

**METASEMANTICS:
On the Limits of Semantic Theory**

T. Parent

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Philosophy

Chapel Hill
2009

Approved by:

William G. Lycan
(Director)

Dorit Bar-On

Thomas Hofweber

Ram Neta

Keith Simmons

© 2009
Ted Parent
ALL RIGHTS RESERVED

ABSTRACT

T. Parent

METASEMANTICS: On the Limits of Semantic Theory (Under the direction of William G. Lycan)

METASEMANTICS is a wake-up call for semantic theory: It reveals that some semantic questions *have no adequate answer*. (This is meant to be the *epistemic* point that certain semantic questions cannot be “settled”—not a metaphysical point about whether there is a fact-of-the-matter.) *METASEMANTICS* thus checks our default “optimism” that any well-formed semantic question can be settled (at least in principle).

Chapter One argues that relative to certain assumptions, a question like “What does ‘Pollux’ denote?” *has no adequate answer*. If an answer is to be non-circular, then any answer ultimately depends on an uninterpreted term—meaning that this term occurs absent an answer to what *it* denotes. This, I argue, makes the answer uninformative in certain crucial respects. The lesson here essentially vindicates Quine’s thesis of ontological relativity (though not his behaviorism or semantic nihilism).

Chapter Two and Three build on this “pessimism” in considering “ontic-idioms,” such as ‘exist’, ‘actual’, etc. If Chapter One entails there is no saying what an ontic-idiom’s *extension* is, these chapters show there is no saying what their *intension* is. Any attempt, I claim, will be equivocal. As corollaries, I show that a univocal statement of Realism about *x* is impossible—as well as a criterion of ontological commitment.

Chapter Four considers truth-conditional semantics, generally speaking. After elaborating Davidson’s claim about the “folly” of defining truth, I counter-balance his

pessimism by showing that an informative analysis of 'true' is still possible (though only for certain translational purposes).

Finally, Chapter Five evaluates a pessimistic argument concerning *mental* content. I argue that under externalism, *a priori* knowledge of content is impossible, at least for knowing whether a concept is about H₂O versus XYZ. But this limit on the *a priori* should be unsurprising; I argue, moreover, that for other purposes we indeed know *a priori* what we think.

*To Jay F. Rosenberg
(1942-2008)
Under whose rigorous tutelage
I greatly benefited.*

And

*To William G. Lycan
Without whom not.*

PREFACE

Jay Rosenberg, to whom half this work is dedicated, was infamous for his cranky, opinionated prefaces which greet the reader upon opening his books. In *The Thinking Self*, he starts with a defense of this practice:

A preface is about the only place left where a serious scholar can write with complete candor about what, *besides* serious scholarship, is on his mind and still be confident that it will appear in print. I see it as a sort of right which one has earned by virtue of having produced the sober and dignified scholarly work which follows. (p. vii)

Of course, I cannot be confident that the present work will go to press, at least in this form. Nor do I feel at liberty to speak with *complete* candor (though I'm sure Jay did). But Jay provides enough of an excuse to speak with more candor than is usual.

We lost Jay to esophageal cancer in February last year, and in this philosophy lost a loyal and brilliant devotee. It has been written, somewhere other than here, that Jay was the most under-appreciated philosopher in recent years. He certainly had an eye for detail like no one else. But Jay unfortunately had a knack for alienating himself from others, mostly due to his frank and plentiful criticism, which he *yelled unmercifully*, standing three feet away. I have no doubt his work would have received more attention if only he had been more personable.

I had the (mis?)fortune of taking a writing-intensive seminar with Jay for first-year graduate students. It was a trying experience. Indeed, Jay's meat-grinder mentoring was responsible for the immediate departure of two of my classmates. But the seminar rebuilt my

brain. His harsh yet penetrating criticism forced me to be highly sensitive to detail, much like Jay was (though I make no claim to be as perceptive as he). I am eternally grateful for his training in philosophical methodology, thus the dedication. On the matter, I am reminded of his own gratitude to Wilfrid Sellars (who apparently was also nasty pedagogue), expressed in the preface to *Linguistic Representation*,

[O]ne normally says something like ‘The debt that this book owes him is immense’. I would say it too, were it not to understate the case. From Wilfrid, I learned to *think*.
(p. xi)

I might say the same about Jay, and his attempt to teach me to *think* was an honor. But for the record, I do not think of myself a third-generation Sellarsian: I am not an inferentialist and I commit the heresy of ranking Quine as the better philosopher. Moreover, I remain unconvinced of the Jay-Sellars claim that good philosophy must be about “everything.” A specialized journal article on a focused topic is often very edifying. I agree, however, that philosophers should not isolate themselves to their specialties, since Jay is right when he says that philosophical problems often illuminate each other. Moreover, I share his lament for the disappearance of “systematic” philosophy. Although it is an ambitious undertaking, philosophical system-building is still possible in the current profession, over the course of a career, while writing in a focused and detail-oriented manner. Not to be flippant, but I think of Quine as a shining example here. (And I find disturbing the recent trend to reduce Quine’s significance to run-of-the-mill naturalism.)

