grouping or kind the star-kind, is not to be found written in the stars"(ibid., p. 435).

do not want to claim at this point that such a view is in fact coherent,⁶⁰ the *prima facie* possibility of such a view which would be completely immune to van Inwagen's challenge is sufficient to indicate that van Inwagen has not even indicated the way an adequate defense of metaphysical realism could go. He has not considered all of the *prima facie* possible ways to deny his Realism. And, as a result, van Inwagen's argument, considered by itself, fails to adequately defend metaphysical realism from all versions of metaphysical antirealism.

4. One last point: since some might think that any view which entailed the denial of the existence of objective truth (which any version of metaphysical antirealism will, as van Inwagen has formulated the notion of objective truth) is absurd, it is important to note that while van Inwagen builds into his account of objective truth the doctrine of metaphysical realism, it is difficult to see what would prevent a correspondence theory of truth paired with metaphysical antirealism (provided that the conjunction is consistent) from counting as a theory of "objective truth" in some sense of the words. According to van Inwagen, what makes truth according to some theory objective truth is that "truth and falsity are conferred on those beliefs and assertions by their objects, by the things they are *about*."⁶¹ The anti-Realism considered by van Inwagen that holds that a statement is true

⁶⁰ In Chapter Six, I will provide some reasons for thinking that such a view can be made coherent. However, it is likely that not all will be convinced.

⁶¹ *Metaphysics*, p. 56.

if it fits in with experience clearly does not meet this condition. Consider again what van

Inwagen has to say in explanation of this epistemic theory of truth:

...these statements fit in with our experience, and their denials go against our experience.... You Realists admit that there is such a status as this. It's just the status that leads you to accept or believe certain statements. And you concede that there are statements that have this status and are nevertheless not what you call 'objectively true', since you concede that a misleading series of experiences could cause someone to accept, say, the statement that lions are herbivorous, which you regard as 'objectively false'. Well, we anti-Realists simply don't see the need for these two additional statuses [sic] 'objectively true' and 'objectively false'. We are content with the statuses 'fits in with our experience' and 'goes against our experience'.⁶²

It is clear why van Inwagen wants to say that on such a theory of truth the truth and falsity of our assertions is not conferred on them by the things that they are about. If no one had had the relevant sort of experience with lions, the statement that lions are herbivorous might very well fit in with our experience and be true on such a theory, even though lions *really* are carnivorous. Since the statement that lions are herbivorous is a statement about lions, it seems quite odd that it could be made true or false independent of any facts about lions themselves. It is not surprising that van Inwagen finds such a theory completely implausible.

However, according to an antirealist correspondence theory such as the one briefly sketched above (if such a view is coherent), the truth or falsity of statements *is* conferred upon them by their objects. 'Lions are herbivorous' will be true if and only if lions, actual lions, eat nothing but vegetables. It is just that on this view actual lions will be inhabitants

⁶² Ibid., pp. 65-66.

of something very like Kant's phenomenal world, rather than the metaphysical realist's mind-independent world. If such a view is in the final analysis coherent, it would seem to be consistent with much of what van Inwagen wants to say about objective truth.

But the metaphysical antirealist is not entirely limited to the adoption of a correspondence theory of truth in this regard. At least one of the important hallmarks of objective truth noted by van Inwagen can be met even by an epistemic theory of truth such as the idealized rational acceptability theory. According to van Inwagen, an important component of truth's being objective is that truth be "one":

If two people, you and I, say, have the same belief about something—perhaps we both believe that Albany is the capital of New York State—then truth or falsity is conferred on our common belief by the features of that one object. Truth is therefore "one"; there is no such thing as a belief or assertion being "true for me" but "not true for you."⁶³

The idealized rational acceptability theory of truth allows for the existence of "objective" truth, thought of as the existence of one and only one set of true statements such that no statement would ever be "true for me" but "not true for you." At the ideal limit of theorizing, there may well be one and only one set of ideally rationally acceptable statements.⁶⁴ If this were the case, there would be no cases wherein a statement was true for me, but not for you. Thus, even the more epistemic idealized rational acceptability

⁶³ Ibid., pp. 56-7.

⁶⁴ This may well not be the case, however. And, it may be thought by some that the notion of objective truth necessitates a unique set of true statements. The mere *possibility* that there will be a unique set of true statements entailed by an idealized rational acceptability theory of truth may not be sufficient for objective truth in this sense.

theory of truth is compatible with the existence of a limited notion of objective (inter-subjective) truth.

So, van Inwagen's arguments have been shown to be ultimately unsuccessful in showing the metaphysical antirealism is incoherent. He has not shown that the metaphysical antirealist is necessarily committed to an unsatisfactory semantic theory. Actually, the fact that some varieties of metaphysical antirealism might turn out to be consistent with the notion of objective truth should not really be surprising; the main point of contention between metaphysical realists and metaphysical antirealists is over the nature of the world itself, and not over whether or not truth is in any sense objective. The metaphysical realist will have to look elsewhere for a defense of his view.

At this point, the metaphysical realist may, quite justifiably, feel that he has no idea to what assumptions he may appeal in defense of his view without a clearer idea of what the position advocated by the metaphysical antirealist *does* entail. What exactly does it mean to say that the existence and features of the individuals that make up the world are dependent on the existence and activities of some set of minds? It is hard to see how such a state of affairs could even be possible. Even if there is nothing inherently problematic in the notion of a mind-dependent entity, if the world contains minds (as surely it does!), how can the existence and features of minds be dependent on the existence and activities of minds? If minds are to play any causal role in the creation and shaping of the world as the metaphysical antirealist would seem to be suggesting, it would seem that minds must

antecedently exist! It begins to look as if the metaphysical realist's best defense may in fact be a good offense. He has every right to a formulation of metaphysical antirealism and every right to believe, in the absence of some evidence to the contrary, that this quite legitimate demand will be nearly impossible for the metaphysical antirealist to meet. In the next chapter, I will explore the possibility that the metaphysical antirealist can meet this demand and present a formulation of his view that is both coherent and fairly substantive.

CHAPTER SIX: HOW TO BE A METAPHYSICAL ANTIREALIST?

In previous chapters, I have argued that the metaphysical realist should be concerned to provide a defense of his view if his is to remain the default view in the debate. I have shown that van Inwagen's attempt at a defense of metaphysical realism fails because it involves unwarranted assumptions about the commitments of the antirealist. But, perhaps, as I suggested at the end of the last chapter, the best defense in this case is a good offense. One promising line of defense for the metaphysical realist is to directly challenge the antirealist to provide details of his positive view. Thus far, we know that the proponents of various kinds of semantic theories are rationally committed to the adoption of an antirealist metaphysic. We also know that the metaphysical antirealist motivates his thesis by pointing to various difficulties with metaphysical realism. But, what exactly is the metaphysical antirealist's thesis? We have been told that the metaphysical antirealist does not deny that there are any individuals, but holds that neither their existence nor any of their features are completely independent of the operations or existence of any mind or minds. This would certainly seem to commit the metaphysical antirealist to the positive thesis that everything that exists is in some way dependent on the workings of minds. But, such a thesis seems, on the face of it, quite absurd. If the metaphysical antirealist cannot establish that he has a viable positive view, it would seem that the metaphysical realist can rest secure until such time as the antirealist is able to do so.

In this chapter, I will examine some ways in which a metaphysical antirealist might try to address some of these worries. I will first examine the antirealist metaphysic, called

'internal realism', developed and defended by Hilary Putnam. This metaphysic, which Putnam has abandoned in his latest writings, arises out of an idealized rational acceptability theory of truth. Mark Johnston¹ has provided two arguments which purport to show that Putnam's internal realism is untenable. In the first, Johnston argues that it is quite plausible that there are sentences which are true or false whose truth or falsity would be beyond our epistemic access even at the ideal limit of theorizing. In the second he argues that Putnam's internal realist cannot explain the fact that a certain statement would be acceptable at the ideal limit of theorizing without appealing to something very like the semantic realist's notion of truth as correspondence with a fact about the world. These arguments will be the subject of the second section of this chapter. While it seems to me that the internal realist does have a response to the first of these arguments, the second seems to constitute a devastating objection to Putnam's view. In the third section, I will examine the constructivist metaphysical program discussed by Paul Grice in the third of his *Carus Lectures*, entitled "Metaphysics and Value,"² and again in his "Reply to Richards."³ This program yields, or is at least consistent with, a form of metaphysical antirealism which may well be able to escape both of Johnston's arguments. The metaphysical picture that results from Grice's program provides, I think, a fairly coherent

¹ "Objectivity Refigured: Pragmatism without Verificationism," in *Reality, Representation and Projection*, John Haldane, ed. (New York: Oxford University Press, 1993), pp. 85-130.

² All page references to these lectures refer to an unpublished manuscript.

³ *Philosophical Grounds of Rationality*, Richard Grandy and Richard Warner, eds. (New York: Oxford University Press, 1986), pp. 45-106.

initial sketch of what a fully worked out version of metaphysical antirealism might look like. To be sure, the view as sketched is not without its difficulties. But, it does not seem obviously incapable of being developed into a view which could be a persuasive contender.

1. Although he seems to have officially given up the last remnants of metaphysical antirealism in his view⁴ in favor of a more purely pragmatic approach to philosophy, Hilary Putnam has had as much influence on the shaping of metaphysical antirealist views as any other contemporary philosopher. In *Reason, Truth, and History*, Putnam defends a version of metaphysical antirealism which has come to be known as internal realism.⁵ In the preface, he writes:

The view which I shall defend holds, to put it very roughly, that there is an extremely close connection between the notions of *truth* and *rationality*....

I shall advance a view in which the mind does not simply 'copy' a world which admits of description by One True Theory. But my view is not a

⁴ In "The Dewey Lectures 1994: Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind," *The Journal of Philosophy* 91 (Sept. 1994): pp. 445-517, Putnam writes in a footnote (p. 448), "I regret having myself spoken of 'mind dependence' in connection with these issues in *Reason, Truth, and History* (New York: Cambridge, 1981)." Putnam expresses the belief that both the notion that the world is mind-independent and his earlier view that the world is somehow dependent on the mind reflect the mistaken notion that there is an interface between our cognitive powers and the external world. Indeed, Putnam seems to go so far as to reject the idea that there is a genuine distinction between the mind and the objects of perception.

⁵ In *Reason, Truth, and History*, Putnam writes "The perspective I shall defend has no unambiguous name. It is a late arrival in the history of philosophy....I shall refer to it as the *internalist* perspective, because it is characteristic of this view to hold that *what objects does the world consist of?* is a question that it only makes sense to ask *within* a theory or description"(p. 49).

view in which the mind *makes up* the world, either (or makes it up subject to constraints imposed by 'methodological canons' and mind-independent 'sense-data'). If one must use metaphorical language, then let the metaphor be this: the mind and the world jointly make up the mind and the world. (Or, to make the metaphor even more Hegelian, the Universe makes up the Universe—with minds—collectively—playing a special role in the making up.)⁶

These passages are, while somewhat suggestive, quite puzzling. It is clear that internal realism will involve some sort of idealized rational acceptability theory of truth, and hence, will be a version of metaphysical antirealism. However, it is far from clear what we should make of Putnam's Hegelian metaphor. Is the metaphor even coherent? In order to make enough sense of Putnam's internal realism to be able to provide a critical analysis of it, it will first be necessary to examine his motivations for the adoption of such a view.

Putnam first used the term 'internal realism' in "Realism and Reason" to refer to a type of view concerning the relation between speakers and the world.

The realist explanation, in a nutshell, is not that language mirrors the world but that *speakers* mirror the world—i.e. their environment—in the sense of *constructing a symbolic representation of that environment*. In 'Reference and Understanding'⁷ I argued that a 'correspondence' between words and sets of things...can be viewed as part of an *explanatory model* of the speakers' collective behavior....[L]et me refer to realism in this sense—acceptance of this sort of scientific picture of the relation of speakers to their environment, and of the role of language—as *internal realism*.⁸

Putnam's appeal to the fact that speakers mirror the world (thought of as their environment) and the explanatory role played by the speakers' use of words to refer to

⁶ pp. x-xi.

⁷ In *Meaning and the Moral Sciences* (London: Routledge and Kegan Paul, 1978). First delivered in Jerusalem, May 1976.

⁸ Ibid., p. 123.

objects in their environment suggests that, in its earliest form, internal realism closely resembled the empirical causal theory of reference discussed in Chapter Four.⁹ The reader will recall that, while I argued that the adoption of such a theory of reference does not sit well with semantic realism¹⁰—the view that (i) (many of) our terms refer to mind-independent objects and their mind-independent features and (ii) (many of) our sentences assert that mind-independent states of affairs obtain—it is, nonetheless, metaphysically consistent with both metaphysical and semantic realism. So, why should we take the internal realism of “Realism and Reason” to be a form of metaphysical antirealism?

The answer to this question lies partly in the fact that Putnam believes himself to have an argument against semantic realism. This argument was discussed in detail in Chapter Four above, but it will be useful to briefly sketch it here. If metaphysical realism is to be a viable philosophical theory, semantic realism must be true—our words, or at least many of them, must refer to entities in a mind-independent world. If semantic realism is true, then there must be a fixed and determinate relation between our words and mind-independent entities. Putnam’s argument against semantic realism attempts to show that the only way such a relation could be fixed is if mental states *intrinsically* refer to various mind-independent entities or states of affairs in such a way that the reference of a mental state may be completely independent of the role the mental state plays in the lives of cognitive agents or, alternatively, if the mind possesses the ability to grasp or intuit

⁹ Above, pp. 129-135.

¹⁰ See especially, pp. 133-135.

mind-independent propositions or forms. Although there are philosophers who are not troubled by the idea of our having such mental powers, Putnam is by no means alone in finding an appeal to such powers somewhat magical or occult. It is for this reason that he rejects semantic realism, and with it, the view that a sentence will (generally) be true if and only if the state of affairs which the sentence asserts to obtain actually does obtain in the mind-independent world.

However, even once semantic realism has been rejected, it will still require some work to show that Putnam's original version of internal realism really amounts to a form of metaphysical antirealism.¹¹ After all, why couldn't metaphysical realism still be true in some sense even if semantic realism were false and a causal or more complex functionalist theory of reference were true? It is important to realize that a causal or functional theory of reference of the sort accepted by Putnam does not in and of itself entail any particular theory of truth. In rejecting semantic realism, we have given up the notion that the truth of many types of sentences necessarily involves correspondence with mind-independent states of affairs. However, we aren't automatically driven to the adoption of a theory of truth, such as a rational acceptability or idealized rational acceptability theory of truth,

¹¹ To be sure, Putnam seems to think that his argument against semantic realism in "Realism and Reason" also shows that metaphysical realism is false. He explicitly says that he takes himself to be arguing against a view one important consequence of which is that "truth is supposed to be radically non-epistemic" ("Realism and Reason," p. 125). So, it would seem that Putnam does take himself to be arguing in favor of an epistemic theory of truth of the sort that yields metaphysical antirealism. However, as I will argue in the remainder of this paragraph, it is not at all obvious that we should simply take Putnam's word that internal realism, as defined in "Realism and Reason" (p. 123), amounts to a version of metaphysical antirealism.

which constitutes a version of metaphysical antirealism. Suppose, for example, that we wanted to retain as much of the semantic realist's notion of truth as possible and defined truth as correspondence between the semantic content of a sentence and a state of affairs in the empirical world whether or not the existence and features of such a world are independent of the working of minds. Now consider the following sentence:

MR: There are things that exist and have (at least some of) the features that they do independently of the existence and workings of any mind or minds.

This sentence will be true if it should happen that any of the objects in the empirical world exist and have (at least some of) the features that they do independently of the existence and workings of any mind or minds; otherwise, it will be false. It seems to me that, despite his explicit statement to the contrary,¹² nothing in Putnam's original formulation of internal realism or in his argument against semantic realism rules out the possibility that MR is true. To rule out the possibility that MR is true given internal realism as defined in "Realism and Reason," one must either explicitly add to the view a theory of truth which entails metaphysical antirealism or a view which entails that nothing in the empirical world will exist independently of the workings of minds.¹³

¹² "The most important consequence of metaphysical antirealism is that truth is supposed to be radically non-epistemic....And it is this feature that I shall attack!" (ibid., p. 125).

¹³ Without a philosophical defense of such a move, Putnam's argument against semantic realism will not constitute a defense of metaphysical antirealism. However, such a move may be motivated by the problems for metaphysical realism discussed in Chapters Two through Four. And, in any case, it is not my goal here to provide a defense of metaphysical antirealism. My aim is instead to present and examine a viable version of metaphysical antirealism. A complete defense of such a view is well beyond the scope of this dissertation.

In *Reason, Truth and History*, Putnam seems to have reinterpreted or enlarged upon internal realism in such a way that it involves both of the above suggestions for making it explicitly into a version of metaphysical antirealism. In places Putnam states that internal realism is the view that truth just is, or is at least coextensive with, idealized rational acceptability:

‘Truth’, in an internalist view, is some sort of (idealized) rational acceptability—some sort of ideal coherence of our beliefs with each other and with our experiences *as those experiences are themselves represented in our belief system*—and not correspondence with mind-independent or discourse-independent ‘states of affairs’.¹⁴

In still other places, he suggests that we construct the world and the objects it contains, and thus, that all of the objects the world contains are to some extent mind-dependent.

If, as I maintain, ‘objects’ themselves are as much made as discovered, as much products of our conceptual invention as of the ‘objective’ factor in experience, the factor independent of our will, then of course objects intrinsically belong under certain labels because those labels are the tools we used to construct a version of the world with such objects in the first place. But *this* kind of ‘Self-Identifying Object’ is not mind-independent.¹⁵

It is to an explication and critique of this seemingly incoherent view that I will now turn.

Although Putnam never quite puts it this way,¹⁶ the internal realism of *Reason, Truth, and History* has two main, interrelated components: (i) an idealized rational

¹⁴ *Reason, Truth, and History*, p. 49.

¹⁵ *Ibid.*, p. 54.

¹⁶ He comes close to putting it this way on p. 201 when he writes, “A fact is something that it is rational to believe, or, more precisely, the notion of a fact (or a true statement) is an idealization of the notion of a statement that it is rational to believe....And I argued that being rational involves having criteria of *relevance* as well as criteria of rational acceptability, and that all of our values are involved in our criteria of relevance.”

acceptability theory of truth, and (ii) a pragmatic relevance condition on rational acceptability. Concerning the first component of internal realism, it will be useful to review here what was said in Chapter Two above. According to an idealized rational acceptability theory, truth is, by its very nature, coextensive with what it would be rational to accept at the ideal limit of inquiry. A sentence would be rational to accept at the ideal limit of inquiry if, for example, it is analytic or would be judged to be the conclusion of a sound deductive argument or to have been directly verified by observation or to contribute to the explanatory power of other well-confirmed theories.¹⁷ As it is this first component of internal realism that is most relevant to our discussion of the extent to which a coherent version of metaphysical antirealism can be formulated, I will have quite a bit more to say about the details of an idealized rational acceptability theory of truth below. But first, I would like to briefly discuss the second component of Putnam's internal realism, the pragmatic relevance condition, and its relation to the first component.

According to Putnam, there is more to being rational than merely being epistemically warranted or justified in holding one's belief. One must also have warranted beliefs which answer questions or address issues that are pragmatically *relevant*, that is, which answer or address a *human* desire or need.