Before he died, I emailed Jay to say that my thesis would be dedicated to him, and he replied graciously. I hope it is true that he did not cringe at the idea, especially since our views are somewhat congenial. Although they are less so these days, thanks to the influence of Bill Lycan.

Bill, of course, is who shares the dedication with Jay. When I think of the hours he has devoted to improving this document, I shudder. Bill is in fact a creature is thought to be mythical in the philosophy world (and probably in the academy generally): He is the dissertation advisor who reads every word with care *and* returns substantive feedback—all within 72 hours. He would occasionally shock me with an apology, in the rare case when he took longer than a week.

But most importantly, Bill deserves to be commended for his stamina. This is one of those dissertations that took entirely too long to write. But contra the usual explanations, this is not because I fell into a five-year drinking binge or became habitually lazy. The fact is that the topic I chose was hard, and I had an advisor who would not let me half-ass the job. (There are close to 50 drafts of Chapter One, not counting drafts with minor edits.) Bill is a paragon of intellectual integrity, and a less principled advisor would have approved my earlier, less rigorous work. He never wavered in his duty to train me, even though our philosophical intuitions and priorities were often at odds. If Jay taught me to think, it was Bill who taught me how to express my thinking properly to a present-day audience of academic philosophers. And in philosophy, expressing your thoughts properly is virtually as important as the thoughts themselves (especially since there is often no clear division between the two).

My debts are hardly limited to Bill and Jay however. Yet I chose to limit the dedication to them, for fear of diluting what little significance the gesture may have. But I would like to note the nearby worlds where I gesture thus to other parties.

In at least half of the nearby worlds, this work is dedicated to my parents—and in half of those worlds, my siblings are also mentioned. I wish I could properly express my appreciation for my family, but that sort of thing is utterly clichéd in a preface; so the task is

near impossible. Yet I will report unabashedly that each member of my family is quite superior to most human beings. My siblings—Brodie, Aimee, Denise, and Marc—each have an uncommon strength of character. I am proud to be related to them. Moreover, I am lucky to have parents that I not only love but also *like*. My father’s generosity is always beyond the call of duty, not only financially but also in his emotional support. This dissertation would not have been possible without him. As for my mother, she is one of my closest companions. She is moreover a true intellectual who has a genuine passion for meaningful reflection—unlike some philosophers I have met in the profession.

I love my family, but I have not dedicated this work to them, mostly because I find it a bit gauche. A dedication might imply that the final product is *for* them. But if I really were writing *for* my family, I would not write on formal semantics. As it is, I wrote this piece *for* Bill Lycan—and Jay gets mentioned because of his profound influence.

Other nearby worlds are ones where I dedicate this work to all the wonderful teachers I have had over the years. In still others, the dedication is to those who helped me intellectually, professionally, and emotionally during the writing of this thesis—including (but not limited to) Jim Baillie, Dorit Bar-On, Sarah Blythe, Jason Bowers, Matthew Chrisman, Patrick Connolly, Felipe De Brigard, Tamra Frei, David Frost, Don Garrett, Hell (the bar, now defunct), Thomas Hofweber, Bryce Huebner, Justin Jeffrey, Betsy Jelinek, Kenny Kupers, Norah Martin, Al Martinich, Ram Neta, the Open Eye Café, the Portland Brew (in Nashville), John Post, Michael Resnik, Dave Ripley, Geoff Sayre-McCord, Joshua Shaw, Keith Simmons, Jeffrey Tlumak, Piers Turner, and (last but certainly not least) Meg Wallace. I would also like to give special mention to the makers of Prozac, Celexa, Lexapro,

Wellbutrin, Effexor (my current favorite), as well as Michael Ryan, Ph.D. (If not for them, I would be dead twice over.)