...being rational involves having criteria of *relevance* as well as criteria of rational acceptability...all of our values are involved in our criteria of relevance. The decision that a picture of the world is true (or true by our present lights, or 'as true as anything is') and *answers the relevant questions* (as well as we are able to answer them) rests on and reveals our

¹⁷ As in Chapter Two, these conditions are not intended to be exhaustive. They are simply intended to give the reader the general flavor of a rational acceptability theory of truth.

total system of value commitments. A being with no values would have no facts either.¹⁸

Putnam is perhaps most clear about the role of what I am calling the pragmatic relevance condition on rationality in his internal realism when discussing the putative fact/value dichotomy. In Chapter Six, Putnam asks us to consider the hypothetical case of the Australian super-Benthamites.¹⁹ Paraphrasing just a bit, the story is this: Australian moral theorists, in conjunction with Australian neurophysicists, have perfected a way of measuring “hedonic tone.” The new system of measurement involves a super-computer which uses a revolutionary new program to predict with amazing accuracy even the fairly long-term consequences of an individual action. In response to this new technology, the government of Australia adopts and enforces a strict Benthamite, act-utilitarian theory as its official moral and legal system. As a result of their adherence to this strict act utilitarian moral theory, whenever the super-computer predicts that the torture of a child or the punishment of an innocent would increase the hedonic tone, the Australians quickly act to torture the child or punish the innocent. Putnam supposes that most of us would find the super-Benthamites to be in the grips of what he calls a “sick system of values.” *We* would most likely reject such a moral theory as failing to capture our moral intuition that individual persons have rights even if the super-Benthamite theory were otherwise rationally acceptable.

Even if none of the statements they make about the situation are *false*, their description will not be one that we will count as adequate and perspicuous;

¹⁸ *Reason, Truth, and History*, p. 201.

¹⁹ See especially, pp. 139-141.

and the description we give will not be one that they would count as adequate and perspicuous. In short, even if we put aside our 'disagreement about the values', we could not regard their total representation of the human world as fully rationally acceptable.²⁰

There is a very real sense in which the super-Benthamite theory just doesn't fit *our* experience of the world. This will not, according to Putnam, be a case where we and the super-Benthamites agree on all of the facts about the world but disagree about values. In adopting the super-Benthamite theory and altering their moral vocabulary to reflect such a theory, the Australians will have, as Putnam wants to put it, altered the texture of the human world. "In the course of time the super-Benthamites and we will end up living in different human worlds."²¹

It is for this reason that he believes that the two components of his view are interconnected, that the rational acceptability of a theory in the sense of epistemic warrant is deeply connected to the notion of pragmatic relevance.

The fact is that, if we consider the ideal of rational acceptability which is revealed by looking at what theories scientists and ordinary people consider rational to accept, then we see that what we are trying to do in science is to construct a representation of the world which has the characteristics of being instrumentally efficacious, coherent, comprehensive, and functionally simple? But why? I would answer that the reason we want this sort of representation...is that having this sort of representation system is *part of our idea of human cognitive flourishing*, and hence part of our idea of total human flourishing, of Eudaemonia.²²

²⁰ Ibid., p. 141.

²¹ Ibid.

²² Ibid., p. 134.

That is, according to Putnam, the theoretic virtues tie in with our concept of the good life. Even at the ideal limit of rational inquiry, there will be an infinite number of theories which will be consistent with any set of empirical evidence. For example, consider the evidence which now seems to empirically verify the theory that most of the living things around us arose by the operation of natural selection and other natural evolutionary processes. We can imagine that at the ideal limit of inquiry the evidence would be overwhelming enough to convince even diehard skeptics. However, other theories would seem to be equally consistent with the evidence—say, the theory that life evolved through the process of natural selection and occasional guidance by a divine being, or through natural selection and the guidance of two divine beings who took turns interfering with the otherwise natural process, or.... For this reason, the rational acceptability of an empirical theory will always be, in part, a matter of its possessing more of various theoretic virtues than any competing theories. According to Putnam, which virtues are relevant will depend upon the purposes to which the theory is to be put and upon various facts about our natures and our concept of human flourishing.

What makes a statement, or a whole system of statements—a theory or conceptual scheme—rationally acceptable is, in large part, its coherence and fit; coherence of ‘theoretical’ or less experiential beliefs with one another and with more experiential beliefs, and also coherence of experiential beliefs with theoretical beliefs. Our conceptions of coherence and acceptability are, on the view I shall develop, deeply interwoven with our psychology. They depend upon our biology and our culture; they are by no means ‘value free’. But they *are* our conceptions, and they are conceptions of something real. They define a kind of objectivity,

objectivity for us, even if it is not the metaphysical objectivity of the God's Eye View.²³

Let us now examine in a bit more detail, the first component of Putnam's internal realism, the acceptance of an idealized rational acceptability theory of truth. We have said that such a theory of truth amounts to the view that truth is coextensive with what it would be rational to accept at the ideal limit of inquiry, and we have said something about what would make a statement rational to accept at the ideal limit of inquiry. However, Putnam is notoriously vague and somewhat inconsistent in his spelling out of the details of this part of his view. Mark Johnston, in his excellent article, "Objectivity Refigured: Pragmatism without Verificationism," develops and critiques several ways of making more explicit the central claim of an idealized rational acceptability theory of truth. Johnston, an advocate of something very like the pragmatic relevance component of Putnam's view, concludes that an idealized rational acceptability theory of truth (which he regards as a version of verificationism) is ultimately untenable. In what immediately follows, I will present some of Johnston's formulations and criticisms of the central thesis of an idealized rational acceptability theory with an eye toward developing the most plausible version of that theory.²⁴ I will reserve a discussion of Johnston's main arguments against the view that emerges for the next section.

²³ Ibid., p. 55.

²⁴ Hereafter, I will use 'IRT' as an abbreviation for 'idealized rational acceptability theory of truth'.

Perhaps the most obvious way to make the definition of IRT more precise is by saying that at the ideal limit of rational inquiry there will be exactly one ideal theory T and that a sentence S is true if and only if it follows from the ideal theory T. Johnston quickly points out that such a definition of IRT is, as a simple matter of metalogic, unacceptable.²⁵ If the ideal theory is to be able to contain many truths of physics, it will have to be consistent and strong enough to express arithmetic. Johnston also gives reasons for thinking that the ideal theory will have to be finitely axiomatizable.²⁶ In that case, it follows from the Tarski result—which states that no consistent, axiomatizable, extension of arithmetic can contain its own truth-predicate—that T cannot contain its own truth-predicate. So, the truth definition provided cannot follow from the ideal theory T. Hence, by its own lights, IRT is false.²⁷ Johnston takes this to be a reductio of the very notion of an ideal theory, a theory that contains all and only truths, and suggests that the best that we can achieve is “an indefinitely extensible hierarchy of theories all of which are *empirically* ideal, a hierarchy such that each member of the hierarchy is better than its predecessor in containing more mathematical and philosophical truths.”²⁸ This is almost

²⁵ “Objectivity Refigured,” pp. 88-9.

²⁶ “If the axioms of T are not finite then there will be syntactic truths of the form ‘S does not follow from the axioms of T’ which we cannot in principle come to recognize. For if the axioms of T are infinite in number then although we can in principle enumerate any finite subset and, depending on the underlying logic, tell that S does not follow from that subset, we can never know whether the finite subset which yields S is just around the corner.... The alienated conception of truth threatens to break out in the Internal Realist’s own backyard”(ibid., p. 89).

²⁷ Compare van Inwagen’s criticisms of what he calls ‘Anti-realism’, discussed above at pp. 159-163.

²⁸ “Objectivity Refigured,” p. 89.

certainly correct. However, this need be of very little concern to Putnam who admits that the ideal theory is unachievable.²⁹ Putnam could, I suspect, happily take Johnston's suggestion that any acceptable version of the IRT must make reference to an improving hierarchy of theories. Putnam's internal realist may still be able to analyze truth in terms of an asymptotic approach to an ideal theory even if the limit thus approached is itself incoherent for the reasons given by Johnston.

Johnston next suggests that due to the role that theoretic virtues play in the rational acceptability of a scientific theory, even the notion of an ideal *empirical* theory, is a chimera. First of all, possession of all of the various theoretic virtues may not be possible by any single theory. Some of the theoretic virtues appear to be incompatible. "[I]mprovements in one respect are essentially to be purchased at the cost of neglecting others theoretic virtues. Some think that baroque architecture is elegant. It achieves this elegance by being far from simple."³⁰ Even if possession of all of the various theoretical virtues by any single empirical theory is not in some sense impossible, even if there is, as Johnston says, some unity of theoretic virtue, it is still possible that for any theory which possesses some degree of the unified theoretic virtue, there is some theory which possesses an even higher degree of theoretic virtue. So, we have some reason to suspect that the best that we can do even with regard to empirical theories is to have a hierarchy of

²⁹ *Reason, Truth, and History*, p. 55. Although, to be sure, Putnam does not think that the ideal theory is unachievable because incoherent.

³⁰ "Objectivity Refigured," p. 90.

improving theories. And, if there is no unity of theoretic virtue, then there will be more than one such hierarchy of improving theories.

Again, this criticism does not cut against Putnam's view. Putnam would seem to have no difficulty with the notion that even an empirically ideal theory is an unachievable ideal given the role that theoretic virtues play in rational acceptability and given the dependence of the *relevant* virtues in a specific context on facts about our interests, goals, and psychology.

Will methodological maxims ('choose the simpler theory') help? Once again, it is not clear that they can....[O]nce again we lack principles for determining a unique preferred trade-off.³¹

What I have been saying is that the procedures by which we decide on the acceptability of a scientific theory have to do with whether or not the scientific theory as a whole exhibits certain 'virtues'....I am assuming that verification in science is a holistic matter, that it is whole theoretical systems that meet the test of experience 'as a corporate body', and that the judgment of how well a whole system of sentences meets the test of experience is ultimately somewhat of an intuitive matter which could not be formalized short of formalizing total human psychology.³²

We use our criteria of rational acceptability to build up a theoretical picture of the 'empirical world' and then as that picture develops we revise our very criteria of rational acceptability in the light of that picture and so on and so on forever.³³

³¹ *Reason, Truth, and History*, pp. 93-4.

³² *Ibid.*, p. 133.

³³ *Ibid.*, p. 134. Some may worry that the idea that we continually revise our criteria of rational acceptability entails that we never know what it is for a theory to be true on Putnam's view. However, this will not be the case if we interpret Putnam as holding the view that a sentence is true if and only if it is rationally acceptable at the ideal limit of rational inquiry where we are operating with the notion of rational acceptability that we would be operating with at that ideal limit of rational inquiry. While we may not know now what notion of rational acceptability we will be operating with at the ideal limit of

To be sure, Johnston is aware of the fact that his criticisms of the IRT discussed so far (as well as the worry to be discussed in the following paragraph) may not cut against Putnam's internal realism.³⁴ But, he suggests that once the commitments of the proponent of an IRT are laid bare, his final criticisms, which I will discuss in the next section, clearly do cut against Putnam's view.

According to Johnston, in order for an IRT to escape the criticisms discussed above, it must define truth in terms of a sentence's holding up in *all* sequences of empirically improving theories.³⁵ Johnston's worry now is that, if truth is defined in terms of a sentence's holding up in all sequences of empirically improving theories *formulatable by us*, bizarre truth-value gaps will appear.

Consider a sentence about an observable state of affairs, for example about the number of cakes on a particular tray at a specific time during a party held years ago. Since, as we may suppose, no one in fact counted the cakes or observed anything which would suffice to determine the number

inquiry, this is certainly no more of a problem for Putnam's internal realist than are the skeptical arguments raised in Chapter Two for the metaphysical realist.

³⁴ On p. 88 of "Objectivity Refigured," Johnston writes, "The objections [to internal realism] can be organized around four motifs. Firstly, there cannot be *an* ideal theory but at most an ideal hierarchy of theories each getting more complete with respect to non-empirical truth. Secondly, if there is one such ideal hierarchy then there are many, with consequent embarrassing indeterminacies for the Internal Realists. Thirdly, if the Internal Realist tries to avoid such indeterminacies by strengthening the evidential base of an ideal theory he gives the game away. Fourthly, even if there were an equivalence between eventually holding up in an ideal sequence of theories and being true, then this equivalence would not hold a priori but as an empirical grace of fortune. Perhaps only the last two clusters of objections [versions of which will be discussed below] cut against Putnam's basic claim that truth is an idealization of rational acceptability."

³⁵ In places, Putnam does seem to accept just such a view. "[T]ruth is expected to be stable or 'convergent'; if both a statement and its negation could be 'justified', even if conditions were as ideal as one could hope to make them, there is no sense in thinking of the statement as *having* a truth-value" (*Reason, Truth, and History*, p. 56).

by inference, equally good theories taking into account all the observations which are ever in fact made could differ over this or each leave this undecided. And no improvements of good theories will have enough observations to settle it either way. But, madness aside, no one in fact would deny that the claim that there were more than ten cakes on the tray is...either true or false.³⁶

Just as we are not currently warranted in believing anything concerning the number of cakes at the party, it does not seem that we would be warranted in believing anything on this matter even at the ideal limit of inquiry. For, even at the ideal limit of inquiry, we will be missing some of the relevant observations. Johnston suggests that “the Internal Realist needs to include along with the sentences actually accepted on the basis of observation, the sentences about observable but unobserved states of affairs which we *would* have accepted had we made the crucial observations we did not in fact make.”³⁷ Further, we must have a broad enough understanding of observations to include “observations” made using various instruments and even using various perceptual mechanisms which we do not in fact possess but of which we could conceive. To this end, Johnston proposes the following definition of IRT:

A sentence S is true if and only if S holds up in all sequences of empirically improving theories based on the sentences accepted by a ubiquitous detector³⁸

³⁶ “Objectivity Refigured,” p. 91.

³⁷ Ibid., p. 91.

³⁸ Ibid., p. 92. It seems to me that this definition of IRT as explicitly formulated by Johnston, rules out the possibility that there are any *a priori* truths. However, I will ignore this worry as it can be fixed with a simple amendment to his proposed definition.

where “[s]omething counts as a ubiquitous detector just in case it delivers a verdict on each eternal sentence, proposition or belief about a detectable state of affairs.”³⁹ The problem here,⁴⁰ says Johnston, is that the notion of truth is required in order to specify the notion of a ubiquitous detector.

The idea of a sentence’s being about observable states of affairs seems to be the idea that if the sentence were true then the causal consequences of its being true would include the sort of causal consequences which could trigger observation. The idea of a sentence’s being about detectable states of affairs seems to be the idea that if the sentence were true the causal consequences of its being true would include the sort of causal consequences which could trigger a detection, that is could be registered on an instrument which extended our observational capacities. *Mutatis mutandis* for propositions or beliefs about observable and detectable states of affairs respectively.⁴¹

So, it would seem that any definition of truth in terms of the sentences accepted by a ubiquitous detector will be circular.

However, circularity should not be a problem for Putnam. Although it is clear that Putnam wants to include among the set of sentences which it would be ideally rational to accept various claims about the past concerning which no observations were actually made and also sentences concerning states of affairs which, strictly speaking, are unobservable by us, Putnam would not accept Johnston’s proposed truth-definition. Putnam explicitly states that he is not attempting to provide a formal *definitional* analysis of truth.⁴² What

³⁹ Ibid.

⁴⁰ Ibid., pp. 92-3.

⁴¹ Ibid., p. 93.

⁴² “...I am not trying to give a formal *definition* of truth, but an informal elucidation of the notion” (*Reason, Truth, and History*, p. 56). See also his criticism of the logical positivist’s attempts to define truth, pp. 105-113.

is crucial for Putnam is that truth and rational acceptability should not be (logically) separable. In other words, his claim is that truth is, by its very nature, coextensive with idealized rational acceptability, that there is an *a priori* epistemic constraint on truth.

Recognizing the possibility of this sort of move on behalf of a proponent of the IRT, Johnston focuses his attentions on the claim that the following bi-conditional (let us call it “ET” since it puts an epistemic constraint on truth), while not a reductive definition of truth, may nonetheless constitute an *a priori* epistemic constraint on truth:

ET: A sentence S is true if and only if S holds up in all sequences of empirically improving theories based on the sentences accepted by a ubiquitous detector.

Against such a version of the IRT, Johnston constructs two arguments: the so-called missing explanation argument and what I will call the enigma argument, both designed to show that ET (or a natural restriction thereof) is not *a priori*. So, it is to an examination of these two arguments that I now turn.

2. I will tackle Johnston’s enigma argument first since it is most easily dealt with in this context. In presenting the enigma argument⁴³ that ET is not *a priori*, Johnston first asks us to focus on the following restriction of ET (which I will call “ST” for *strongly theoretic truth*):

⁴³ See “Objectivity Refigured,” pp. 96-7.

ST: S is a true strongly theoretical sentence iff S is a strongly theoretical sentence and S holds up in all sequences of improving theories based on the sentences accepted by a ubiquitous detector.⁴⁴

ST requires some explanation. According to Johnston, a sentence is a strongly theoretical sentence if and only if it is a sentence that is not about a detectable state of affairs⁴⁵ and not a sentence of logic or mathematics.⁴⁶ Since the set of strongly theoretical sentences will obviously be a subset of all the sentences that there are, ST is a logical consequence of ET. So, if Johnston can show that ST is not *a priori*, it will follow that ET is not *a priori*.

Johnston next asks us to consider the following sentence:

S: There are enigmas, i.e., there are entities essentially undetectable by us.

The enigma argument can now be restated as follows:⁴⁷

1. S is a strongly theoretical sentence—and so, of course, either true or false.⁴⁸

⁴⁴ Ibid., p. 93.

⁴⁵ Johnston claims that a sentence is about a detectable state of affairs if it “involves the obtaining of a state of affairs with causal consequences which could be detected by any being who had all the (metaphysically) possible idealizations of our powers of detection”(ibid., p. 97).

⁴⁶ Ibid., p. 93.

⁴⁷ See ibid., pp. 96-97. Thanks are due to Robert Coburn who helped me recognize the structure of Johnston’s argument.

⁴⁸ Johnston’s argument clearly depends on the claim that strongly theoretical sentences are either true or false. But why should a proponent of ET accept that claim? To be sure, a proponent of ET may want to hold that any sentence which does not hold up in all sequences of empirically improving theories based on the sentences accepted by a ubiquitous detector are false. If he does, then he will have no quarrel with Johnston’s claim that a strongly theoretical sentence must be either true or false. However, a proponent of ET may want to hold that in order for a sentence S to be false, its negation must hold up in all sequences of empirically improving theories, etc., etc. Since it is possible that neither S nor its negation would hold up in all sequences of improving theories, the proponent of ET need not accept the claim that S must be either true or false.

2. S is a sentence for whose truth no (adequate) empirical evidence is possible.
3. It cannot be settled *a priori* whether or not S is true.
4. Thus, there are strongly theoretical sentences that are true or false and yet whose truth or falsity is beyond our epistemic reach.
5. If ST is *a priori*, there will not be strongly theoretical sentences that are true or false and yet whose truth or falsity is beyond our epistemic reach.
6. Therefore, ST is not an *a priori* truth (and hence, neither is ET).

The unhappy conclusion for the proponent of ET is a logical consequence of steps 4 and

5. It is difficult to see how an internal realist who accepts ET (and so, ST) could reject step 5 of the argument above and step 4 does seem to follow from the first three premises.

Since premise 2 of the enigma argument is analytically true, it would seem that if the internal realist or other proponent of ET is going to object to the enigma argument, he must find reason to reject either premise 1 or premise 3.

In support of premise 1 of the argument, the claim that S is a strongly theoretical sentence, Johnston argues that sentence S (i) has content, (ii) is not about a detectable state of affairs, and (iii) is not a sentence of logic or mathematics. From this, along with Johnston's analysis of the notion of a strongly theoretical sentence mentioned above, it follows that premise 1 is true. Johnston anticipates that the internal realist or other proponent of ET will be tempted to reject this argument for premise 1 by arguing that S is devoid of content since it cannot be used in a warranted assertion. He replies to the anticipated argument by suggesting that it turns on "too strong a premise."

There are very long sentences which we would never in fact employ in warranted assertions just because of their length. These sentences are

contentful because they can be built up compositionally from a finite list of predicates, names, functors, etc. which do find use in warranted assertions. So, if there is a plausible constraint on contents then it had better allow that sentences can have content in virtue of being built up from predicates, names, functors, etc. which figure in sentences which we can recognize to be made probable or warranted by certain kinds of evidence. [Sentence S] is such a sentence. Accordingly there is no good reason to deny it a content.⁴⁹

While the internal realist might be able to find and exploit a weakness in Johnston's reasoning in support of the claim that S has content,⁵⁰ I will not pursue this line of response here. It seems to me that Johnston's argument for premise 1 of the enigma argument is at least *prima facie* fairly cogent. And, in any case, it seems to me that there is an even greater weakness in Johnston's enigma argument that can be exploited by the proponent of ET.

⁴⁹ "Objectivity Refigured," p. 96.

⁵⁰ It seems to me that the proponent of ET might be able to object to Johnston's move from the claim that it is absurd to deny that some very long sentences which could never be used by us in a warranted assertion have content to the claim that any sentence which can be built up compositionally from linguistic elements which have content itself must have content. It seems at least possible that one could maintain that there is a warranted assertability constraint on the content of entire sentences even in the face of Johnston's alleged counter-example by drawing a distinction between those sentences which would never actually be used by us in a warranted assertion (such as the very long sentences Johnston mentions) and sentences (like S) which could never be used in a warranted assertion even by a being that had, in addition to all of the (metaphysically) possible idealizations of our powers of detection, an unlimited amount of time in which to make assertions. While it is not obvious that this line of defense can be made to work, it is not entirely obvious that it cannot. A more detailed examination of the semantic issues involved in settling this issue is a task well beyond the scope of this chapter.

Why should a proponent of ET accept premise 3 of Johnston's argument—why should he accept the claim that S's truth or falsity is not knowable *a priori*? Johnston offers very little by way of support for this claim.

Notice that it is not analytically false to suppose that there are Enigmas. It cannot be settled *a priori* whether there are or are not Enigmas.... What might Enigmas be like? No one of course knows. But it is possible to fill out logically coherent scenarios which put some flesh on the formal observation that there is nothing logically or analytically incoherent about the supposition that there are Enigmas. One (bleak) scenario is this. Materialism in some form is true of us. Although there are immaterial things we never interact with anything immaterial thanks to a causal caste system in which, as a metaphysically necessary matter, the immaterial interacts only with itself. Since the immaterial realm gives off no signals that might affect us or any instrument we could develop, no idealization which abstracts away from our contingent limitations on our powers of observation and detection would have us detecting the immaterial realm. The whole immaterial realm then would be enigmatic in the intended sense.⁵¹

Johnston's story about what enigmas might be like goes a long ways toward motivating his claim that S is not analytically false, that the idea that there are enigmas is not logically or analytically incoherent. But Johnston must do more than show that S is not analytically or logically false. He must show that S cannot be known to be false (or be known to be true for that matter) *a priori*—and that he has given us no reason to think.

Some may claim that to show that S is not *a priori* it is sufficient to show that S is neither analytic nor a sentence of logic or mathematics. However, it is far from clear that the set of analytic sentences conjoined with the set of sentences of logic and mathematics does exhaust the set of sentences whose truth or falsity is knowable *a priori*. Many will

⁵¹ "Objectivity Refigured," p. 97.

want to hold, along with Kant, that, e.g., certain ethical sentences, say, ‘it is wrong to torture small babies merely for entertainment purposes’, are knowable *a priori* even though they are neither analytic nor sentences of logic or mathematics. And, in any case, it is clear that Johnston wants to accept as *a priori* certain sentences that are neither analytic nor sentences of logic or mathematics.⁵²

While there is some intuitive pull to the idea that S is not knowable *a priori*, this is by no means obvious. If Johnston’s enigmas argument is to give us good reason to think that Putnam’s internal realism or any other view which embraces ET is ultimately untenable, it seems to me that more needs to be done to support premise 3 of that argument. It is, after all, open to the proponent of ET to run Johnston’s argument backward, claiming that since one consequence of his view is that there can be no strongly theoretical sentence whose truth or falsity is beyond our epistemic reach, and since S is a strongly theoretical sentence for the truth of which no (adequate) empirical evidence is possible, either S or its denial must be knowable *a priori*.⁵³ Without further support for Johnston’s claim that S cannot be known *a priori*, it seems to me that there is no obvious reason to prefer running the argument the one way rather than the other.

So, the enigma argument does not, after all, establish the conclusion that ST (and hence ET) is not *a priori*. However, before the internal realist can reject Johnston’s

⁵² See his sentences (20) and (21), *ibid.*, pp. 114–115.

⁵³ Note that the proponent of ET need not claim that either S or its denial is currently known *a priori*, only that one or the other is knowable *a priori*.

conclusion that ET is not knowable *a priori*, he must somehow address Johnston's missing explanation argument. That argument, in main outline, is as follows:⁵⁴

- 1) If a sentence S about a detectable state of affairs holds up in all improving empirical theories based upon the deliverances of the ubiquitous detector, then there will be a causal explanation of why this is the case that is of the form: S holds up in all sequences of improving empirical theories based upon the deliverances of the ubiquitous detector *because S is true*.
- 2) But if ET is an *a priori* truth, there can be no explanation of this form as to why S holds up in all sequences of theories based on the sentences accepted by a ubiquitous detector (whatever S is about).
- 3) Therefore, ET is not *a priori*.

Again, the argument is deductively valid. If the internal realist is to avoid the conclusion, he must find reason to reject one or the other of the premises. Unfortunately for the internal realist, both premises of Johnston's missing explanation argument are quite plausibly true. Premise 1 is plausible in that, if S asserts that a detectable state of affairs holds, then this state of affairs will have a causal impact on the ubiquitous observer. And, if this is the case, then there should be a causal explanation of this fact along the lines suggested by Johnston. Premise 2 is plausible in that, as Johnston points out, there are no causal explanations of the form [A because B] where A and B are equivalent as an *a priori* and necessary matter.⁵⁵ So, it does look as if Johnston's missing explanation constitutes a powerful argument against an IRT such as Putnam's internal realism.⁵⁶

⁵⁴ See "Objectivity Refigured," pp. 93-96. This formulation is again due to Robert Coburn.

⁵⁵ Ibid., p. 95.

⁵⁶ Robert Coburn has suggested that the internal realist may have a response to Johnston's missing explanation argument along the following lines. Johnston claims that the

So, can we now conclude that there is apparently no coherent version of metaphysical antirealism to be had, and so no objection to the metaphysical realist maintaining the default position in the debate despite the problems raised for his view in the chapters above? Or is there some other version of metaphysical antirealism, one that isn't susceptible to the missing explanation argument or an analogue thereof? In his unpublished *Carus Lectures* and in his "Reply to Richards," Paul Grice develops a metaphysical view which I will call 'metaphysical constructivism' according to which the world consists largely (if not entirely) of constructed, mind-dependent entities.⁵⁷ While the picture that Grice sketches does not in and of itself entail metaphysical antirealism,⁵⁸ it seems to me that much of what Grice has to say on this topic is consistent with and congenial to metaphysical antirealism. In the section that follows, I will present a sketch

proponent of ET cannot provide a causal explanation of why a sentence S about a detectable state of affairs, say the sentence 'there is a rabbit in front of me', holds up in all improving empirical theories based upon the deliverances of the ubiquitous detector. However, this claim turns on an unwarranted assumption concerning the form that such a causal explanation must take. Coburn suggests that the internal realist might be able to explain the success of S, not in terms of the truth of S, but rather in terms of the truth of some other set of sentences (say, some sentence A about the history of a rabbit and some sentence B about my history) which, together with the causal laws of the universe, entail S. Such an explanation, if spelled out in more detail, would plausibly escape the charge that its *explanans* and *explanandum* were *a priori* equivalent. If this is the case, then there may be no reason to think that Putnam's internal realism is any less capable of development into a viable alternative to metaphysical realism than is the Gricean metaphysic discussed below.

⁵⁷ Grice nowhere call his view 'metaphysical constructivism'; instead, he claims to be following a "constructivist metaphysical program." However, I think this name for his view is apt.

⁵⁸ Because, as we will see below, Grice does not want to rule out the possibility that there may be unconstructed, primitive entities.

of the various components of Grice's constructivist metaphysical program. While the picture of the mind-dependent world that emerges, as even Grice admits, "contains much that is obscure, fragmentary and ill-defended (where indeed it is defended at all),"⁵⁹ it seems to me that it is sufficient to show that we cannot simply rule out the possibility that a coherent and viable account of the notion of a mind-dependent empirical world can be developed along the lines suggested by Grice. If the metaphysical antirealist is able to make sense of a mind-dependent empirical world and hold that empirical truth consists in correspondence with facts about the empirical world, then, as we will see, no analogue of the missing explanation argument leveled against such a version of metaphysical antirealism will be able to get off the ground.

3. In the third of his Carus Lectures entitled "Metaphysics and Value," Grice describes his metaphysical program as follows:

[T]he metaphysical program which I shall be seeking to follow will be a constructivist program and not a reductionist program. The procedure which I envisage, if carried out in full, would involve beginning with certain elements which would have a claim to be thought of as metaphysically primary, and then to build up from these starting-points, stage by stage, a systematic metaphysical theory or concatenation of theories. It would be *no* part of my plan to contend that what we end up with "is *really* only such-and-suches".... I would not seek to exhibit anything as 'boiling down' to any complex of fundamental atoms.

In order to pursue a constructivist program of the kind which I have in mind, I will need three things: first, a set of metaphysical starting points, things which are metaphysically primary; second, a set of recognized construction routines or procedures, by means of which non-primary items

⁵⁹ "Carus Lecture III: Metaphysics and Value," pp. 73-4.

are built up on the basis of more primary items; and third, a theoretical motivation for proceeding from any given stage to a further stage, so that the mere possibility of applying the routines would not be itself enough to give one a new metaphysical layer; one would have to have a justification for making that move; it would have to serve some purpose.⁶⁰

In this section I will discuss in more detail what Grice perceives to be the three stages of his constructivist program: the identification of metaphysical primitives, the procedures of metaphysical construction, and the motivations for proceeding from one stage to another. After the discussion in this section, the reader should have a fairly clear idea of the basic structure of Grice's constructivist program. In the next section, I will go on to explain how a coherent and more fully worked out execution of this metaphysical program, should one be forthcoming, would not be undone by Johnston's missing explanation argument which seems to undermine Putnam's internal realism. I will also raise and discuss several worries concerning the program itself and various ways of amplifying the details of the program into a coherent view. (Many of these worries are raised by Grice himself in his "Reply to Richards." Others were raised in his lectures at the University of Washington in 1980 and 1983.) Finally, I will argue that, despite the problems raised for Grice's metaphysical constructivism, it is at least not obvious that they cannot be surmounted. That is, it is by no means obvious that the program cannot be coherently carried out. Without some further defense of his view—perhaps an argument that any way of filling out the details of Grice's program which would be congenial to the metaphysical

⁶⁰ "Metaphysics and Value," pp. 48-9.

antirealist will be incoherent—the metaphysical realist cannot simply lay claim to the default position in the debate.

As Grice sees it, the procedure of metaphysical construction begins with the identification of metaphysically primitive entities or perhaps properties. In order to point in the direction of how we would come to identify the metaphysically primary entities, Grice introduces the notion of theory-theory. Theory-theory amounts to a characterization of what a theory is, what purposes theories serve, and what virtues theories have. The idea is that, as a result of our theory-theorizing, one kind of theory (or one theory) might come to be identified as more basic or primary and the entities and properties postulated by this (type of) theory (if any there be) would then be the metaphysically primitive entities and properties. Although he has very little to say on the subject, Grice hints at two directions in which our theory-theorizing might lead.

It might, for example, turn out to be establishable that every theory has to relate to a certain range of subject-items, has to attribute to them certain predicate attributes, which in turn have to fall within one or another of [a] range of types or categories. In this way the inquiry might lead to recognized metaphysical topics, such as the nature of being, its range of application, the range of predication, and a systematic account of categories.⁶¹

A second approach would focus not on the idea of the expressibility of the outcomes of rational enquiry in theories, but question what it is, in such inquiries that we are looking for, why they are of concern to us....what as theorists we strive for is not (primarily) further facts, but *rational knowledge* or *understanding*, of the facts we have, together with whatever further facts our investigations may provide for us. Metaphysics will have as its concern the nature and realizability of those items which are involved in any successful pursuit of understanding; its range would include the

⁶¹ “Reply to Richards,” p. 88.

nature and varieties of explanations (as offered in some modification of the Doctrine of Four Causes), the acceptability of principles of logic, the proper standards of proof, and so on.⁶²

Although he initially presents these as different approaches to the initial stage of his metaphysical program, he allows that the two approaches may not be genuinely distinct. “[I]t seems plausible to suppose that explanations, if fully rational, must be systematic and so must be expressible in theories. Conversely, it seems plausible to suppose that the function of theories is to explain, and so whatever is susceptible to theoretical treatment is thereby *explained*.”⁶³

Despite the fact that Grice’s metaphysical program is itself a constructivist program, it is quite possible that the model of explanation at which we might arrive as a result of our theory-theorizing would be, in a sense, a reductionist one.⁶⁴ That is, we might come to view every macro-event as, at least in principle, susceptible to a “bottom up” explanation in terms of the more elementary ingredients that make it up. On such a model of explanation, the metaphysical primitives might be identified with the entities or properties postulated by our most complete physical theory. Having identified the primary entities or properties, we would then proceed in a series of theoretic stages which involve the construction of a series of more complex entities and properties. “Each newly introduced entity-type would carry with it a segment of theory which would supplement

⁶² Ibid.

⁶³ Ibid.

⁶⁴ But of course it needn’t be. I use this example because it is a model of explanation that will be familiar to most readers.

the body of theory already arrived at, and which would serve to exhibit the central character of the type or types of entity associated with it.”⁶⁵ Perhaps if we identify the primary entities as, among others, say, electrons and quarks, we would proceed by constructing a series of theories postulating first atoms, then molecules, then amino acids, proteins, DNA, living organisms, etc. While the series of constructed entities would be made of (and, in that sense, reducible to) the primary entities, this would nonetheless be compatible with there being special laws that govern the constructed entities postulated in the later stages of theory.

However, it is important to keep in mind that the series of entities and properties at which we arrive through this process of metaphysical construction depends crucially on what we come to identify as the result of our theory-theorizing as the First Theory.⁶⁶ Grice himself imagines, as another example of what might occur, a more strictly *metaphysical* series of theories than the one I sketched above, perhaps postulating at one level (the first?) non-enduring particulars, at the next level continuants, followed by “a specially privileged kind of continuants, namely substances and so on.”⁶⁷ But, regardless of what we have identified as the relevant metaphysical primitives, the procedure of

⁶⁵ “Metaphysics and Value,” p. 50.

⁶⁶ It is important to note that if there is no single unified purpose or goal to theorizing or no single unified notion of explanation, there may well be no single theory identifiable as *the* first theory, but rather, there will only be theories identifiable as the first theory relative to this or that theoretic purpose.

⁶⁷ “Metaphysics and Value,” p. 50.

constructing new entities on top of the metaphysically primary ones will involve the use of one or more “routines of construction.”

It is pretty obvious that metaphysical construction needs to be disciplined; but this is not because without discipline it will be badly done, but because without discipline it will not be done at all. The list of available routines determines what metaphysical construction *is*; so, it is no accident that it employs these routines.⁶⁸

Grice identifies three routines of metaphysical construction: Humean Projection, Category Shift, and Metaphysical Transubstantiation. However, he nowhere indicates that he believes this list of routines to be exhaustive.⁶⁹ A discussion of these three routines will not only give us a clearer picture of what metaphysical constructivism amounts to, it will also give us some indication of the directions in which to look for further routines of metaphysical construction.

The construction routine of Humean Projection was first discussed by Grice in his “Methods in Philosophical Psychology.”⁷⁰ “[T]his operation consists in taking something which starts life, so to speak, as a specific mode of thinking, and then transforming it into an attribute which is ascribed not to thinking but to the thing thought about and indeed is, in a given case, attributed either correctly or incorrectly.”⁷¹ According to Grice, there are four either real or apparent stages of the routine of Humean Projection. In the first stage

⁶⁸ “Reply to Richards,” pp. 96-97.

⁶⁹ Indeed, Robert Coburn recalls that he mentioned several other construction routines in his University of Washington lectures of 1980 and 1983.

⁷⁰ *Proceedings and Addresses of the American Philosophical Association* 68 (November, 1975), pp. 23-53.

⁷¹ “Metaphysics and Value,” p. 70.

(which Grice says may not always be present) we begin with a critical idea or concept which is an “intuitive and unclarified element of our conceptual vocabulary.”⁷² Grice uses as an example of such an intuitive and unclarified concept the concept expressed by the word ‘or’. The second stage involves reaching a specific mental state “in the specification of which it is possible though maybe not necessary to use the name of the initial concept as an adverbial modifier; we come to ‘or-thinking’ (or disjoining)....”⁷³ In stage three of Humean Projection, we replace reference to the specific mental states with general psychological verbs such as ‘thinking’ together with an operator corresponding to the specific state and meeting certain restrictions.⁷⁴ So at this stage we arrive at reference to, e.g., thinking that p or q. In the final stage, restrictions imposed on the operators are removed and we get, e.g., our idea of thinking that if p or q, then r.

Grice makes five observations about the routine of Humean Projection: (i) this routine delivers concepts (or properties or relations) rather than objects; (ii) the provision in stage four of “full syntactico-semantic freedom for the operators which correspond to the initial concepts is possible only via the provision of truth-conditions, or of some different but analogous valuations, for statements within which the operators appear”;⁷⁵ (iii) because the need to provide truth-conditions (or something analogous) is required at

⁷² “Reply to Richards,” p. 97.

⁷³ Ibid.

⁷⁴ The operator should “[appear] within the scope of the general [psychological] verb, but is itself allowed only maximal scope within the complement of the verb, and cannot appear in sub-clauses” (ibid.).