I decided to write about semantic theory because it seemed to me that such theorizing these days is often done in a rather unreflective manner. It also struck me that some puzzles in semantics, e.g., substitutivity puzzles, could be resolved if we take into account the limitations of our semantic theorizing. (Roughly, these phenomena, I believe, are due to deficiencies in a speaker's "background language.") Unfortunately, I have been unable to develop these points in the present work, and mostly have had to rest content with describing the basic phenomena that would provide the basis of such insights. (Sometimes having a "grand vision" of the philosophical landscape only sets you up for a "grand" amount of work.) I might pursue these matters in the future, though perhaps not until tenure is safely behind me.

T. P.
Chapel Hill
May 2009

TABLE OF CONTENTS

LIST OF FIGURES.....	xii
Chapter:	
I. WHAT DOES THE MODEL-THEORETIC ARGUMENT SHOW?.....	1
Preliminary Sketch of the Argument.....	5
Nonstandard Interpretations (Part One).....	7
Nonstandard Interpretations (Part Two).....	10
Indeterminacy and the Lewis Objection.....	15
Another Model-Theoretic Argument.....	20
Pessimism about Semantics.....	25
Contrast with Cartesian and Quinean Scepticisms.....	42
Semantic Pessimism and Anti-Realism.....	45
The Significance of Pessimism.....	47
II. SEMANTIC PESSIMISM AND ONTIC-IDIOMS.....	49
Five-and-a-Half Ways to be Innocent.....	50
Defining the Commissive Use.....	54
Lewisian Proposals.....	58
Another Proposal from Lewis.....	60
The Significance of the Preceding.....	61

III. SEMANTIC PESSIMISM AND ONTOLOGY	63
The Problem and a Proposal.....	66
The Metaphysics of MIOs.....	70
Modality and Negative Existentials.....	73
Some Limits of Interpretation.....	76
Wittgensteinian Understanding and Primitive Ontic-Idioms.....	81
Ontological Commitment.....	83
Closing Remarks.....	86
IV. SEMANTIC PESSIMISM AND TRUTH-CONDITIONAL SEMANTICS.....	88
The Determination Argument.....	88
Tarski-Davidson Circles.....	94
A Proposal.....	98
Semantic Explanation.....	102
V. WHAT DOES THE SLOW-SWITCH ARGUMENT SHOW?	105
Externalism and Anti-Individualism.....	107
The Slow-Switch Argument.....	108
The Relevance Objection.....	110
The Objection from SVJs.....	113
Brown’s Reply.....	115
Cartesian Self-Knowledge and “Knowing What”.....	119
The Charitable Skeptic.....	124
BIBLIOGRAPHY.....	128

LIST OF FIGURES

Figure 1: Two Models for (14).....	31
Figure 2: (15) in Model A and Model B.....	35
Figure 3: Tarski-Davidson Circles.....	98

**FIRST CHAPTER:
WHAT DOES THE MODEL-THEORETIC ARGUMENT SHOW?**

Some men are so impressed by what science knows that they forget what it does not know; others are so much more interested in what it does not know than in what it does that they belittle its achievements. Those who think that science is everything become complacent and cocksure, and decry all interest in problems not having the circumscribed definiteness that is necessary for scientific treatment... On the other hand, those who pooh-pooh science revert, as a rule, to some ancient and pernicious superstition... Both these attitudes are to be deplored, and it is philosophy that shows the right attitude, by making clear at once the scope and limitations of scientific knowledge.

—Bertrand Russell, “Philosophy for Laymen”

Consider Anderson who shows her student Max a map of Iceland. As is usual, the map has dots and lines for the major cities and roads, respectively. However, the map lacks a legend explaining these marks, and there are no labels to indicate what the names of the cities and roads might be. Indeed, there are no words whatsoever on the map, so that it is not even stated that the map is a map of Iceland.

Suppose Anderson attempts to explain the unmarked map to Max by drawing a picture of Iceland—but like the map, the picture also lacks any labels or legend. Nevertheless, suppose Anderson tries to explain the map just by pointing to a dot on the map, and then pointing to the corresponding area in the picture: “See, this dot designates *this* locale.” Suppose she also tries to explain which island is depicted by the map, by demonstrating his entire drawing and simply stating: “See, the map depicts *this* island *here*.” Now clearly, Max won’t learn much from this. If Max cannot identify Iceland as represented on the map, then presumably he cannot identify Iceland by the picture. And so, Anderson will be unsuccessful in teaching Max what location the map depicts, what cities are found there, etc.

In Max's case, the problem is that he is trying to learn about one representation (the map) via another representation (the drawing) for which he had no prior understanding. This, I claim, is somewhat analogous to a semanticist who tries to gain a certain theoretical understanding of a term, by using some other term that is left theoretically untreated. In one kind of case, the semanticist understands a name like 'Anderson' by using the term 'Anderson' itself—as in the interpretation ' 'Anderson' denotes Anderson.' Or following Russell (1905), the name may be analyzed into a description such as 'the teacher with the unmarked map of Iceland'. But either way, the semanticist understands a representation only by some other (possibly type-identical) representation. And like Max, if there is no appropriate grasp of the latter representation, it may seem that we do not learn much from the semanticist. Of course, her analysis will indicate how one representation is semantically related to another. But as in the case of Max, one might feel that we have not been properly acquainted with *the object* that 'Anderson' denotes.

In the course of this chapter, I shall try to make this idea more precise. Essentially, I shall argue that certain fundamental questions about denotation *lack adequate answers*. Alternatively, the claim is that in one important sense a question like "What does 'Iceland' denote?" *cannot be answered*, even in principle. Of course, I think such a question *can* be answered in other, important senses. Thus, I do not intend to suggest there is no fact-of-the-matter as to what 'Iceland' denotes. My point here is more epistemic than metaphysical: The claim more precisely is that there is no sentence of the language which can serve as an *informationally adequate* answer, to certain questions about denotation. Or better, no sentence of the language will be informative *in the right*

way—much like the drawing of Iceland does not appropriately inform Max of what the unmarked map represents.

What drives this claim is something like the metaphor that we cannot “step outside” the language to see what our terms correspond to in the world. In particular, I think we can ask a question like ‘What does ‘Iceland’ denote?’ in a way that would (*per impossibile*) require us to “step outside” the language. However, I want to replace the metaphor here with something more rigorous, viz., with the so-called “model-theoretic” argument from Putnam (1976), (1977), and (1981). This argument, I want to suggest, does not show the *metaphysical* thesis of anti-realism, as Putnam claims, but rather the epistemic-informational thesis that I want to advocate. That is, the model-theoretic argument reveals a way of asking ‘What does ‘Iceland’ denote?’, which makes an informationally adequate answer impossible.

I am not the first to argue for such a thesis. Quine made this kind of claim in the second half of “Ontological Relativity,” also based on the model-theoretic argument (though in Quine it is better known as the “proxy-function” argument). The problem, however, was that Quine’s thesis about the unanswerability of semantic questions arose in a context where he assumed not only behaviorism, but also his semantic nihilism. And in such a context, it can be easy to dismiss Quine’s remarks. Yet I believe that one can use the same sort of argument, to the same effect, *without* such contentious assumptions. So in one sense, my aim is to resuscitate the second half of “Ontological Relativity,” for those of us who are neither behaviorists nor semantic nihilists.

To take another angle, the present chapter can be seen as limiting (what we may call) the “default stance” in science, or in semantics more precisely. Our default stance is

that any question which can be formulated in the language has an adequate answer, except perhaps cases of, e.g., presupposition failure (“Have you stopped beating your wife?”). Now I think the “optimistic” stance toward questions is normally the appropriate one to take in a field of inquiry. Unless we have special reason to think otherwise, it is productive to assume that any question in a field of inquiry has an answer. Even so, the present chapter aims to show that we *do* have special reason to reject this stance, when it comes to questions about denotation.

So in pointing to certain limits of semantic theorizing, my view could be called “pessimistic” as regards semantic theory. But I do not intend to belittle semantic theory as an enterprise, as Russell warned against in the opening quotation. In particular, I do not mean to advocate a kind of anti-scientism, or that semantics “rests on a mistake,” or that semantics is a failed “social construct” arising in a “Eurocentric paradigm.” Rather, I take the present chapter to be part of the scientific enterprise, and in particular, part of an inquiry into the epistemic condition of human beings.

Since the model-theoretic argument is the key element in all this, much of the chapter will be spent clarifying what the argument is, and what exactly it shows. I shall first give a careful and detailed presentation of (what I take to be) Putnam’s anti-realist version of the argument. Then, I shall argue that this version does not survive an objection from David Lewis (1984).¹ Next, I will consider a different, though less noted, version of model-theoretic argument in Putnam, which also attempts to debunk realism. This version also fails, we shall see. However, this second version serves as a starting point from which I will show that some semantic questions are indeed unanswerable in the sense discussed above.

¹ References to Lewis throughout will be in relation to this paper.