⁷⁵ Ibid., p. 98.

stage four, the difference between the second and third stages is apparent rather than real;⁷⁶ (iv) the development of stages is not arbitrary and must serve some purpose;⁷⁷ (v) “[s]ubject to these provisos, application of this routine to an initial concept...does furnish one with a metaphysical reconstruction of that concept; or if the first stage is missing, we are given a metaphysical construction of a new concept.”⁷⁸

The next routine of metaphysical construction is what Grice calls ‘Category Shift’, although he allows that it might have been called ‘subjectification’ or perhaps even ‘objectification’,⁷⁹ since its role is to introduce new objects or new subjects of discourse. Although Grice is not as explicit concerning the distinct stages of this routine, we can perhaps spell out the details of Category Shift using the stages of Humean Projection as a model. In the first stage of Category shift we would have substances, various things predicable of substances, and various “slots” or categories into which predicates of substances may fit. Grice thinks that these “slots” will resemble Aristotle’s categories of substance, quality, quantity, etc.⁸⁰ “[O]ne [slot] is substance itself (secondary substance), in which case the predicate is intracategorial and *essential*; and there are others into which

⁷⁶ “The third stage provides only a notational variant of the second stage, at least unless stage four is also reached”(ibid.).

⁷⁷ This is connected to the third component of metaphysical construction, a motivation for proceeding from one level to another, and will be discussed further below.

⁷⁸ “Reply to Richards,” p. 98.

⁷⁹ Ibid., pp. 98-99.

⁸⁰ However, Grice allows that the members of the list might not be fully co-ordinate. “[T]he development of the list might require not one blow, but a succession of blows; we might for example have to develop first the category of attribute, and then the subordinate categories of quantitative attribute (quantity) and non-quantitative attribute (quality), or again the category of event before the subordinate category action”(ibid., p. 99).

the predicates assigned in non-essential or accidental predication may fall....”⁸¹ In the second stage, an item or perhaps a complex of a set of items which occupied one or more of the non-substantive categories at stage one comes to occupy the first, substantial slot. That is, in this stage, what was formerly a predicable comes to be the subject of predication. For example, suppose that our substances include persons and that among the various things predicable of persons is a capacity for laughter. Then risibility may be treated as a subject of predication. We may say, for example, that risibility is a desirable trait, or one that only members of the species *Homo Sapiens* truly possess. These new subjects of predication, as Grice says, would not be substances proper and their predicables would not be qualities or quantities, etc. *simpliciter*.⁸² However, when properly motivated, a use of Category Shift produces a substantial which has some analogue of secondary substance predicable of it. Grice again emphasizes the fact that, as with Humean Projection, a use of the routine Category Shift must be motivated if it is to be a case of entity *construction* rather than a mere *façon de parler*.

The third routine of metaphysical construction described by Grice is one which he calls ‘Metaphysical Transubstantiation’. According to Grice, Metaphysical Transubstantiation involves the idea that there can be two continuants S_1 and S_2 such that S_1 exists at time T_1 and time T_2 and S_2 exists at T_1 (at which time $S_1 = S_2$), but does not exist at T_2 .

⁸¹ Ibid., p. 98.

⁸² Ibid., p. 99.

Execution of the maneuver of metaphysical transsubstantiation [sic] consists in taking a certain sort of substance S_1 , to which a certain property or set of properties P would be essential, and then introducing a type of substance S_2 an instance of which may possess property or properties P , but if it does, does not possess them *essentially*. What will be essential to S_2 will be some other set of properties P' , properties which might attach, though not *essentially*, to some, or even to all instances of S_1 .⁸³

Here, it is perhaps easiest to think of the traditional metaphysical example of the hunk of clay and the statue. The statue has, as an essential property, a certain shape, say, that of Michaelangelo's *David*. It cannot lose this general shape and continue to exist. The hunk of clay also has the property of having the shape of Michaelangelo's *David*. But for the hunk of clay, this is a merely contingent property; it would have existed even if it had never been formed into the shape of Michaelangelo's *David* and will continue to exist even when pounded into a lump and reshaped.

According to Grice, a use of Metaphysical Transsubstantiation does not alter the totality of attributes which each entity possesses. Rather, it affects which properties attach to which types of entities essentially and thereby allows us to distinguish new entity-types and to formulate further theory concerning the new entity-type.⁸⁴ This feature of Metaphysical Transsubstantiation can clearly be seen in one example of a use of the routine discussed by Grice. In this example, we start with the previously constructed entity type *Homo Sapiens* instances of which possess, as an essential property, a certain genotype (say), and also possess, as an accidental property, rationality. What then

⁸³ "Metaphysics and Value," p. 63.

⁸⁴ "Reply to Richards," p. 102.

happens is that *Homo Sapiens* uses his rational capacity to (in a sense) reconstruct himself.

The result is the creation of the entity type *persons* instances of which possess, as an essential property, rationality, and, as an accidental property, a certain genotype.⁸⁵

While each human is standardly coincident with a particular person (and is indeed, perhaps, identical with that person over a time), logic is insufficient to guarantee that there will not come a time when that human and that person are no longer identical, when *one* of them, perhaps, but not the other, has ceased to exist. But though logic is insufficient, it may be that other theories will remedy the deficiency.⁸⁶

Once again, it will be essential to a proper use of Metaphysical Transubstantiation that the introduction of the new entity-type serve some theoretical purpose.

Much like the ideal theory, *the* complete set of acceptable routines of metaphysical construction might be a chimera. It could well be that as our total theory evolves we are forced to revise our list of acceptable construction routines and even, in certain cases, to invent new ones. Grice would, I think, be amenable to the suggestion that we look to our actual, evolving practices of theory construction and postulation of objects in order to develop an ever more complete list of acceptable construction routines.

As we have seen, regardless of what type of routine of metaphysical construction is to be used, for that use to be legitimate, that is, for the use to really involve the *construction* of a new concept, property, or entity-type, there must be a theoretical or

⁸⁵ Notice that Grice is perfectly willing to consider giving up Leibniz's Law (at least in its unrestricted and unqualified form) concerning such cases. Cf. George Myro, "Identity and Time," in *Philosophical Grounds of Rationality* (Oxford: Claredon Press, 1986), pp. 383-409, where Myro spells out some of the details of Grice's suggestion.

⁸⁶ "Reply to Richards," p. 102.

explanatory purpose served by such a use. This is the third component of Grice's metaphysical constructivist program. To move through the stages of a construction routine arbitrarily, without proper motivation, is just to invent a new way of speaking about already existing concepts, properties, or objects, and not to construct new ones. Grice discusses several types of theoretic motivation for the use of a routine of construction: "the possibility of opening up new applications for existing modes of explanation,"⁸⁷ the availability of a transcendental argument that demonstrated that such and such a kind of entity, property, or concept is rationally required, and the existence of what might be called a legitimate psychological need or demand. In what immediately follows, I will discuss in more detail these types of theoretic motivation. It is important to keep in mind that these examples are most likely intended by Grice to illustrate but not to exhaust the range of acceptable theoretical motives.

According to Grice, "[i]n setting out some theory or theory-stage B, which is to succeed and include theory or theory-stage A, it may be that one introduces some theoretical apparatus which provides one with a redescription of a certain part of theory A."⁸⁸ In such a case, we say that there is overlap between theory A and theory B. Grice uses the example of introducing the theory of the positive and negative integers which would, among other things, provide a redescription of (and hence overlap) the earlier theory of the natural numbers. Introducing such a theory is not intended to reduce or

⁸⁷ Ibid., p. 99.

⁸⁸ "Metaphysics and Value," p. 57.

entirely replace the earlier theory (the earlier theory may be much more economical to use when we are dealing only with positive integers). Instead the new theory is intended, perhaps, to enlarge the range of arithmetical laws. "While such an overlap may be needed to secure intelligibility for [the new] theory B, theory B would be pointless unless its expressive power transcended that of theory A...."⁸⁹ The ability of the new theory to enlarge the range of arithmetical laws and operations in the area beyond the overlap is, as Grice puts it,⁹⁰ the "*ratio essendi*" of, or the reason for, the introduction of the new theory of positive and negative integers. The idea then is that one way in which we would be justified in introducing a new theory (thereby constructing the entities or properties postulated and described by the theory) is when the new theory significantly overlaps an old one and opens up new and fruitful applications for existing modes of explanation or operation in the area beyond the overlap with the old theory.

Another motivation for a legitimate employment of a construction routine would be provided by a transcendental argument that demonstrated a rational need for such and such a kind of concept, property, or object. According to Grice, what gives a transcendental argument its transcendental flavor is the suggestion that "something is being *undermined* by one who is skeptical about the conclusion which such an argument aims at establishing."⁹¹ Grice sketches three transcendental argument-patterns.⁹² The

⁸⁹ "Reply to Richards," pp. 91.

⁹⁰ See "Metaphysics and Value," p. 58.

⁹¹ "Reply to Richards," p. 94.

⁹² *Ibid.*, pp. 94-5.

first pattern, which Grice says fits Descartes' *cogito* argument in the *Meditations*, shows that the skeptic's procedure for motivating doubt is self-destructive. The second pattern, which Grice says fits arguments against, e.g., Hume's skepticism about causality and the phenomenalist's skepticism about material-objects, attempts to show that "the rejection of the 'over-and-above' entities (causality in the case of Humean skepticism or material objects in the case of phenomenalism) is epistemically destructive of the entities with which the skeptic deems himself secure; if material objects or causes go, sense-data and dateable events go too."⁹³ The third pattern is exemplified, according to Grice, by an argument against the view that we have no free will. Such an argument in broad outline would go: freedom is a pre-condition for any exercise of rationality. So, if freedom were allowed to fall, so too would rationality.

The last motivation or justification for the use of a routine of metaphysical construction discussed by Grice can be characterized as the presence of a legitimate psychological need or rational demand which is met by the construction of some new theory or the use of some procedure of construction.

[G]iven that such a creature is equipped to formulate a legitimate demand for solutions to [some set of] questions, then any set of procedures which it could devise which are such that there are no *objections* to them, and which if they were accepted as proper procedures could then be used to deliver answers to these questions, will be procedures which it will be reasonable for the creature to accept as *proper*, provided that there are no other equally unobjectionable candidates with equally good prospects of delivering answers to the same questions.⁹⁴

⁹³ Ibid., p. 95.

⁹⁴ Ibid.

[G]ranted that there is a rational demand for[e.g.] absolute value, we can then perhaps argue that within whatever limits are imposed by metaphysical constructions already made, we are free to rig our metaphysics in such a way as to legitimize the conception of absolute value; what it is proper to believe to be true may depend in part on what we would like to be true. Perhaps part of the Kantian notion of positive freedom...is the freedom not merely to play the metaphysical game but, within the limits of rationality, to fix its rules as well.⁹⁵

Grice dubs the principle which gives credence to the adoption of such a set of procedures for satisfying a legitimate psychological need the “Metaphysical Principle of Supply and Demand.”⁹⁶ So, the third motivation for a specific use of a metaphysical construction procedure will involve an argument that in some ways resembles a transcendental argument. However, in this case, what would be undermined by a failure to acknowledge as legitimate a certain construction procedure or a specific use thereof is the satisfaction of some rational demand or legitimate psychological need.⁹⁷

This concludes my discussion of the three components of Grice’s metaphysical constructivism: the identification of metaphysical primitives, the use of acceptable routines of construction, and the motivation for the use of such routines. In the next section, I will show how Grice’s metaphysical program would allow the metaphysical antirealist to provide a positive view which would be able to escape Johnston’s criticisms and which might well constitute a viable alternative to metaphysical realism, at least if it is capable of

⁹⁵ Reply to Richards,” p. 105.

⁹⁶ “Metaphysics and Value,” p. 68.

⁹⁷ Unfortunately, Grice does not have much to say about what would make a psychological need “legitimate” for the purposes of motivating the use of a metaphysical construction routine.

coherent execution. But, this is a big ‘if’. There will be, as we will see, many lingering worries concerning the possibility of filling out Grice’s picture into a coherent metaphysical view. However, in the absence of an argument which shows that the difficulties that emerge are insuperable, there seems to me to be as much reason to explore various ways of executing Grice’s program to see if any can adequately address these worries as there is to pursue possible responses on behalf of the metaphysical realist to the difficulties raised above for his view.

4. Before we go on, I think that a brief summary of the ground we have covered so far is in order. In the first two sections of this chapter, I presented Putnam’s version of metaphysical antirealism, internal realism, and Johnston’s “missing explanation” and “enigmas” arguments against it. We saw that, although Putnam plausibly has a response to the enigmas argument, the missing explanation argument constitutes an apparently damning objection to his view. However, I suggested that a metaphysical antirealist might be able to escape the missing explanation argument provided that he could make sense of the notion of a mind-dependent empirical world. In the previous section, I presented the elements of Grice’s metaphysical constructivist program in an attempt to provide some initial content to the idea of such a mind-dependent empirical world. Although the content provided is, at this stage, purely programmatic and remains to be executed in detail (assuming that it can be coherently executed), we are now in a position to see how such a

version of metaphysical antirealism, when and if fully executed, would be able to avoid the missing explanation argument which proved the undoing of Putnam's view.

Recall that the difficulty for Putnam's idealized rational acceptability theory of truth (or at least the version of that view developed by Johnston) stemmed from the fact that it held the set of true sentences to be coextensive as a necessary and *a priori* matter with the set of sentences that would hold up in all sequences of empirically improving theories based on the sentences accepted by a ubiquitous detector. Unfortunately for Putnam, this appears to rule out the possibility of providing a causal explanation for the fact that a sentence S holds up in all sequences of empirically improving theories based on the sentences accepted by a ubiquitous detector, even though some explanation of that sort would seem to be called for. What we want to say is that it is *because* S is a member of the set of true sentences and the truth of S has various causal impacts on the ubiquitous detector that S is a member of the set of sentences which hold up in all sequences of empirically improving theories.... However, such an explanation is nonsense if the two sets of sentences are *a priori* coextensive.

What a metaphysical antirealist needs to do in order to avoid this difficulty is to abandon an idealized rational acceptability theory of truth (at least for sentences concerning detectable states of affairs). Grice's metaphysical constructivism allows the metaphysical antirealist to do just that. We can think of the empirical world as a world made up of properties, relations, and entities all of which are the result of Gricean construction. We can then define empirical truth in such a way that it involves a

correspondence between the state of affairs asserted to hold by the sentence and a state of affairs which actually does hold in the empirical world. Such a theory of truth has much in common with that favored by the semantic realist. The difference is that in this case truth involves correspondence with a world that is constructed, and hence, mind-dependent. So, it would seem that if Grice's constructivist program can be coherently developed, the theory of empirical truth which defines truth in terms of correspondence with the constructed empirical world should have no more difficulty with the missing explanation argument than would the more traditional semantic realist theory of truth in terms of correspondence with a mind-independent world.

But can Grice's metaphysical constructivism be coherently developed? It is well beyond the scope of this chapter to provide a complete defense of Grice's metaphysical program or to attempt to develop any single version of the program in sufficient detail to establish once and for all that it is coherent. What I want to do in the remainder of this section is instead to raise and discuss a number of questions (many of which were raised by Grice himself⁹⁸) concerning Grice's program that will need to be answered if any way of executing the program is to constitute a viable alternative to metaphysical realism. Few, if any, of these questions will be satisfactorily answered here. However, it is not obvious that they cannot be answered and, especially given the problems faced by the metaphysical realist, there is, I think, ample reason for philosophers to devote both time

⁹⁸ See especially "Reply to Richards," pp. 89-93. Robert Coburn also recalls that many of these questions, and perhaps others as well, were raised by Grice in his lectures at the University of Washington in 1980 and 1983.

and energy toward examining the extent to which these problems can be addressed by the metaphysical antirealist.

The first worry one might have about a constructivist program such as the one developed by Grice is that it is susceptible to an analogue of the regress argument against the view, presented in Chapter Five,⁹⁹ that, e.g., the height of mount Everest is a social construct. The argument, in outline, was this: if the sense in which an entity, property or concept is mind-dependent is that it is the result of the operation of social or linguistic conventions, or in Grice's case, of rationally motivated uses of one or more metaphysical construction routines, on prior entities, properties, or concepts, then either a regress threatens, or there must be some basic unconstructed, and hence plausibly mind-independent, entities or properties. In either case, we do not have a complete and coherent explanation of the mind-dependent nature of the world as a whole. It seems to me that, unlike the social construct view discussed in Chapter Five, Gricean constructivism (or at least one way the program might be executed) is plausibly immune to the charge that it threatens an infinite regress or accepts some unconstructed entities.

Recall that the first component of metaphysical constructivism would be, according to Grice, theorizing about the role and purpose of theorizing in an effort to identify what might be called the First Theory. First Theory would amount to a full-dress characterization of "the ideas and materials presupposed by any theory."¹⁰⁰ Any entities

⁹⁹ See pp. 156-158.

¹⁰⁰ "Reply to Richards," p. 87.

or properties postulated by First Theory would be able to serve as the metaphysical primitives in an ontological scheme (a constructive ordering of categories of properties and entities) and might well be regarded as original, unconstructed, and mind-independent. However, Grice is willing to flirt with the idea that it may not be the case that our theory-theorizing would yield the result that every theory has to relate to a certain range of subject items and attributes. In such a case, the question, "which are the primitive, unconstructed entities?," would be one that it only makes sense to ask from within an ontological scheme.

Are we, if we lend a sympathetic ear to constructivism, to think of the metaphysical world as divided into a constructed section and a primitive, original, unconstructed section? I will confess at once that I do not know the answer to this question. The forthright contention that if there is a realm of constructs there has to be a realm of non-constructs to provide the material upon which the earliest ventures in construction are to operate, has its appeal....But I am by no means sure that it is correct....Part of my trouble is that there does not seem to me to be any good logical reason calling for a class of ultimate non-constructs. It seems to me quite on the cards that metaphysical theory...might consist of a package of what I will call ontological schemes in which categories of entities are constructively ordered, that all or most of the same categories may appear within two different schemes with different ordering, what is primitive in one scheme being non-primitive in the other, and that this might occur whether the ordering relations employed in the construction of the two schemes were the same or different. We would then have no role for a notion of *absolute* primitiveness; all we could use would be the relative notion of primitiveness-with-respect-to-a-scheme.¹⁰¹

If it is, as Grice says, quite on the cards that there would be no role for absolute primitiveness to play, then there need be no absolutely original, unconstructed entities and

¹⁰¹ Ibid., pp. 91-92.

properties. Such a version of Grice's metaphysic, if capable of coherent development, would avoid the regress argument raised in Chapter Five.

However, Grice realizes that his metaphysical program will be charged with circularity if it allows for the possibility that there is no set of original, unconstructed entities and properties. It may appear that this way of executing the constructivist program avoids a potentially vicious regress only by embracing an equally vicious circle. For this reason, Grice recognizes the need to "get clear about the nature and forms of real or apparent circularity, and to distinguish those forms, if any, which are innocuous from those which are deadly."¹⁰² He proposes to do so by examining the notion of *priority*, that is examining whether "the supposition that A is prior to B allows or disallows the possibility that B may also be prior to A either in the same, or in some other dimension of priority."¹⁰³ To this end, Grice uses the example of legal and moral rights. There are obvious ways to take the legal concept of a right as prior to the moral. However, that does not rule out the possibility that the moral concept is "valuationally" prior to the legal concept; "the range of application of the legal concept *ought to be* always determined by criteria which are couched in terms of the moral example."¹⁰⁴ A second example used by Grice involves the properties of sense-data and material objects.

It might be, perhaps, that the properties of sense-data,...(and so sense-data themselves) are posterior in one sense to corresponding properties of

¹⁰² Ibid., p. 92.

¹⁰³ Ibid., pp. 92-93.