I. Preliminary Sketch of the Argument.

The model-theoretic argument, as I understand it, comes in three stages. The *first* stage is the distinctly “model-theoretic” portion. In Putnam (1976), (1977) this model-theoretic stage purports to show:

(C1) A true scientific theory has at least one nonstandard (or “unintended”) interpretation which preserves its truth.

However, in Putnam (1981), the first stage ends with the different claim that:

(C1.1) For each possible world, sentences in a theory can be given a nonstandard interpretation that conserves the truth-value of each sentence in that world.

The key difference here is that (C1.1) concerns interpretations of a theory in *every* possible world.

Either way, the *second* stage of the argument uses (C1) or (C1.1) to present a case against metaphysical realism. Putnam’s claim in particular is that, given the model-theoretic contentions, realism will entail the view that:

(C2) The reference (or denotation) of terms is indeterminate.

Here, terms like ‘determinate’ and ‘fixed’ express metaphysical notions, where x is determined or fixed by y just in case x is a (metaphysically) sufficient condition for y . Relevant substitutions for ‘ x ’ in the present case would include ‘a single interpretation of a term’ and ‘a term’s denoting an object o ,’ etc.

Yet for your typical realist, (C2) would be absurd—and so, Putnam takes the argument to show that realism is “incoherent” (1976, p. 124). So, as the third stage of the argument, Putnam then concludes his brand of anti-realism:

(C3) Sentences in an ideal scientific theory cannot be false.

Here, 'ideal' means "epistemically ideal" (or some such thing); an ideal theory is one that is empirically adequate, and maximizes simplicity, conservativeness, utility, etc.

But as Lewis also notes, it is dubious that (C3) separates realists from anti-realists. Indeed, it seems to me that even a Berkeleyan may count as a "realist" in Putnam's sense. For if the Berkeleyan holds that if the world consists of nothing but God's ideas, she might still believe that truth is determined by *facts* about God's ideas, and not by *what our best theory says* about those ideas.

However, the denial of (C3) at least is what Putnam sees as most relevant to the realist view which the model-theoretic argument attacks. Putnam (1976) writes "The most important consequence of metaphysical realism is that *truth* is supposed to be *radically non-epistemic*...the theory that is 'ideal'...*might be false*." (p. 125). Yet because of our Berkeleyan, (C3) apparently does not characterize anti-realism in general; rather, it strikes me as a precisification of a *pragmatist* viewpoint, of the sort William James (1907) advocated. If, as Putnam claims, an ideal scientific theory cannot be false, then truth seems to be something like "what works in the way of belief."

Still, this pragmatist view would be sufficient for the falsity of realism. For as Putnam states, the realist "assumes a theory-independent fact of the matter as to what a term in a given theory corresponds to" (1976, p. 136). However, if (C3) is right, then such theory-independent facts would not exist; the facts would be *dependent* on what the ideal theory says. So even if (C3) is not characteristically anti-realist, it would still undermine realism (perhaps *inter alia*).

Since my aim is to determine what exactly the model-theoretic reasoning shows, I shall first examine the case for (C1) and (C1.1) in the next two sections, respectively. It

turns out that Putnam's case for (C1) is lacking, as well as his argument for (C1.1) [though part of the former argument can be made good.] However, readers may turn to section IV if they wish to skip this technical material and get to more philosophical matters. From section IV on, I show that the model-theoretic argument (and one of its variants) does not show Anti-Realism, but rather that certain semantic questions lack informationally adequate answers.

II. Nonstandard Interpretations (Part One)

In Putnam (1976) and (1977), argument for (C1) seems to be based on the Löwenheim-Skolem theorem; this theorem states that if a theory has a model of whatever cardinality, then it also has a countable model.² Interestingly, then, the theorem guarantees the existence of *countable* models for set theory, even though the standard model for set theory is uncountable *and* contains elements which are themselves uncountable sets. [Skolem's (1920) "paradox"]. The sentences of set theory can be made true in a countable model, basically, because the predicates can be *reinterpreted* so that the sentences say something different from what they standardly say. For example, 'uncountable' can construed as "uncountable in the model," or more exactly, "not enumerated by any function in the model." In this manner, the sentence 'there exist uncountable sets' could be made true in the model if the model does not contain a one-one pairing between natural numbers and members of some set.

² I follow the standard habit of model theorists to use the term 'model' ambiguously, between models and *domains* of models. In truth, any model has a cardinality of *two*, since models are pairs with a domain and an interpretation-function. Yet talk of uncountable and countable models is meant with respect to the *domain* of a model. Model theory has done just fine with the ambiguity, presumably because disambiguation occurs in contexts of use. So under that assumption, I follow the standard usage.

Does (C1) follow from the Löwenheim-Skolem theorem? Observe that if we begin with a theory that has a countable model, then the theorem is utterly trivial. So in particular, the theorem does not demonstrate a model with a *nonstandard* interpretation in these cases. But that means *we cannot show (C1) from the Löwenheim-Skolem theorem.*

One may think, however, that (C1) holds if we are first given that a theory has an uncountable model. Yet even that does not follow. Consider a theory which names countably many objects, and whose quantificational claims are either trivial or satisfied just by the named objects. Then, if the model happens to contain an uncountable number of additional, unnamed objects, the theory will obviously have a countable and an uncountable model. Nonetheless, the language of the theory may have the *same interpretation* between these two models.

Thus: If we are merely given that a theory has an uncountable model, then for all that tells us, the model is uncountable in virtue of having uncountably many “extraneous” objects, such as the model described above. Accordingly, if we then use Löwenheim-Skolem to show the theory has a countable model, this does not show that the language of the theory has a *nonstandard* interpretation-function which preserves truth. For the uncountable model we started with may just be “overabundant” though it shares the same interpretation-function as the countable model.

A nonstandard interpretation is guaranteed by Löwenheim-Skolem only when a theory contains sentences *about* uncountably many objects, on the standard interpretation. That is why set theory is of special interest. But few of our scientific theories are like set theory; it is rare for us to stake claims about uncountably many

objects. So in fact, Löwenheim-Skolem guarantees nonstandard interpretations for very few, if any, of our scientific theories. That is a far cry from (C1).

In Putnam (1976), however, part of the argument for (C1) seems to be separable from considerations relating to Löwenheim-Skolem.³ Putnam writes:

Pick a model M of the same cardinality as THE WORLD. Map the individuals of M one-to-one into the pieces of THE WORLD. The result is a satisfaction relation SAT—a “correspondence” between the terms of [a language] L and sets of pieces of THE WORLD—such that the theory T₁ comes out *true*—true of THE WORLD—provided that we just interpret ‘true’ as TRUE(SAT). (p. 126)

The idea is first to map objects in model M to pieces of the world, via some relation G, which is (presumably) non-equivalent to identity.⁴ Then, pick a relation R between terms and the worldly pieces such that: a term *t* bears R to *y* iff [*t* denotes *x* in M, where *x* is in turn is mapped to *y* (by G)].⁵ Next, we use R to define the denotation-predicate so that *t* now denotes *x* instead of *y*. (N.B., in some but not all instances, *x* can be identical to *y*.) In this, we have the makings of a model with the set of *x*s as domain, and R as the interpretation function. And in this model, it will be correct to say the theory is satisfied by pieces in the world. But since at most one of these interpretations can be the standard interpretation, this shows the theory has a nonstandard interpretation which preserves its truth. This, I think, is a sound argument for (C1).⁶

Of course, semantic terms such as ‘denotes’ and ‘satisfies’ will mean something different than what we mean by these terms. Yet possible reinterpretations of ‘satisfies’

³ Putnam (1976) does not regard these considerations as separate (see his note 3). But my objections to each individually also applies to them when conjoined.

⁴ N.B., there will be such a relation even in uncountable models, since for finite *x* in a model, we can define a relation which maps these *x* to something other than *x*—and then map all other *y* to themselves.

⁵ Here I use ‘denotes’ so that a predicate denotes *members* of its extension, rather than the extension itself. For instance, I assume ‘dog’ denotes members of the set of dogs, not the *set* as such.

⁶ This argument, or something like it, seems to be what’s behind Lewis’ slogan “(almost) any world can satisfy (almost) any theory,” though Lewis also mentions Löwenheim-Skolem unnecessarily (p. 229-230).

is exactly what the argument exploits. Even if semantic terms are located in a different language, a metalanguage, both languages can subject to various interpretations.

To illustrate, consider the following simple, little theory:

- (1) Everest is a mountain
- (2) London is not a mountain.

Consider now the standard model for (1)-(2), where:

- (3) 'Everest' denotes Everest.
- (4) 'London' denotes London.
- (5) 'Mountain' denotes mountains.