¹⁰⁴ Ibid., p. 93. Compare John Stuart Mill's discussion of the relation between legal and moral rights in the last of chapter of his *Utilitarianism* (Indianapolis: Hackett, 1979), especially pp. 42-43.

material things (and so to material things themselves); properties of material things, perhaps, render properties of sense-data intelligible by providing a paradigm for them. But when it comes to the provision of a suitably motivated *theory* of material things and their properties, the idea of making these *definitionally* explicable in terms of sense-data and their properties may not be ruled out by the holding of the aforementioned conceptual priority in the reverse direction.¹⁰⁵

Grice takes these examples to show that it is at least plausible that the claim that a set of entities A may be prior to another set B in one ontological scheme and posterior to B in another is, while circular, not viciously so.¹⁰⁶ While it is clear that much more needs to be said to fully defend the idea that there may be no role for the notion of absolute primitiveness to play and that the circularity that results is not vicious, it is far from clear that such a defense cannot be provided, and so far from clear that metaphysical constructivism would be susceptible to any attempted analogue of the regress argument which charged it with vicious circularity or the recognition of unconstructed entities.

However, even if we have found a plausible way of adequately addressing the previous concern, there are still many other questions raised by Grice in connection with his program—questions to which he has no worked out and well-defended answers. To begin with, what exactly is constructed? Individuals? Types? States of affairs (i.e., facts about individuals or types)? Presumably it will depend upon which routines of construction are used. Metaphysical Transubstantiation, for example, seems to construct

¹⁰⁵ "Reply to Richards," p. 93.

¹⁰⁶ To be sure, we have seen other, apparently non-vicious, explanatory circles above. Recall Shoemaker's circular analysis of the notions of causality, properties and identity of continuants discussed in Chapter Three, pp. 65-70.

entity types, while Humean Projection constructs concepts or properties. But if, for example, types of entities are constructed, does it follow that individual instances of those types are also constructed? That is, if, for example, material objects are constructed, does it follow that this table or that tree is constructed? It is not at all clear what should be said here.

Another series of questions raised, but not answered, by Grice is the following: Do there have to be constructors? Who (or what) does the constructing? Individuals? All of us together somehow? Robert Coburn recalls that Grice even suggested that perhaps types, rather than individuals, do the constructing.¹⁰⁷ And who (or what) constructs the constructors? Or are *they* necessarily unconstructed entities? This last question is of particular importance to the metaphysical antirealist. It would seem that it must be answered in the negative if Grice's metaphysical program is to yield a version of metaphysical antirealism at all (as opposed to a version of subjective idealism). The fact that Grice regarded this as an open question shows that he, at least, did not think that it was impossible that the metaphysical constructors were themselves constructed entities.

Perhaps an indication of how and why Grice thinks such a thing would be possible can be gleaned from his willingness to entertain the possibility of an ontological scheme in which sense-data and other mental particulars serve as the metaphysical building blocks for further entities including material objects.¹⁰⁸ The idea might be that what is primitive

¹⁰⁷ In lectures at the University of Washington in 1980 and 1983.

¹⁰⁸ See "Reply to Richards," p. 93, passage quoted above.

in this ontological scheme (though not necessarily in others) are sense-data and other mental particulars and that from these primitives minds, among other things, will be constructed. However, assuming that there must be constructors for metaphysical construction to occur, it would seem that the constructors must be rational creatures of some sort or another in that a successful use of a routine of construction must be, according to Grice, rationally motivated. As it is clear that the primitives acknowledged in this ontological scheme (sense-data, etc.) cannot be doing the constructing, the constructors must be rational creatures, and yet, minds are among the constructed entities.¹⁰⁹ There is a sense in which the minds have constructed themselves. Many will be quick to point out that it is far from clear that such a picture even makes sense. However, it seems to me, and so it seemed to Grice, that it is far from clear that such a line cannot be coherently developed, particularly if one can help oneself to the idea that what is primitive in one ontological scheme may be a constructed entity in another.¹¹⁰

¹⁰⁹ This way of thinking about Grice's example was suggested to me by Robert Coburn.

¹¹⁰ It is interesting to note in this connection that Grice remarks, "...I once invented (although I did not establish its validity) a principle which I labeled as *Bootstrap*. The principle laid down that when one is introducing the primitive concepts of a theory formulated in an object language, one has freedom to use any battery of concepts expressible in the meta-language, subject to the condition that counter-parts of such concepts are subsequently definable or otherwise derivable in the object-language" ("Reply to Richards," p. 93). Perhaps Grice would be willing to consider some analogue of this Bootstrap principle which would allow a constructed entity to, in a sense, be its own constructor. Perhaps the idea would be that if there is a legitimate ontological scheme wherein minds are primitive, then any other ontological scheme in which minds are constructed entities can inherit legitimacy from the first.

There are still other questions which must be addressed by anyone who would attempt to pursue Grice's program. Does metaphysical construction occur in time? If not, how should we understand what's going on in metaphysical construction? After all, if construction doesn't occur in time, then there aren't constructors in the sense of agents actively doing the construction. And, in this case, it becomes difficult to understand in what sense we have a *constructivist* program. Grice discusses one way of answering at least some of these questions using the analogy of a fictional character who may be "constructed" in 1985 but whose fictional life spans 1764-1798. When dealing with fictional characters and events, we have no difficulty with the idea that a constructed entity or event exists (in some sense) at a time prior to its construction. Reflection on this fact may cause us to wonder why we are bothered by such an idea concerning metaphysical construction. What exactly is different about it?

We may perhaps find it tolerable to suppose that a particular great mathematician should in 1968 make it true that (let us say) ultra-lunary numbers should exist timelessly or from and to eternity. We might even, should we so wish, introduce a 'depersonalized' (and 'detemporalized') notion of construction; in which case we say that in 1968 the great mathematician...not only constructed the timeless existence of sub-lunary numbers, but also *thereby* the depersonalized and detemporalized construction of the timeless existence of sub-lunary numbers, and also the depersonalized construction of the depersonalized construction,... of sub-lunary numbers in this way we might be able in one fell swoop to safeguard the copyrights both of the mathematician and of eternity.¹¹¹

So, Grice is willing to entertain the possibility that, at least in the case of mathematical and other timeless entities, once construction has taken place (which it does in its personalized

¹¹¹ Ibid., p. 90, my italics.

sense—that is, the sense in which the mathematician is the constructor—at a particular time, say 1968), the entities constructed have always existed and have existed quite independently of human beings.¹¹² Again, while it is not obvious that such a version of the program is capable of coherent execution, the fact that a philosopher of Grice's stature suspected that it might be should give us reason not to dismiss the possibility outright and to welcome any serious attempts at providing the necessary details of such a view.

Yet another series of questions arises given the fact that, as Grice views the project of metaphysical construction, it is answerable to, e.g., our evolving views of rational acceptability and the roles served by our theories, since it is these (among other things) that inform our theory-theorizing and serve as the motivation for any instance of metaphysical construction. As we saw above, if a use of one of the construction routines is not properly motivated, it fails at its goal of metaphysical construction and amounts to little more than a new way of speaking. Since our views of what is rationally acceptable and what roles various theories are supposed to serve may change over time, a use of one of Grice's construction routines which at one time seemed properly motivated, and hence to result in the metaphysical construction of a new set of concepts, properties, or entities, may come to seem unmotivated. What should we say in such a case? Is de-construction possible to correct for our mistake? Do those previously constructed entities simply “pop” out of existence once we come to realize that the use of the construction routine

¹¹² Note that, if this story makes sense, it also provides the metaphysical antirealist with a model for an answer to van Inwagen's geological argument, discussed in Chapter Five, pp. 150-152.

that created them was not, by current standards, properly motivated? Or were we wrong in thinking that the previous “entities” had been constructed at all? Were we, perhaps, misled by a mere *façon de parler*?

In his lectures at the University of Washington, Grice indicated some attraction to one way of developing a version of his program according to which metaphysical truth and scientific truth would be distinct. On this picture, the role of science is the discovery of scientific truth where scientific truth might well be construed as involving correspondence to a pre-existing reality, although one which has been metaphysically constructed. Metaphysical truth, on the other hand, would concern the world as constructed and would not be answerable to any pre-existing reality. One way this version of the view might be further developed which would address some of the questions raised in the previous paragraph would be to define metaphysical truth in terms of (or make metaphysical truth coextensive with) the set of strongly theoretic sentences (in Johnston’s sense¹¹³) that would be accepted at the ideal limit of theory-theorizing and metaphysical construction. If the notion of detemporalized construction makes sense, it could be the case *now* that such and such concepts, properties, or objects exist because they result from the use of one or more metaphysical construction routines (ones with which we need not now be familiar) which would be properly motivated at the ideal limit of metaphysical construction. Despite the fact that such a view would amount to the acceptance of something very like an idealized rational acceptability theory of metaphysical truth, it would quite plausibly

¹¹³ See above, pp. 204f.

escape the missing explanation argument which proved the undoing of Putnam's internal realism. It was, after all, the inclusion of sentences about detectable states of affairs in the set of sentences for which an idealized rational acceptability theory of truth is correct that caused the difficulties for Putnam's view. This way of developing Grice's picture avoids that difficulty since it, unlike Putnam's view, maintains a sharp distinction between metaphysical and empirical or scientific truth. However, it does so at the cost of a somewhat counter-intuitive bifurcation of the notion of truth.¹¹⁴

So, there are many questions concerning Grice's program to which he has no worked out and well-defended answers. These questions, and probably many more, must be addressed by the metaphysical antirealist who would adopt Gricean metaphysical constructivism. There are, to be sure, legitimate worries about whether the program can in fact be coherently executed in such a way as to yield a version of metaphysical antirealism. However, it is by no means obvious that the program is not susceptible of coherent execution in a way that would result in a philosophically viable alternative to metaphysical realism.

As we saw in the earlier chapters of this dissertation, the metaphysical realist too faces a number of difficulties. If his view is to be consistent with some of our most basic intuitions about the world, he must be able to provide a coherent and detailed account of

¹¹⁴ To be sure, it is not obvious that even the metaphysical and semantic realist can ultimately avoid such a bifurcation in the notion of truth. Correspondence with the mind-independent world is all well and good for sentences expressing empirical states of affairs. But what exactly is it for a metaphysical picture to correspond with the mind-independent world? It is not clear that such an idea makes much sense.

how it is that we are able to have semantic and epistemic access to a completely mind-independent world and also provide us with reasons for rejecting various very bizarre ontologies. While it is not at all clear that the metaphysical realist cannot rise to the challenges that face his view, neither is it obvious that he can. It seems to me that, given the position in which he finds himself, coupled with the fact that several very good philosophers have been willing to lend a sympathetic ear to metaphysical antirealism, the metaphysical realist should be concerned to do more than merely point out that it is not at present clear that Grice's program can be developed into a viable version of metaphysical antirealism—at least if he is to adequately defend his view and show that he should retain the default position in the debate. The metaphysical realist would do well to either provide an argument that Grice's program *cannot* be coherently executed in such a way as to yield metaphysical antirealism or to develop a theory concerning our semantic and epistemic access to the mind-independent world which is capable of escaping the criticisms raised in Chapters Two through Four.¹¹⁵ Until one or the other is forthcoming, and so long as we can continue to see some promise in the possibility of developing a coherent and complete Gricean metaphysic,¹¹⁶ it does not seem to me unreasonable to declare that the debate has reached some sort of balance in that it may well be a matter of

¹¹⁵ Or perhaps to provide an argument for the conclusion that any coherent and viable version of metaphysical antirealism will inherit all or most of the difficulties faced by the metaphysical realist.

¹¹⁶ Or, if the internal realist can develop a response to the missing explanation argument along the lines suggested by Coburn (see above, p. 210, footnote 56), some promise in the possibility of further developing Putnamian internal realism.

philosophical judgment which view, metaphysical realism or metaphysical antirealism, has the more severe problems.

However, this balance is an extremely delicate one. Even if the metaphysical realist does nothing to provide a defense for his view, it could well be that, say, after a series of unsuccessful attempts, philosophers will give up on Gricean metaphysical constructivism. It seems not implausible to assert that unless some progress is made toward filling in the details of a coherent metaphysical antirealist view, the metaphysical realist should be awarded the benefit of the doubt. The *coherence* of his view was not in question; what was in question was whether or not metaphysical realism could be made consistent with some of our basic intuitions about the world and our relation to it. So, there may well be a sense in which, even if the metaphysical realist is not given the benefit of the doubt by default, his is still the dialectically stronger position in the debate. The metaphysical realist has (at least for the time being) the luxury of being able to sit and wait while the metaphysical antirealist, it would seem, must make philosophical progress just to maintain the balance in the debate. Even so, I think that the evidence I have presented here suggests that it is well worth struggling to see if such progress can be made.

SELECTED BIBLIOGRAPHY

- Anderson, David L. "What is Realistic about Putnam's Internal Realism?." *Philosophical Topics* 20 (Spring, 1992): 49-83.
- Anderson, David L. "What is the Model-Theoretic Argument?." *The Journal of Philosophy* 90 (June, 1993): 311-322.
- Aristotle. *Metaphysics*, Revised Oxford Translation. Edited by J. Barnes. *Complete Works of Aristotle*. Princeton: Princeton University Press, 1984.
- Armstrong, D. M. *A Combinatorial Theory of Possibility*. Cambridge: Cambridge University Press, 1989.
- Atlas, Jay David. *Philosophy Without Ambiguity*. Oxford: Clarendon Press, 1989.
- Bennett, Jonathan. *Events and Their Names*. Indianapolis: Hackett, 1988.
- Berkeley, Bishop. *Three Dialogues Between Hylas and Philonous*. London: 1713.
- Berkeley, Bishop. *A Treatise Concerning the Principles of Human Knowledge*. London: 1910.
- BonJour, Laurence. *The Structure of Empirical Knowledge*. Cambridge: Harvard University Press, 1985.
- BonJour, Laurence. "Against Naturalized Epistemology." Edited by Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, *Midwest Studies in Philosophy*, vol. 19, *Philosophical Naturalism*. Minneapolis: University of Minnesota Press, 1994: 283-300.
- Boyd, Richard. "How to be a Moral Realist." Edited by Geoffrey Sayre-McCord, *Essays on Moral Realism*. Ithaca: Cornell University Press, 1988.
- Boyd, Richard. "On the Current Status of Scientific Realism." Edited by Richard Boyd, Philip Gasper, and J.D. Trout, *The Philosophy of Science*. Cambridge: MIT press 1991: 195-222.
- Boyd, Richard. "Introductory Essay: Confirmation, Semantics and the Interpretation of Scientific Theories." Edited by Richard Boyd, Philip Gasper, and J.D. Trout, *The Philosophy of Science*. Cambridge: MIT press 1991: 3-37.
- Bridgman, P. W. *The Logic of Modern Physics*. New York: Macmillan, 1927.
- Brody, Baruch A. *Identity and Essence*. Princeton: Princeton University Press, 1980.

- Brown, Curtis. "Internal Realism: Transcendental Idealism?" Edited by Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, *Midwest Studies in Philosophy*, vol. 12, *Realism and Antirealism*. Minneapolis: University of Minnesota Press, 1988: 145-155.
- Carnap, R. *Meaning and Necessity*. Chicago: University of Chicago Press, 1947.
- Coburn, Robert. "Evolution and Skepticism." *Pacific Philosophical Quarterly* 71 (1990): 1-13.
- Coburn, Robert. *Strangeness of the Ordinary: Problems and Issues in Contemporary Metaphysics*. Savage, MD: Rowman and Littlefield, 1990.
- Copi, I. M. "Essence and Accident." Edited by S. Schwartz, *Naming, Necessity and Natural Kinds*. Ithaca: Cornell University Press, 1977.
- Davidson, Donald. *Inquiries into Truth and Interpretation*. New York: Oxford University Press, 1984.
- Devitt, Michael. "Review of Putnam's *Reason, Truth, and History*." *Philosophical Review* 92 (1984): 274-277.
- Devitt, Michael. *Realism and Truth*. Cambridge: Basil Blackwell, 1991.
- Dummett, Michael. "What is a Theory of Meaning." *Mind and Language*. Oxford: Clarendon Press, 1976: 105-108.
- Dummett, Michael. *The Logical Basis of Metaphysics*. Cambridge: Harvard University Press, 1991.
- Edwards, Paul (ed.). *The Encyclopedia of Philosophy*. New York: Macmillan and the Free Press, 1967.
- Einstein, Albert. "Autobiographical Notes." Edited by Paul Schillip, *Albert Einstein: Philosopher Scientist, The Library of Living Philosophers*, vol. 7 (Evanston, 1949): 1-96.
- Field, Hartry. "Realism and Relativism." *The Journal of Philosophy*. (1982): 553-567.
- Boolos, George and Richard Jeffrey. *Computability and Logic*. Cambridge: Cambridge University Press, 1989.
- Grice, Paul. "Method in Philosophical Psychology (From the Banal to the Bizarre)." *Proceedings and Addresses of the American Philosophical Association* 48 (November, 1975): 23-53.

- Grice, Paul. *The Conception of Value*. New York: Claredon and Oxford University Press, 1991. Given as *The Carus Lectures*, 1983.
- Grice, Paul. "Reply to Richards." Edited by Richard Grandy and Richard Warner, *Philosophical Grounds of Rationality*. Oxford: Oxford University Press, 1986: 45-106.
- Grice, Paul. *Studies in the Way of Words*. Cambridge: Harvard University Press, 1989.
- Hare, R. M. *The Language of Morals*. New York: Oxford University Press, 1952.
- Heller, Mark. "Putnam, Reference, and Realism." Edited by Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, *Midwest Studies in Philosophy*, 12, *Realism and Antirealism*. Minneapolis: University of Minnesota Press, 1988: 113-127.
- Hirsch, Eli. *The Concept of Identity*. Oxford: Oxford University Press, 1982.
- Horwich, Paul. *Probability and Evidence*. London: Cambridge University Press, 1982.
- Johnston, Mark. "Objectivity Refigured: Pragmatism Without Verificationism." Edited by John Haldane and Crispin Wright, *Reality, Representation, and Projection*. New York: Oxford University Press, 1993: 85-130.
- Johnston, Mark. "Constitution is Not Identity." *Mind* 101 (January, 1992): 89-105.
- Johnston, Mark. "Reasons and Reductionism." *The Philosophical Review* 101 (July, 1992): 589-618.
- Johnston, Mark. "Verificationism as Philosophical Narcissism." Edited by James Tomberlin, *Philosophical Perspectives*, vol. 7, *Language and Logic*. Atascadero, CA: Ridgeview, 1993: 307-330.
- Kornblith, Hilary. *Inductive Inference and Its Natural Grounds*. Cambridge: The MIT Press, 1993.
- Kornblith, Hilary. "Naturalistic Epistemology and Its Critics." Unpublished, (1994).
- Kripke, Saul. "Identity and Necessity." Edited by Donald Davidson and Gilbert Harman, *Semantics of Natural Language*. Dordrecht: D. Reidel, 1972.
- Kripke, Saul. "Naming and Necessity." Edited by Donald Davidson and Gilbert Harman, *Semantics of Natural Language*. Dordrecht: D. Reidel, 1972. Published in book form, Cambridge: Harvard University Press, 1980.
- Kuhn, Thomas S. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press, 1962.