Suppose now we map Everest to the number 3 and London to 4. From this, we can define a model M^* where the terms in (1)-(2) are interpreted as follows:

- (3.1) 'Everest' denotes 3.
- (4.1) 'London' denotes 4.
- (5.1) 'Mountain' denotes prime numbers.

And under this interpretation, (1) and (2) will still express truths, viz.:

- (1.1) 3 is prime.
- (2.1) 4 is not prime.

Thus, the theory has a nonstandard interpretation which preserves its truth. And basically, (C1) guarantees that this can happen with *any* scientific theory.

III. Nonstandard Interpretations (Part Two)

As I mentioned, the first stage of the model-theoretic argument is notably different in Putnam (1981). Here, the aim is to show the claim that:

- (C1.1) For each possible world, sentences in a theory can be given a nonstandard interpretation that conserves the truth-value of each sentence in that world.

Again, (C1.1) differs from (C1) in that the former explicitly concerns interpretations in any model whose objects are exactly those of some possible world (hereafter, any

“model corresponding to a world”).⁷ [Also, (C1.1) is formulated in terms of preserving truth *and* falsity. But since we assume bivalence, this difference is only superficial.]

Now in Putnam (1981), the argument for (C1.1) appeals to facts about *permutations* instead of the Löwenheim-Skolem theorem. (Reminder: a function is a permutation on a set S iff it is one-one into and onto S .) Let us first define a “nontrivial” predicate as one whose extension is neither empty nor universal in a given world.

Putnam then defines a certain kind of function, where a function f is of this kind iff:

- (6) For any predicate with extension E in a world W , if the predicate is nontrivial, then $f(E) \neq E$; otherwise, $f(E) = E$.

When f meets this condition, Putnam suggests that f can replace the interpretation-function in the standard model corresponding to W (p. 217). The result, he says, is a model that has f as a nonstandard interpretation-function onto objects in W (ibid.). In addition, Putnam contends that f will be a permutation on these objects (ibid). This, he claims, is sufficient for an isomorphism between the f -model and the standard model. If so, then they are models for exactly the same sentences. Further, since f is nonstandard, then we know there is a nonstandard interpretation of terms which conserves truth-values assigned by the standard interpretation in W (assuming that the theory has a predicate which is nontrivial in that world) (p. 218).⁸ This is meant to be sufficient for (C1.1).

⁷ N.B, (C1.1) entails (C1) [but not vice-versa]. Proof: Given a true scientific theory, its standard model will correspond to some possible world W . (C1.1) says, then, that there is a nonstandard interpretation which conserves the truth-values it has in W . (In this case, the truth-value conserved for each sentence would be “true”). So besides the standard interpretation, the theory also has a nonstandard interpretation which preserves truth, which is what (C1) says.

⁸ In Putnam’s first statement of the thesis, the point is restricted to the theories with a nontrivial predicate in *some world or other*. But this seems to be a slip, since Putnam gives a deviant interpretation to a predicate in a world only if the predicate is nontrivial *in that world*. So if we want a deviant interpretation for *every* world, the theory must have a nontrivial predicate in every world.

However, I find the argument here confusing in several respects. I think my first two concerns can be resolved if we revise the argument (albeit in considerable respects). But the third concern I think is more formidable.

First Concern: The reader may have noticed that f s is defined on certain *sets* of objects in a world (viz., extensions). Yet then, the function is not a permutation of the *objects* themselves, as Putnam seems to imply.

But maybe this is no big deal. Perhaps this was really the intention all along. In that case, you might worry whether one-one correspondence is preserved between the f -model and the standard model (which is a condition on isomorphism). Yet this worry disappears once we remember that the domain is *constant* between the two models.

Even so, one may wonder how f is an interpretation-function at all. For such a function is defined on *predicates*, whereas f takes *sets* as inputs. But charitably, it is not f itself which is a deviant interpretation-function, but rather a function that *defines* a deviant function. In particular, given such an f , we can define an interpretation-function g such that, for a predicate P_n with E as its standard extension, $g(P_n) = \{x \mid x \in f(E)\}$. So for a given world, it appears that $g(P_n)$ will be different from E , since $f(E) \neq E$ for nontrivial predicates. And so, g will be a nonstandard interpretation-function.

Second Concern: Yet whatever the details here, Putnam assumes that *any* function like f gives us the desired sort of interpretation. That is, given a world W , Putnam thinks that such a function will afford us an interpretation-function such that, when paired with the standard domain, gives us an *isomorphic* model corresponding to W , i.e. a model that makes precisely the same sentences true. *But that just isn't so.* For

there is not yet a guarantee that his interpretations will “co-ordinate” the predicates properly, so to speak. The point can be illustrated by following:

(7) Everest is a mountain & Everest is snow-capped.