- Lepore, Ernest, and Barry Loewer. "A Putnam's Progress." Edited by Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, *Midwest Studies in Philosophy*, vol. 12, *Realism and Antirealism*. Minneapolis: University of Minnesota Press, 1988: 459-473.
- Lewis, David. "Putnam's Paradox." *Australasian Journal of Philosophy* 62 (September, 1984), 221-236.
- Lewis, David. *On the Plurality of Worlds*. Oxford and New York: Basil Blackwell, 1986.
- Loar, Brian. "Truth Beyond All Verificationism." Edited by, Barry Taylor, *Michael Dummett: Contributions to Philosophy*. Dordrecht: Nijhoff: 81-116.
- Locke, John. *Essay Concerning Human Understanding*. Edited by P.H. Nidditch. Oxford: Clarendon Press, 1975; First published in 1689.
- Lowe, E. J. *Kinds of Being*. Oxford: Basil Blackwell, 1989.
- Mackie, J. L. *Ethics: Inventing Right and Wrong*. London: Penguin Books, 1977.
- Mackie, J. L. "De What Re Is De Re Modality?" *The Journal of Philosophy* 71 (September, 1974): 551-61.
- Merill, G. H. "The Model-Theoretic Argument Against Realism." *Philosophy of Science* 47 (1980): 69-81.
- Musgrave, Allan. "Realism, Truth, and Objectivity." Unpublished.
- Nagel, Thomas. *The View from Nowhere*. New York: Oxford University Press, 1986.
- Nerlich, Graham. *The Shape of Space*. Cambridge: Cambridge University Press, 1976.
- Owen, David. "Locke On Real Essence." *History of Philosophy Quarterly* 8 (April, 1991): 105-118.
- Plantinga, Alvin. "Actualism and Possible Worlds." *Theoria* 42 (1976): 139-160.
- Plantinga, Alvin. *Warrant: the Current Debate*. New York: Oxford University Press, 1993.
- Putnam, Hilary. "The Analytic and the Synthetic." *Philosophical Papers*, vol. 2, *Mind, Language and Reality*. Cambridge: Cambridge University Press, 1975: 33-69.
- Putnam, Hilary. "Language and Reality." *Philosophical Papers*, vol. 2, *Mind, Language and Reality*. Cambridge: Cambridge University Press, 1975: 272-290.

- Putnam, Hilary. "The Meaning of 'Meaning'." *Philosophical Papers*, vol. 2, *Mind, Language and Reality*. Cambridge: Cambridge University Press, 1975: 215-271.
- Putnam, Hilary. "The Refutation of Conventionalism," *Philosophical Papers*, vol. 2, *Mind, Language and Reality*. Cambridge: Cambridge University Press, 1975: 153-191.
- Putnam, Hilary. "Realism and Reason." *Meaning and the Moral Sciences*. London: Routledge and Kegan Paul, 1978: 123-140. First given as the Presidential Address to the eastern Division of the American Philosophical Association, Boston, Mass, 29 December, 1976.
- Putnam, Hilary. *Meaning and the Moral Sciences*. London: Routledge and Kegan Paul, 1978.
- Putnam, Hilary. *Reason, Truth, and History*. Cambridge: Cambridge University Press, 1981.
- Putnam, Hilary. "Models and Reality." *Journal of Symbolic Logic* 45 (1985): 464-482.
- Putnam, Hilary. *The Many Faces of Realism*. La Salle, IL: Open Court, 1987.
- Putnam, Hilary. *Realism with a Human Face*. Cambridge: Harvard University Press, 1990.
- Putnam, Hilary. *Realism and Reason: Philosophical Papers*, vol. 3, *Realism and Reason*. Cambridge: Harvard, 1992.
- Putnam, Hilary. "The Dewey Lectures 1994: Sense, Nonsense, and the Senses: An Inquiry into the Powers of the Human Mind." *The Journal of Philosophy* 91 (Sept. 1994): 445-517.
- Putnam, Hilary. *Pragmatism*. Oxford: Blackwell, 1995.
- Quine, W. V. O. "Identity, Ostension and Hypostasis." *From a Logical Point of View*. Cambridge: Harvard University Press, 1966: 65-79.
- Quine, W. V. O. "Truth by Convention." *The Ways of Paradox*. Cambridge: Harvard University Press, 1966: 9-99. First published in O.H. Lee, ed., *Philosophical Essays for A. N. Whitehead*. New York: Longmans, 1936.
- Quine, W. V. O. "Two Dogma's of Empiricism." *From a Logical Point of View*. Cambridge: Harvard University Press, 1953: 20-46.
- Quine, W. V. O. *Word and Object*. Cambridge: MIT Press, 1960.

- Quine, W. V. O. "On What There Is." *Review of Metaphysics* 2 (Spring 1948): 21-38.
- Resnik, Michael. "You Can't Trust an Ideal Theory to Tell the Truth." *Philosophical Studies* 52 (1987): 151-160.
- Schlesinger, George. "Natural Kinds." *Boston Studies in the Philosophy of Science*, vol. 3. Dordrecht: D. Reidel, 1967.
- Schwartz, Robert. "I'm Going to Make You a Star." Edited by Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, *Midwest Studies in Philosophy*, vol. 11, *Studies in Essentialism*. Minneapolis: University of Minnesota Press, 1986: 427-439.
- Shoemaker, Sydney. "Causality and Properties." *Identity, Cause and Mind*. Cambridge: Cambridge University Press, 1984: 206-233.
- Shoemaker, Sydney. "Identity, Properties, and Causality." *Identity, Cause and Mind*. Cambridge: Cambridge University Press, 1984: 234-260.
- Shoemaker, Sydney. "On What There Are." *Philosophical Topics* 16 (Spring 1988): 201-223.
- Sidelle, Alan. *Necessity, Essence, and Individuation*. Ithaca: Cornell University Press, 1989.
- Sidelle, Alan. "Rigidity, Ontology, and Semantic Structure." *Journal of Philosophy* 89 (August, 1992): 410-430.
- Sosa, Ernest. "Putnam's Pragmatic Realism." *The Journal of Philosophy* 90 (December, 1993): 605-626.
- Stalnaker, Robert. "Anti-Essentialism." Edited by Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, *Midwest Studies in Philosophy*, vol. 4, *Studies in Metaphysics*. Minneapolis: University of Minnesota Press, 1979: 343-355.
- Stevenson, C. L. *Ethics and Language*. New Haven: Yale, 1944.
- Strawson, P. F. *The Bounds of Sense*. London: Methuen and Company LMT, 1966.
- Van Cleve. "Semantics, Supervenience, and Referential Indeterminacy." *The Journal of Philosophy* 89 (July 1992): 344-361.

- Van Inwagen. "On Always Being Wrong." Edited by Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, *Midwest Studies in Philosophy*, vol. 12, *Realism and Antirealism*. Minneapolis: University of Minnesota Press, 1988: 95-111.
- Van Inwagen, Peter. *Material Beings*. Ithaca: Cornell University Press, 1990.
- Van Inwagen, Peter. *Metaphysics*. Boulder: Westview Press 1993.
- Walker, Ralph C. S. *Kant*. Edited by Ted Honderich, *The arguments of the Philosophers*. London: Routledge & Kegan Paul, 1978.
- Wiggins, David. *Sameness and Substance*. Cambridge: Harvard University Press, 1980.
- Wittgenstein, Ludwig. *Philosophical Investigations*. New York: MacMillan, 1958.

APPENDIX: SOME HEROIC ATTEMPTS TO AVOID SKEPTICISM

In Chapter Two, section 1, I suggested that it might be possible for a metaphysical realist to meet the skeptic's challenge by adopting what Nagel has called a heroic epistemology, one that is able to bridge the gap between the mind-independent world and the mind-dependent appearances. In this appendix I will examine two relatively recent attempts at such a heroic project: a version of the coherence theory of justification, which is developed by Laurence BonJour in *The Structure of Empirical Knowledge*,¹ and a version of so-called externalist epistemologies. I shall give reason to doubt that either of these attempts is wholly successful against the skeptic. But, even if the difficulties I point out can be handled, at least it should be clear that the metaphysical realist has a much bigger job on his hands than many metaphysical antirealists have—at least if we assume that the theories I consider here are among the most promising of those that metaphysical realists have developed for meeting the skeptical challenge.

According to BonJour's coherentist theory, a belief B is justified for a doxastic agent S only if it is a member of S's system of beliefs and that system is coherent. For S's system of belief to be coherent it must be both logically consistent (it must be logically possible that all of the beliefs are true) and have a high degree of probabilistic consistency (a doxastic system will be probabilistically inconsistent if it contains both the belief that P and the belief that P is likely to be false).² The coherence of a doxastic system is further

¹ Laurence BonJour, *The Structure of Empirical Knowledge* (Cambridge: Harvard University Press, 1985).

² *Ibid.*, p. 95.

increased in proportion to the number and strength of inferential connections among its component beliefs and decreased to the extent to which it is divided into inferentially unconnected subsystems of beliefs and by the presence of unexplained anomalies in the system.³

However, a high degree of coherence in a doxastic system at a given time is not sufficient for that system's conferring justification on any of its component beliefs, according to Bonjour. If the coherence of the doxastic system is to confer any justificatory status on the component beliefs, such coherence must be truth-conducive. As it seems likely that a doxastic system could enjoy a high degree of coherence at a given moment simply by accident, it is clear that momentary coherence will not be truth-conducive. It is for this reason that Bonjour adds the further requirements that the agent's doxastic system be coherent for a sufficiently long run.⁴ Yet a further restriction is required if the theory is to give us an account of *empirical* knowledge:⁵ it must allow for putative input from the empirical world into the doxastic system of the agent. That is, it must meet what Bonjour dubs the "Observation Requirement" as well. The Observation Requirement is briefly stated by Bonjour as follows:

[I]n order for the beliefs of a cognitive system to be even candidates for empirical justification, that system must contain laws attributing a high degree of reliability to a reasonable variety of cognitively spontaneous

³ *Ibid.*, pp. 98-99.

⁴ While it is still possible that even the long-run coherence of a doxastic system is an accident, it is presumably far less likely.

⁵ This requirement is also needed to preclude the possibility that a doxastic system is long-run coherent because completely static.

beliefs (including in particular those kinds of introspective beliefs which are required for the recognition of other cognitively spontaneous beliefs).⁶

A belief is cognitively spontaneous if it is not the result of any inference. Instances of such cognitively spontaneous beliefs would include perceptual and memory beliefs. Such spontaneous beliefs, when there is reason from inside the doxastic system to regard them as generally reliable,⁷ function as input into the doxastic system. Given such input, the continued coherence of a doxastic system over a sufficiently long-run is claimed to be much more conducive to truth—make it much more likely that the beliefs which make up the doxastic system are true—than the coherence of a static, or momentary, doxastic system. Unlike the case of momentary coherence, the continued coherence of a dynamic doxastic system would cry out for an explanation, particularly if the doxastic system were to converge on a specific world view and “thereafter remain relatively stable, reflecting only those changes (such as the passage of time and the changes associated with it) which are allowed or even required by the general picture of the world thus presented.”⁸

And yet, the long-run coherence and fulfillment of the Observation Requirement of S’s belief system will not confer justification on any of the component beliefs unless S has a reflective grasp of the long-run coherence of the system and realizes that such coherence is truth-conducive. That is, for any of my empirical beliefs to be justified, according to

⁶ *The Structure of Empirical Knowledge*, p. 141.

⁷ Notice that the spontaneous beliefs need only be *regarded from inside the system as reliable*. Bonjour adds no requirement that they actually *be reliable*. This is one important respect in which his view differs from the reliabilist view which will be examined below.

⁸ *The Structure of Empirical Knowledge*, p. 170.

BonJour, I must be in implicit possession of a justificatory argument of the following form:

1. B is one of my beliefs B_1, \dots, B_n and that doxastic system consisting of B_1, \dots, B_n has been coherent for a sufficiently long run while fulfilling the Observation Requirement.
2. If B is a member of S's system of beliefs and S's system of beliefs has been coherent for a sufficiently long run while fulfilling the Observation Requirement, then B is likely to be true.
3. Therefore, B is likely to be true. That is, B has justification for me.

BonJour's coherentist account shares with the epistemological foundationalist theory⁹ an adherence to epistemic internalism, the view that I am justified in a belief B only if I am in implicit cognitive possession of a good reason for thinking that B is true. So, it will not be sufficient that the premises of the justificatory argument are true if I am not also justified in believing that they are true. Since B is, *ex hypothesi*, an empirical belief, it does not seem possible that premises 1 and 2 can both be justified *a priori*; an empirical belief cannot be justified completely *a priori*. BonJour takes premise 2 to be justifiable *a priori*¹⁰ and denies that premise 1 is justified *a priori*. But how then is this premise to have justification for me? The answer seems to be, according to BonJour, that it isn't really justified at all. We must, instead, *presume* that my grasp of my system of beliefs is at least approximately correct. "[T]hough questions can be raised and answered

⁹ According to BonJour "the central thesis of epistemological foundationalism...is the twofold thesis: (a) that some empirical beliefs possess a measure of epistemic justification which is somehow immediate or intrinsic to them...and (b) that it these 'basic beliefs'...which are the ultimate source of justification for *all* of empirical knowledge"(ibid., p. 17).

¹⁰ Ibid., p. 158.

with regard to particular aspects of my grasp of my system of beliefs, the *approximate* accuracy of my overall grasp of that system must be taken for granted in order for coherentist justification to even begin.”¹¹ The assumption that my overall grasp of my own doxastic system is at least approximately accurate is what Bonjour calls the Doxastic Presumption.

While Bonjour does not¹² view this appeal to the Doxastic Presumption as particularly problematic, some have held that Bonjour’s reliance on the Doxastic Presumption yields the result that his view is essentially skeptical in nature. This claim has been made quite convincingly by Alvin Plantinga in his book *Warrant: the Current Debate*.¹³ I will briefly paraphrase his argument. In order for the above justificatory argument to confer justification on my belief B, I must be justified in believing that my system of beliefs has been coherent for a sufficiently long run (call this belief B_c). This metabelief is a further empirical or *a posteriori* belief.¹⁴ But B_c cannot get justified through my grasp of a metajustificatory argument for B_c since it must itself be one of the premises of any such argument. Thus B_c will be an *a posteriori* belief that is completely unjustified on a Bonjourian view. But if B_c is completely unjustified, how can I make use

¹¹ Ibid., p. 127.

¹² Or at least did not at the time of writing *The Structure of Empirical Knowledge*.

¹³ Alvin Plantinga, *Warrant: the Current Debate* (New York, Oxford: Oxford University Press, 1993). See especially, pp. 107-109.

¹⁴ While a belief concerning the momentary coherence of my doxastic system might be thought by some (though not by Bonjour) to be *a priori*, a belief concerning the long-run coherence of my doxastic system is clearly *a posteriori* involving as it does a memory belief.

of it in justifying any of my other *a posteriori* beliefs? An appeal to the Doxastic Presumption does not help at this point. The Doxastic Presumption is the claim that I am entitled to presume that my metabeliefs, including B_c are true, but it does not confer any justification on them. But as Plantinga points out “It isn’t sufficient that my metabeliefs happen to be true; if I am not justified in those beliefs, then my justificatory argument confers no justification upon B *for me*.”¹⁵ Thus, it seems that BonJour’s view entails that no one is actually justified *simpliciter* in believing any of the usual sorts of empirical beliefs, e.g., that there is a tree in front of me, that I was in Jerusalem in 1971, etc. The only empirical beliefs that one is justified in accepting on a BonJourian view are conditional in form—I am only justified in believing that *if* my grasp of my belief system is relatively accurate, *then* there is a tree in front of me. I would be justified *simpliciter* in believing the consequent of this conditional only if I could be justified in believing that the antecedent were true. But, on BonJour’s theory, although I may presume that the antecedent is true, I can never be justified in believing that this is the case. The view seems for this reason to fail to meet the skeptical challenge.

There is, however, a further problem that arises for a BonJourian coherentist, a problem that only arises if metaphysical and semantic realism are taken to be true. As I mentioned earlier, BonJour subscribes to what is called an internalist theory of justification. On any internalist theory of justification, justification must be conferred on an agent’s beliefs, if it is to be conferred at all, in virtue of some facts to which the agent

¹⁵ *Warrant: the Current Debate*, p. 109.

has cognitive access. That is, the justificatory features of the belief must be in some sense *internal* to the agent's doxastic system. It is for this reason that it is not sufficient that the premises of the justificatory argument for my belief be true, but is also necessary that I be internally justified in believing that they are true. At the same time, these justificatory features must be truth-conducive—they must make my belief likely to be true. However, there is a natural tension between these two requirements if truth is taken to be some sort of correspondence with the metaphysical realist's mind-independent world. Metaphysical realism entails that a world, e.g., inhabited entirely by Cartesian souls who have beliefs about the world that are very much like our own is a genuine metaphysical possibility.¹⁶ Any features that would serve as justificatory criteria for an agent's belief that are completely *internal* to the doxastic system of the agent are criteria that could be met for a Cartesian soul, most of whose beliefs are false according to the metaphysical and semantic realist. If this is the case, it is difficult to see how such justificatory criteria can be said to be truth-conducive, how they could confer epistemic justification on any belief. After all, if we consider all of the possible worlds in which the relevant criteria are met by an agent, it seems plausible that there will be at least as many of these worlds in which the agent's beliefs are false (because, e.g., he is a Cartesian soul) as there are in which his beliefs are true.¹⁷

¹⁶ Or if there is some inherent difficulty with the notion of a nonmaterial soul, we can make use of some other skeptical hypothesis.

¹⁷ To be sure, this objection trades on the assumption that probability is to be understood by reference to the relative frequency of one event given another. While it is a *prima facie* plausible assumption to make about the nature of probability, there are, of course,

BonJour's response to this objection is contained in his account of the justification of the second premise of the justificatory argument above, that if B is a member of an agent's system of beliefs and that system has been coherent for a sufficiently long-run while meeting the Observation Requirement, then B is likely to be true. If this account is successful, then the tension between internalist justification and metaphysical realism is avoided by a BonJourian coherentist view. So, it is to an analysis of this account that we will now turn.

BonJour attempts to argue that, with a little reflection, we can tell *a priori* that, given the long-run coherence of a belief system, the claim that the component beliefs correspond to the way the mind-independent world is, and hence are true, is more probable than any competing skeptical hypothesis.¹⁸ That is, although a world of deluded Cartesian souls is a metaphysically possible world (perhaps), it is *a priori* improbable that the long-run coherence of any doxastic system is the product of either random chance or of the manipulations of an evil demon who desires to deceive the believer, etc. Thus, the second premise of the justificatory argument is itself justified, according to BonJour, *a priori*; we appeal to the greater *a priori* probability that the correspondence hypothesis is true, that the long-run coherence of any doxastic system is the result of a correspondence between the components of that system and the mind-independent world.

alternatives; see Paul Horwich, *Probability and Evidence* (Cambridge: Cambridge University Press, 1982), esp. pp. 16-51.

¹⁸ *The Structure of Empirical Knowledge*, p. 181.

The probability that Bonjour appeals to is an objective probability. He reminds the reader that it is a theorem of the probability calculus that the probability of a given hypothesis H on evidence E is equal to the probability of E given H times the probability of H divided by the antecedent probability of E. That is,

$$\text{Prob}(H \text{ given } E) = \text{Prob}(E \text{ given } H)\text{Prob}(H)/\text{Prob}(E).$$

We are interested in the relative values that result where H1 is the correspondence hypothesis, H2 is the skeptical hypothesis and E is the evidence of long-run coherence of my doxastic system. Since we want to compare the *a priori* probability of two competing hypotheses on the same evidence, we can ignore the denominator, 'Prob(E)'. To determine which of the two competing hypotheses, the correspondence hypothesis or the skeptical hypothesis, has a greater *a priori* probability on the evidence of a long-run coherent doxastic system, what we need to do is to compare the probability of such coherence relative to each hypothesis (i.e., Prob(E given H1) and Prob(E given H2)), and the antecedent probability of each hypothesis (i.e., Prob(H1) and Prob(H2)).

BonJour proceeds in two stages: first he argues that the *a priori* probability of long-run coherence in a doxastic system is very low relative to the hypothesis that the features of the doxastic system are the product of chance or of a completely unspecified evil demon (i.e., nothing further is said to specify the intentions of the evil demon).¹⁹ Whereas, the probability of long-run coherence in a doxastic system is quite high relative

¹⁹ Bonjour dubs such unspecified demon hypotheses 'simple demon hypotheses'.

to the correspondence hypothesis. After all, we would expect a doxastic system to be relatively coherent if it accurately reflected the goings on in the world.

Like pure chance, such an unspecified demon is capable of producing, and equally likely to produce, virtually any configuration of beliefs, and the simple demon hypothesis provides no reason at all for expecting him to confine himself to those which will fit coherently into my cognitive system. The upshot is that the continued coherence-cum stability of my system of beliefs is excellent, though not totally conclusive, evidence *against* simple demon hypotheses, simply because it is a result which would be extremely unlikely to occur if such hypothesis were true.²⁰

Obviously, the probability that my beliefs would be long-run coherent merely by chance is *a priori* quite low. Since the demon hypothesis, according to BonJour, adds nothing to the chance hypothesis other than that my beliefs are caused by the demon (it says nothing about how or why he causes my belief), BonJour concludes that the continued coherence of my system of beliefs given a simple demon hypothesis is the same as it would be given a chance hypothesis, very low. Thus, since it is highly unlikely relative to both the chance and simple demon hypotheses that my beliefs would be coherent in the long-run, the overall probability of either the chance or the skeptical hypothesis on the evidence of coherence will be *a priori* quite low.

In the second stage of his argument, BonJour addresses skeptical hypotheses, which he calls elaborated demon hypotheses, on which it is highly probable that a given agent would have a coherent doxastic system. Such an elaborated demon hypothesis would be the traditional Cartesian story in which it is the expressed wish of an omnipotent

²⁰ *The Structure of Empirical Knowledge*, p. 184.

demon to delude the poor Cartesian souls into believing that the correspondence hypothesis is true. Relative to such an elaborated demon hypothesis, the probability of the coherence of the doxastic system should be just as high as (or even higher than) the probability of such coherence relative to the correspondence hypothesis itself—very nearly

1. However, BonJour argues that the antecedent probability of an elaborated demon hypothesis will be quite low relative to the antecedent probability of the simple demon hypothesis. Speaking of what he calls an elaborated chance hypothesis, an hypothesis according to which my observational beliefs which fit into my long-run coherent doxastic system are produced merely by chance, BonJour writes:

[There is an] internal tension or probabilistic incompatibility between (a) the claim that the observational beliefs are produced by chance, and (b) the claim that they continue even in the long run to satisfy the complicated and demanding pattern required in order to be coherence-conducive. If these two claims are considered separately, the latter is unlikely relative to the former, and hence also the former relative to evidence constituted by the latter; whereas if they are combined into one hypothesis [an elaborated chance hypothesis], this same incompatibility, now internal to the hypothesis itself, makes that hypothesis unlikely to be true on a purely *a priori* or intrinsic basis.²¹

BonJour thinks that a parallel argument can be offered for the conclusion that an elaborated demon hypothesis will have a low *a priori* probability.²² If we are to take this suggestion seriously, we should assume that the elaborated demon hypothesis is more or less equivalent to the conjunction of the simple demon hypothesis and the claim that the demon has certain specific intentions which guarantee the long-run coherence of my

²¹ Ibid., pp. 182-3.

²² Ibid., pp. 184-5.

system of beliefs.²³ If this is the case, the antecedent probability of the elaborated demon hypothesis should be equivalent to the probability that a simple demon hypothesis is true times the probability that the demon has the relevant intentions to guarantee that my doxastic system, complete with putative observational input, is long-run coherent. However, according to Bonjour, “[t]he unlikelihood that a demon would have just such desires and purposes (and that these would not change) seems no less great than the unlikelihood that an unspecified demon would produce just such observations.”²⁴ So, the very same tension that makes the probability of long-run coherence given a chance or simple demon hypothesis quite low will affect the antecedent probability of an elaborated demon hypothesis. If Bonjour’s reasoning is correct, I can know *a priori* that the probability of any skeptical hypothesis on the evidence of the long-run coherence of my doxastic system is very low.²⁵

Unfortunately, there seems to me to be a problem with the reasoning. Bonjour thinks that the simple demon hypothesis is analogous to the chance hypothesis in that an

²³ Ibid., p. 183. “[I]t is important to distinguish between [the simple and elaborated demon hypotheses].... The first postulates merely that there is an all-powerful evil demon who causes my experience...the second form postulates in addition, as a part of the explanatory hypothesis itself, that the demon has certain specific desires...in virtue of which he will single-mindedly continue to produce in me, even in the long run, coherence-conducive observations.”

²⁴ Ibid., p. 185.

²⁵ One worry at this point, which was suggested to me by Robert Coburn and which I won’t pursue here, is that Bonjour has not considered skeptical hypotheses of the Kantian variety, according to which the way the world is *an sich* transcends our abilities to conceive it. For more on Kantian skeptical hypotheses, see Chapter Two, part 2, below.

unspecified demon is equally likely to produce virtually any configuration of beliefs.²⁶

However, this is not as obvious as Bonjour would seem to suggest. One of the things that seems to cloud the issue is a certain ambiguity concerning the simple demon hypothesis.

In distinguishing the simple and elaborated demon hypotheses Bonjour writes:

The [simple demon hypothesis] postulates merely that there is an all-powerful evil demon who causes my experience, that is, my cognitively spontaneous beliefs, without saying anything more about the demon's motives and purposes or about what sorts of beliefs he is inclined to produce; whereas the [elaborated demon hypothesis] postulates in addition, as a part of the explanatory hypothesis itself, that the demon has certain specific desires, purposes, and so on, in virtue of which he will single-mindedly continue to produce in me, even in the long run, coherence-conducive observations.²⁷

It seems to me that there are two ways of understanding this distinction. On one way of understanding the simple demon hypothesis, the demon is held to be causing my experiences to fulfill some rational end, but his specific desires and intentions are left unspecified. The elaborated demon hypothesis then adds a specification of his intentions which would more or less guarantee the coherence of my doxastic system. On the other way of understanding the simple demon hypothesis, the demon need not be causing my experience intentionally or to fulfill any rational purpose, he need not even be aware that he is doing so. On this interpretation, the elaborated demon hypothesis adds to the simple demon hypothesis both the claim that the demon is intentionally producing my experience

²⁶ See *The Structure of Empirical Knowledge*, pp. 183-4.

²⁷ *Ibid.*, p. 183.

to fulfill some rational end and some claim about what his specific desires and intentions are.

If we understand the simple demon hypothesis to postulate that my cognitively spontaneous beliefs are caused by a demon in order to fulfill some unspecified rational purpose, the fact that my spontaneous beliefs are caused by a rational agent on the simple demon hypothesis but not on the chance hypothesis would seem to be an important disanalogy between the two. To be sure, the demon, much like chance, would be capable of producing in me virtually any configuration of beliefs. Even so, it seems at least plausible to suppose that the demon, being a rational creature himself, would be more likely to produce belief systems that are fairly coherent, at least in the short-run. After all, rational creatures may not tend to tell stories that are true, but they do tend to tell stories that make sense. And in any case, it seems somewhat unlikely that a demon would have a desire to produce in me a wildly arbitrary and gerrymandered set of cognitively spontaneous beliefs. While not obviously the case, it does seem plausible to suppose that the *a priori* probability that an unspecified demon would produce a coherent doxastic system is significantly higher than the probability that such a system would be produced by chance. And, if this is the case, it is not at all clear that there is a significant internal tension in the elaborated demon hypothesis which renders its *a priori* probability very low.

If we understand the simple demon hypothesis in the second way, that hypothesis seems to be little more than a fancy version of the chance hypothesis. Although my cognitively spontaneous beliefs are held to be caused by a demon on this understanding of

the hypothesis, the fact that he is a demon and hence a rational creature plays no role in the explanation. The probability that a demon (thought of as a purely neutral producer of beliefs) would produce in me a coherent doxastic system is quite plausibly low. However, it is not at all clear that when calculating the antecedent probability of the elaborated demon hypothesis that hypothesis should be construed as a conjunction of the simple demon hypothesis interpreted as little more than a fancy chance hypothesis and some story about the demon's desires that results in the long-run coherence of my system of beliefs. It would be clearly wrong to view the antecedent probability of the elaborated demon hypothesis as the probability of the conjunction of the *chance* hypothesis and the truth of the claim that the demon has certain specific intentions to delude me, intentions which he is capable of fulfilling if reference to his intentions is to be at all relevant to the probability of the long-run coherence of my doxastic system (i.e., the beliefs are not produced by chance at all). But, that probability should be zero since it is tantamount to the probability that a contradiction is true (that my beliefs both are and are not produced by chance). Similarly, it is not obviously appropriate to hold that the elaborated demon hypothesis is a conjunction of the simple demon hypothesis (understood as little more than a fancy version of the chance hypothesis) and the story about the demon's intentions.

To the extent that we do view the elaborated demon hypothesis as the conjunction of the simple demon hypothesis and the claim about the demon's intentions, we cannot treat the simple demon hypothesis simply as a version of the chance hypothesis. The fact that the demon is a rational creature must play some role, even if an unspecified one, in the

explanation of the production of my beliefs. And, in such a case, we have no reason to suppose that the *a priori* probability of the simple demon hypothesis on the evidence of the coherence of my doxastic system will be fairly low.

So, it is far from clear that we can agree with Bonjour that we can know *a priori* that the probabilities of both the simple and elaborated demon hypotheses on the evidence of the coherence of my doxastic system will be very low. But there is a further problem; I will be justified in accepting the second premise of the justificatory argument, *only if* I can tell *a priori* that the probability of the correspondence hypothesis on the evidence of coherence is significantly higher than it is on any of the competing hypotheses. Both the correspondence hypothesis and elaborated demon hypothesis make it highly probable that my doxastic system would be coherent over the long-run. So, in order to be justified in accepting the second premise of the justificatory argument, I need to be able to tell *a priori* that the antecedent probability of the correspondence hypothesis is higher than that of the elaborated demon hypothesis. Bonjour's argument only shows that the antecedent probability of the elaborated demon hypothesis will be lower than that of the simple demon hypothesis (whatever the antecedent probability of the simple demon hypothesis, assuming that it is non-zero, the antecedent probability of the elaborated demon hypothesis will be even lower since the relative improbability of coherence given a simple demon hypothesis gets transferred to the antecedent probability of the elaborated demon hypothesis). Bonjour has said nothing in support of the claim that the antecedent probability of the elaborated demon hypothesis is *a priori* lower than that of the

correspondence hypothesis.²⁸ Had he succeeded in establishing the relatively low probability that my doxastic system is coherent given a simple demon hypothesis (and hence, the relatively low *a priori* probability of the elaborated demon hypothesis), this might not have been a difficulty. But, as I argued above, the comparatively low *a priori* probability of the elaborated demon hypothesis was not convincingly established.

Plantinga has raised a related objection to Bonjour's analysis of the relative *a priori* probabilities of the skeptical and correspondence hypotheses on the evidence of the coherence of my doxastic system. Plantinga questions our ability to determine *a priori* the antecedent probabilities of the two hypotheses.

...Bonjour's argument requires not merely that propositions do have *a priori* probabilities; it also requires the premise that the Correspondence Hypothesis, for a given coherent structure S of beliefs, be more probable, *a priori*, than any of the skeptical explanations of the coherence of S...; and of course that too must be knowable *a priori*. This is monumentally dubious. Even if such a hypothesis as [the correspondence hypothesis] and these skeptical explanations do have an *a priori* probability, a probability on necessary truths alone, it is surely anyone's guess what that probability might be. Assuming there is such a thing as *a priori* probability, what would be the *a priori* probability of our having been created by a good God who (all else being equal) would not deceive us? What would be the *a priori* probability of our having been created by an evil demon who delights in deception? And which, if either, would have the greater *a priori*

²⁸ Though to be sure, Bonjour does address the worry that the correspondence hypothesis "involves a considerable internal complexity of its own, and some reason accordingly needs to be given for thinking that this complexity does not render it for basically parallel reasons just as unlikely, or even more unlikely, to be true than are the demon hypotheses....The principle [sic] point at which the correspondence hypothesis seems to be vulnerable...is in its assertion that the cognitively spontaneous beliefs which are claimed within the system to be reliable are systematically caused by the kinds of external situations which they assert to obtain....It is unlikely, relative to all the possible ways in which beliefs could be caused by the world, that they would be caused in the specific way required by the correspondence hypothesis"(ibid., p. 185).

probability? Short of being able to argue that God exists necessarily (in which case the first would have a probability of 1), how could we possibly tell?²⁹

Even if such philosophical hypotheses do have probabilities on necessary truths alone, nothing at all guarantees that we would be able to know *a priori* all of the necessary truths relevant to the *a priori* probability of any given hypothesis. If we cannot know *a priori* what the antecedent probabilities of the two hypotheses are, then we will not be able to know on purely *a priori* grounds the probabilities of either on the evidence of the long run coherence of my structure of beliefs. Thus, the second premise of the justificatory argument will itself remain unjustified. Plantinga's objection turns, at least in part, on the idea that our having *a priori* access to the relevant antecedent probabilities of the hypotheses in question is itself "monumentally dubious."

While Plantinga's objection rides almost completely on this intuition, the intuition is, I think, quite strong. To be sure, Bonjour *does* seem to have motivated the claim that we can know *a priori* something of a philosophical hypothesis's antecedent probability *relative to other hypotheses that are similar or that vary in various significant respects*. For example, an hypothesis, H1, might be less probable relative to another, H2, given its internal complexity as we saw with the elaborated demon hypotheses. However, this will not show that H1 is antecedently less probable than H2 *simpliciter*. There might be other considerations relative to which H2 will be seen to be *a priori* less probable than H1. For this reason, Plantinga's claim that it is quite unlikely that we would ever be able to know

²⁹ Warrant: *The Current Debate*, p. 109.

on purely *a priori* grounds all of the relevant considerations which go into determining the antecedent probability of a given philosophical hypothesis seems quite plausible.³⁰

In order to justify the claim that long-run coherence is truth-conducive, I need to know *a priori* that the correspondence hypothesis is more likely to be true given the evidence of the coherence of my system of beliefs than any of the competing skeptical hypotheses. But, if I can't know *a priori* all of the necessary truths which determine even the relative size of the antecedent probabilities of the relevant hypotheses, I will not be able to know *a priori* the relative probabilities of the hypotheses given the evidence of the coherence of my doxastic system. Thus, it is highly unlikely that the second premise of a justificatory argument can be adequately justified for a given doxastic agent on purely *a priori* grounds. Bonjour's coherentism is not able to resolve the tension that exists between internalist theories of justification and metaphysical and semantic realism.

Let us now turn to a different sort of epistemological theory to which a metaphysical realist might appeal in a heroic effort to avoid skepticism. In the previous

³⁰ Plantinga is right to point to the relation between elaborated demon hypotheses and theistic hypotheses. If we can know *a priori* the relevant necessities involved in determining the antecedent probability of a demon hypothesis, then we should be able to know *a priori* the relevant necessities involved in determining the antecedent probability of a theistic hypothesis. But we do not seem to be able to determine *a priori* the probability that an omniscient, omnipotent, omnibenevolent being exists. If we could, there would be many more (or perhaps many fewer) takers on Pascal's wager. This fact, I would submit, gives us good reason to believe that we do not have *a priori* access to the antecedent probabilities of either the correspondence hypothesis or any of the skeptical hypotheses.

discussion of Bonjour's coherentist epistemology, we saw that one of the problems for the view is that it requires an agent to be in implicit cognitive possession of a justificatory argument in order for any of his empirical beliefs to be justified. Perhaps the metaphysical realist will have more success combating skepticism if he instead adopts an externalist theory of justification and knowledge. According to an externalist theory of justification, justification is a matter of the obtaining of certain states of affairs that are external to the agent's doxastic system. The agent need not be able to "tell" from inside his doxastic system that the relevant states of affairs obtain for his beliefs to be justified; it is sufficient simply that they do obtain. Although there are many varieties of externalist theories of justification, one can get a sense for the general "flavor" of externalism by examining a typical externalist theory, reliablism.

Very simply, reliablism is the view that a belief is justified if and only if it is produced by a reliable cognitive mechanism. A cognitive mechanism, such as introspection, sense perception, or memory, is reliable if it tends to produce a preponderance of true beliefs.³¹ Now, ideally our cognitive and other sciences would be able to provide a naturalistic account of the reliable functioning of introspection, sense perception, memory, etc., on the basis of which an agent could in principle be informed that his cognitive mechanisms were reliably functioning at any given time. It is for this or similar reasons that many, though not all, externalist epistemologies are also called

³¹ On most reliabilist theories, the cognitive mechanism need not be completely reliable, or infallible. It is enough that the cognitive mechanism produces true beliefs a sufficiently large percentage of the time, or in a significant number of situations.

naturalized epistemologies. However, it is important to remember that an agent need not have any kind of conscious access to the fact that his cognitive mechanisms are reliably functioning in order for his beliefs that are the products of those mechanisms to be justified on a reliabilist account of justification.

Knowledge of the mind-independent world looks relatively easy if the reliabilist's theory of justification is correct. If some of the agent's cognitive mechanisms are reliable indicators of what is happening in the mind-independent world, then the agent's beliefs about the mind-independent world will be justified, and if his beliefs are, in addition, true they will be known to be so. The agent need not be in any position to know or even believe that the relevant mechanisms are reliable. If they are reliable, the agent can, and in many cases does, have knowledge of the mind-independent world. Although much more can be said about reliabilist theories in particular and externalist theories more generally, this brief sketch of the reliabilist's account of knowledge of a mind-independent world should suffice for our purposes. It should be clear that if an externalist account of justification is acceptable, the main skeptical worry associated with metaphysical and semantic realism concerning the very possibility of knowledge of a mind-independent world can be, at least *prima facie*, addressed.

But is the adoption of an externalist theory of justification acceptable? One forceful objection to an externalist account of the justificatory criteria for empirical knowledge of a mind-independent world can be found in a fairly recent article by

BonJour.³² Although the article is explicitly aimed only at naturalist epistemologies and not at epistemic externalism in general, it seems to me that it is the naturalist's commitment to an externalist theory of justification and the having of good epistemic reasons that is in many ways the main cause for concern.³³ While I agree with the thrust of BonJour's argument, I believe that it opens itself up to charges of begging the question against epistemic externalism, and hence naturalism, in that it does not make explicit that it is a naturalized epistemology *when held in conjunction with metaphysical and semantic realism*, rather than naturalized epistemology simpliciter, that is problematic.³⁴ In what follows I will briefly sketch BonJour's argument as well as the reply to this argument offered by Hilary Kornblith on behalf of the epistemic naturalist in his unpublished

³² "Against Naturalized Epistemology," *Midwest Studies in Philosophy*, 19, *Philosophical Naturalism*, Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, eds. (Minneapolis: University of Minnesota Press, 1994), pp. 283-300.

³³ In a footnote BonJour writes, "I have formulated the argument in terms of a reason for thinking that the belief is true, rather than in terms of the belief's being (epistemically) justified, because I do not want to enter here into the somewhat vexed controversy between externalist and internalist conceptions of justification. My view is that the result arrived at in the text is enough to constitute epistemological disaster whether or not the beliefs in question may be said to be justified in some other sense of justification that does not involve our having a reason to think that they are true" ("Against Naturalized Epistemology," p. 300). However, it seems to me that the externalist will want to claim that his sense of justification does involve our having a reason to think that they are true even though this reason is not one that need be internally accessible. Just as there are internalist and externalist senses of justification, there are internalist and externalist senses of having a good reason to think that something is true. It seems to me that the argument BonJour raises for the naturalist is not as independent of the vexed controversy between internalists and externalists as he would have it.

³⁴ BonJour seems to take the truth of metaphysical and semantic realism for granted in this objection and does not even consider the possibility that there might be reason for adopting an externalist epistemology if metaphysical antirealism is true.

manuscript, "Naturalistic Epistemology and its Critics."³⁵ I will then argue that a Kornblithian response is not adequate if we are to interpret the view as providing the metaphysical realist with a way of avoiding skepticism.

BonJour objects to a naturalized approach to an account of empirical knowledge by focusing on what he believes to be inherent problems for naturalized epistemology itself. Naturalized epistemologists reject the view that certain beliefs are justified by being knowable *a priori*.³⁶ Instead, naturalists generally hold the externalist view discussed above that a belief is justified if it is, e.g., produced by a reliable perceptual or other psychological mechanism.³⁷ BonJour argues³⁸ that, if the naturalist is serious about rejecting all forms of *a priori* justification, the price is a fairly robust skepticism. I will now briefly paraphrase his argument.

Granting for the moment that we have reason for thinking beliefs produced by a reliable mechanism are true, let us consider a non-observational belief. What reasons could we have for thinking such a non-observational belief to be true?³⁹ Any reasons

³⁵ Hilary Kornblith, "Naturalized Epistemology and its Critics," TD, 1994.

³⁶ Motivations for the rejection of *a priori* justification range from specific worries about the notion of *a priori* itself (see Quine's "Two Dogmas of Empiricism," *From a Logical Point of View* (Cambridge: Harvard University Press, 1953), pp. 20-46) to what appears to be a general frustration with a seeming failure of traditional *a priori* approaches to make answering the Cartesian skeptic possible.

³⁷ To be sure, if we have a generally reliable *a priori* mechanism, *a priori* beliefs will be justified even on a reliabilist account of justification. The important difference is that such beliefs are not justified simply in virtue of being *a priori*. Rather, they are justified because the *a priori* faculty is generally *reliable*.

³⁸ "Against Naturalized Epistemology," p. 296.

³⁹ The reliabilist will probably want to object right here that his view does not require the cognizing agent to be in possession of any reasons for thinking a belief that P is true for

offered will either be entirely independent of direct observation, in which case the reason is, contra the naturalist, *a priori*, or will involve an inference from other directly observational beliefs. According to Bonjour, for a reason of the second sort to be a cogent reason, we must have good reasons for believing that a conditional whose antecedent is a conjunction of directly observational propositions corresponding to the directly observational beliefs, and whose consequent is the non-observational belief in question is true. But any reasons we have for accepting this conditional could only be *a priori* given that all relevant observational propositions are already included in the antecedent of the conditional. If this argument is cogent we may conclude with Bonjour that “if, as the naturalist claims, there are no *a priori* reasons for thinking anything to be true..., the inevitable result is that we have no reason for thinking that any of our beliefs whose content transcends direct observation is true.”⁴⁰ Unfortunately for the naturalist, beliefs whose content transcends direct observations include not only beliefs about the remote past, the future, and present states of affairs wherein no observer is present, but also most of the beliefs accepted by the theoretical sciences, beliefs to which the naturalist will want to appeal in his account of the psychological and perceptual processes of human cognizing agents. Bonjour concludes that naturalized epistemology is “self-referentially

the agent to be justified in believing that P. I will turn to such a reliabilist objection after I have finished my explication of Bonjour’s argument.

⁴⁰ “Against Naturalized Epistemology,” p. 296.

inconsistent” in that “its own epistemological claims exclude the possibility of there being any cogent reason for thinking that those claims are true.”⁴¹

Despite Bonjour’s claim that the above argument seems as “obvious and compelling as any in the whole of philosophy,” Kornblith remains unmoved, claiming that a central premise in the argument effectively begs the question against naturalized epistemologies.

BonJour explicitly assumes that if an observation statement O is to provide good reason for believing some further claim T, then the conditional ‘if O, then T’ must itself be a claim which the agent has good reason to believe. But naturalists would deny this premise. Naturalists believe that human beings are so provided by nature that they are inclined to make certain kinds of inferences which are in fact reliable, long before they have evidence that those inferences are reliable. On the naturalistic account, such inferences constitute cases of good reasoning....BonJour is simply taking for granted certain constraints on good reasoning which the naturalist rejects. So there is a substantive dispute here as to what good reasoning consists in, not as BonJour portrays things, a simple case of self-referential inconsistency.⁴²

As I mentioned earlier, the naturalist typically subscribes to a reliabilist theory of justification. A belief T will be justified if it is caused in the right way by a reliable mechanism whether or not the agent is aware that it is so caused, or has any argument in his possession whose conclusion is T. That is, an agent S can be justified in his belief that T without himself being in possession of any evidence that supports T, so long as T was in fact caused by a reliable mechanism. Kornblith is correct to point out that on a naturalist view the particular agent S who has the belief that T on the evidence O does not need to

⁴¹ Ibid.

⁴² Kornblith 1994, pp. 29-31.

be in cognitive possession of good reasons for thinking that the conditional ‘if O, then T’ is true in order for O to provide good reason (in the naturalist’s sense) for S to believe that T; the mere fact that O obtains and is causally related to the belief in the right way is sufficient. If BonJour’s argument does turn on a premise which requires that S be in cognitive possession of good reasons for accepting the conditional in order for S to have good reason to believe that T, it clearly begs the question against naturalized epistemology.

However, it is not clear that BonJour’s argument does turn on this question-begging assumption. BonJour need not deny the naturalist’s presumption, that the agent S can be justified (or have good reason in the naturalist’s sense) in believing that T on the basis of O without S’s knowing that the inference from O to T is reliable, for his objection to have force. Indeed, his argument, as opposed to Kornblith’s reconstruction of it, makes no reference to any particular agent at all. The naturalist (Kornblith included) generally assumes that we live in a world populated in part by discrete three dimensional objects of various natural kinds, and further that we have reliable psychological and perceptual mechanisms that are sensitive to such objects.⁴³ Given this assumption, it makes sense for the naturalist to claim that S has good reason to believe that T on the basis of O because his belief that T was in fact produced by such a reliable mechanism even if S is not aware of this fact (and indeed, even if S is not aware of any reasons at all for his belief that T).

⁴³ Indeed, the naturalist doesn’t merely assume this; he thinks that he *knows* this. And, if a naturalized epistemology is correct, and if this belief is the result of the proper functioning of a reliable mechanism, he does know it.

All this is well understood by Bonjour. But what makes the naturalist's account plausible as an epistemological theory is that *we* can tell ourselves stories about how S was caused to believe T by some state of affairs in the world, a story that makes reference to the standard environment that agents like S find themselves in and some reliable perceptual mechanisms or other. While it is not necessary for S to possess reasons himself for believing that T is true to be justified in his belief that T, what makes the naturalist's account at all plausible as a philosophical theory is that *we*, as scientists and epistemologists, can be in cognitive possession of epistemic reasons for thinking that S's belief that T is true.⁴⁴ Such reasons may not be required for S's belief to count as *knowledge* if the naturalist's account is correct, but surely are required for the acceptance of naturalism as a viable philosophical theory. Bonjour's objection to naturalism is perhaps best viewed as the claim that the naturalist can provide us with no *philosophically acceptable* reasons for believing that the naturalist's basic assumption, that we live in a

⁴⁴ To be sure, this is not a requirement for knowledge on a naturalistic epistemology. Third-person assessment need not be possible for an agent to have knowledge on such a view. However, it does seem to me that such third-person assessment must be possible in at least some cases for an externalist epistemology to be a viable epistemic alternative to the more traditional foundationalist and coherentist epistemologies. To see why, consider a version of externalism according to which my beliefs are produced in me by the God of traditional theism. Such beliefs would be reliable, since such a God would not deceive me, and so, if true, would constitute knowledge. However, if a proponent of such an externalist view could provide us with no compelling reasons to accept that there was a God, it is hard to see that we would have any reason to view this epistemology as a philosophically viable alternative to a more traditional internalist one. Similar reasoning should call into question the philosophical viability of a more standard version of reliabilism. This fact is often overlooked since the assumption that my sensory and other cognitive mechanisms are reliable seems much less problematic than the assumption that God exists.

world populated by discrete three dimensional objects of various kinds with which our psychological and perceptual mechanisms dovetail, is true. And so, the naturalist can give us no philosophically acceptable reasons for thinking that naturalism is an epistemological theory that enables us to defeat skepticism.

The naturalist will probably still want to object along the lines indicated by Kornblith that even if *no one* can provide such “philosophically acceptable” reasons for thinking that our basic world-view is correct and that we have at least some reliable access to the world, this does not entail that we are not *epistemically* justified in making such assumptions. To rule this out of hand is again to beg the question against naturalism.

Kornblith points out that:

[C]hildren and animals may reason quite well without having the evidence, or in some cases even the conceptual repertoire, which would license those inferences. This is not to say that such evidence is inevitably and for all time out of their reach. Non-human animals are unlikely to be able to assess the reliability of their own inferences. But children do grow up, and when they are in a position to raise the question of the reliability of their own inferences, they are also in a position to gather evidence on the issue, and, in some cases resolve it.⁴⁵

However, having said this, Kornblith is free to point out that there are possible worlds otherwise exactly like the actual world in which children never grow up, or, having grown up, never raise the question of the reliability of their own inferences, or in which there is only the more basic reasoning of non-human animals. Surely if a naturalist account is correct, and non-human animals or children have justified beliefs (or engage in good

⁴⁵ Kornblith 1994, p. 30.

reasoning of any sort) in the actual world where grownup scientists and epistemologists can in fact provide reasons for thinking these beliefs are likely to be true by pointing to the make-up of the world and the various reliable perceptual mechanisms employed, non-human animals and children would be justified in these other possible worlds as well. After all, on a naturalist view, it is the make-up of the world and the reliability of the perceptual and other cognitive mechanisms which accounts for the justification of the beliefs, and not the story told by the scientists and epistemologists. Thus, no one, not even scientists and philosophers, need be in cognitive possession of reasons for holding the belief in question for the belief to be justified given a naturalistic epistemology. The naturalist will claim that if our world-view is largely correct, and if we have reliable access to the world, our belief that our world-view is correct and that we have reliable access to the world will be epistemically justified.

Such a reply is, I think, sufficient to show that Bonjour's claim that naturalized epistemology is "self-referentially inconsistent" is incorrect. However, it is far from clear that such a reply adequately addresses Bonjour's main worry that a naturalized epistemology will be of no help to the metaphysical realist in avoiding skepticism. To see why the reply is unsatisfying, let us consider two metaphysically possible worlds on the metaphysical realist's view: (a) one a mind-independent world populated in part by discrete three-dimensional objects of various kinds (much like the ones we believe to exist in our world), including various creatures whose psychological and perceptual mechanisms are sensitive to these natural kinds, but wherein the average human life-span is no more

than nine years and wherein no developed scientific view of the world exists, and (b) another world inhabited entirely by immaterial Cartesian souls who erroneously believe that they inhabit a world very much like the first world. Suppose that in both worlds the inhabitants who believe themselves to be human also have the further belief that W, i.e., that they live in a mind-independent world populated by three-dimensional objects of various kinds that more or less corresponds to the objects that they ostensibly perceive.⁴⁶ In neither world is any agent in *conscious* possession of good reasons for his belief that W, but this is not relevant, given a naturalistic account of knowledge. According to the naturalist, humans in the first world have good reason for their belief that W and are justified in believing that W because, let us suppose, this belief is caused in them in the right way by the functioning of a reliable cognitive mechanism. Because it is also true that W in this world, the inhabitants can be said to know that W. But in the second world, the “human” inhabitants are not justified in believing that W because their belief that W has not been caused in them in the right way by a reliable mechanism; it has, rather, been caused by some unfortunately unreliable aspect of their psychological make up.⁴⁷ The difficulty is that even in the world where the agents do have knowledge of a mind-independent reality on a naturalistic account, it is impossible for any agent to tell *from the*

⁴⁶ Although this may not be an explicitly formulated belief for any of the agents in question, we may assume for the purposes of this example that it is at least an operating background assumption.

⁴⁷ Perhaps the thought of being a non-embodied Cartesian soul in a completely non-physical universe is so horrifying that the inhabitants of the second world have been equipped by a merciful god with a psychological make-up that disposes them to the

inside that he lives in a world where knowledge of mind-independent reality is even possible. If some agent in one of these worlds were to begin to doubt that W was correct⁴⁸ or came to believe that W was false, nothing short of a viable *internalist* epistemological theory would ever convince him that he was justified in believing that W. To tell him that his belief is justified because caused by a reliable belief-forming mechanism in no way provides *him* with any reason to believe that he has such a reliable mechanism.

If metaphysical realism is true, and my belief that there is a tree in front of me that exists and has the features that it does independently of my (or anyone else's) mind is reliably caused by the state of affairs that it depicts (that is, by a mind-independent tree located in front of me), then on a naturalized or externalist epistemology I know that there is, in the mind-independent world, a tree in front of me. I also know that I know that there is, in the mind-independent world, a tree in front of me, and know that I know that I know..., etc., provided that these further beliefs are also caused in the right way. Unfortunately, if I begin to doubt, despite the observational evidence, that there is, mind-independently, a tree in front of me, it will not help to tell me that *if* a naturalized epistemology is the correct epistemology and my belief is produced by a reliable cognitive mechanism, then I can know that there is a tree in front of me, and know that I know, etc.

delusion that they live in a physical world with recognizable natural kinds and allows them to avoid the torture of insanity.

⁴⁸ Never mind that it is highly unlikely that a nine-year-old would ever have such sophisticated philosophical worries.

Nothing short of a philosophical argument which provides me with reasons, internalist reasons, for thinking that the antecedent of the conditional is true (and hence that our commonsense and scientific world views are largely correct), will shake me from my skepticism.⁴⁹ It is for this reason that it is seemingly impossible for naturalized epistemologies to provide philosophically acceptable reasons for thinking that we do, in fact, have knowledge of the mind-independent world. For someone who is truly worried about the possibility of knowledge of a mind-independent world, an externalist theory of justification will provide him little more comfort than glibly assuring him that if he has knowledge of the mind-independent world, then he has knowledge of the mind-independent world—and knowledge that he has knowledge..., etc.

It may seem somewhat unfair to require that the metaphysical realist who would favor a naturalistic approach to epistemology produce reasons for thinking that we do, in fact, have knowledge of the reliability of our cognitive mechanisms. After all, a naturalist will not generally attempt to argue that we have knowledge of any kind; that there is genuine human knowledge is a basic presupposition of most, if not all, naturalized epistemologies. Kornblith himself explicitly states his commitment to the presupposition that there are genuine cases of human knowing:

On my view, knowledge is a natural phenomenon, and it is this natural phenomenon that is the subject matter of epistemology. Not the concept of

⁴⁹ To be sure, this problem is not unique to the conjunction of an externalist epistemology and metaphysical realism. It is just that, as we saw in Chapter Two, the same skeptical worry will not arise if certain forms of metaphysical antirealism are antecedently accepted as true. The skeptical worry that creates this problem seems only to get a foothold if one is a metaphysical realist.

knowledge, but knowledge itself....The phenomenon of knowledge is ubiquitous. It may be found in simple perceptual situations, where an epistemic agent confronts a table in good light and clear viewing conditions, as well as in the more complex interactions between agent and environment found in the scientific laboratory. I am assuming that there is a single phenomenon here to be studied, that these clear-cut cases of knowledge constitute a natural kind....I will thus assume that, at least pretheoretically, there is a robust phenomenon of human knowing, and that this phenomenon is susceptible to investigation.⁵⁰

However, it is precisely because of this presupposition that the naturalized epistemologies should not be seen as providing a response to the skeptic. The naturalized epistemologies, in presupposing that we do have empirical knowledge, simply *presupposes* exactly what was in question in the skeptical arguments. So, a naturalized epistemology cannot legitimately be used to respond to a skeptical argument whose conclusion is that we do not have empirical knowledge. The naturalist has a somewhat different project in mind. He is interested, not in refuting the skeptic, but rather in investigating the phenomenon of our *prima facie* knowledge of the natural kinds postulated by our pretheoretic world view and whose existence is confirmed by the various physical sciences.

While coherentist and externalist epistemologies are not the only candidates for the heroic epistemology that would answer the skeptical worries that the metaphysical realist faces, they are among the most plausible candidates to date. As we have seen, neither of them is successful in this capacity. This should give us ample reason to claim that the metaphysical realist has a problem dealing with the skeptical challenges that is, at least

⁵⁰ Kornblith 1994, pp. 12-14.

prima facie, considerably more serious than at least some kinds of metaphysical antirealism face.

Curriculum Vitae
Deborah Colleen Smith

Department of Philosophy
University of Washington
Seattle, WA 98195
(206) 543-5855

1815 N.E. 190th
Shoreline, WA 98155
(206) 365-7786
dcsmith@u.washington.edu

EDUCATION

Ph.D.	1996	University of Washington (philosophy)
M.A.	1992	University of Washington (philosophy)
B.A.	1990	Pomona College (philosophy)

Dissertation: "Metaphysical Realism and Antirealism: an Analysis of the Current Debate"

Committee: Robert Coburn (chair), Laurence Bonjour, S. Marc Cohen, Kenneth Clatterbaugh.

AREAS OF SPECIALIZATION

Metaphysics, Epistemology, Philosophy of Language

AREAS OF COMPETENCE

Ethics (theoretical and applied), Logic, British Empiricism, Wittgenstein

PROFESSIONAL ACTIVITY

"Naturalism and Physicalism," comment on "Meaning, Normativity, and Naturalism" by Eric Gampel, presented at the Pacific Division Meeting of the APA, 1996.

"Moral Realism, Skepticism and Anti-Realism: A Critical Analysis of the Criteria for Moral Realism," presented at the Pacific Division Meeting of the APA, 1995. *Graduate Student Award winner.*

"Van Inwagen, Metaphysical Antirealism, and the Height of Mount Everest: Can Linguistic Conventions Make a Mountain out of a Molehill?," presented at the Northwest Conference on Philosophy, 1995.

"Kantian Universalizability and Contradictions in Willing," comment on "Scrooge and the South Sea Islanders" by Warren Harbison, presented at the Northwest Conference on Philosophy, 1994.

"The Rights of Humans Versus the Rights of Animals," comment on "Should Feminists Be Vegetarians?" by Kathryn Paxton-George, presented at the Northwest Conference on Philosophy, 1991.