As it stands, (7) is true in our world. Yet thanks to the relatively weak constraints on f , an f -model could make the extensions of ‘is a mountain’ and ‘is snow-capped’ *completely disjoint*, e.g., if ‘mountain’ denotes water and ‘snow-capped’ denotes rocks. And in a model which assigns such extensions, (7) will inevitably be false, no matter what ‘Everest’ denotes.

Thus, since isomorphism is not guaranteed, I conclude that Putnam’s argument for (C1.1) fails. Nonetheless, considering Putnam’s proof of (C1), one might have an educated guess as to what is meant to occur in the present case: Putnam may be just trying to generalize the earlier proof of (C1) to an arbitrary world. At this point, this would mean defining the deviant function g not just by any permutation f ; rather, given a world, g would be defined as the relation R mentioned earlier.

More exactly: Suppose M is the standard model corresponding to an arbitrary world W . Then, map the objects of M to objects of W , by some relation non-equivalent to identity. [But note that the objects of M and of W are the same, since M corresponds to W —so the mapping does turn out to be a permutation $f(x)$ on W -objects.] Now as before, consider the relation R such that: a term t bears R to y iff t denotes x in M , where $f(x) = y$. Since $f(x)$ is non-equivalent to identity, it is not the case that tRy iff tRx in a given world, at least for some t . So R can stand as a nonstandard interpretation-function in a model M^* corresponding to W , so that a term t denotes y in M^* iff tRy . And since R is basically *defined* to ensure isomorphism, M^* will make exactly the same sentences

true as in the standard model for W . Thus, for an arbitrary world, we can get a nonstandard interpretation of a theory which conserves truth-values in that world.

Third Concern: However, there is a third objection which affects to both versions of the argument to (C1.1). Note that (C1.1) is a claim about *any* theory in *any* possible world. But at least with the first version of the argument, it is assumed that at least one predicate of a theory will be *nontrivial* in a given world, on the standard interpretation. But consider, e.g., that all predicates in (7) will be trivial in a world where there exists only a snow-capped Everest (or worlds with any number of additional snow-capped mountains). Yet that means Putnam's strategy will not change the interpretation of the predicate in that world. So for all Putnam has shown, there are worlds where some theories have no nonstandard interpretation.⁹

So much for the first version of the argument, but does the second version fare better? Unfortunately, a bit of reflection reveals that R *must* assign the same extension to a trivial predicate in a world. First, suppose the predicate p is *empty* in the standard model M . Then since any term t is such that tRy iff t denotes x in M (where $f(x) = y$), and given that p denotes no x in M , it follows that p will not bear R to anything. Hence the predicate will remain empty when R is the interpretation-function.

Second, suppose a predicate p is *universal* in a world W . Then, its extension will include everything in W . Note, however, that the permutation $f(x)$ maps objects in W to objects in W . Because of this, R must assign p a universal extension in W . For again, a term t bears R to y iff t denotes x in M (where $f(x) = y$). And since p denotes *every* x in M

⁹ Note that a similar issue does not arise with (C1). In the proof of (C1), but not with (C1.1), our nonstandard interpretation can be nonstandard in virtue of having a *nonstandard domain* of objects. And in such a case, even universal predicates will get a nonstandard interpretation. Still, this does not guarantee that a theory using only empty predicates will have a nonstandard interpretation which preserves truth. But charitably, one may assume that "scientific" theories of concern use at least one nonempty predicate.

(and since $f(x)$ is one-one) p will bear R to every y in the nonstandard model. Yet the domain is constant between the standard and nonstandard model in these cases, since they are both supposed to be models which correspond to W . So p will denote precisely the same objects in both models, viz., all the objects in W .

Is there a way to show (C1.1) in its full generality? There is the thought that we could define nonstandard interpretations, where each predicate is assigned its anti-extension in a world. Then, we could invert the meaning of ' \sim ', so that its *absence* means that the predicate is unsatisfied by the object, and its presence means that it is satisfied. But at least for Putnam, the argument is supposed to work by keeping the logical constants constant. Besides, it looks like (C1.1) would be quite *trivial* if ' \sim ' can be reinterpreted. Of course, this may be grist for Putnam's mill, at least as concerns the upcoming argument for (C2). But in any case, it is unclear how one could show (C1.1) just by reinterpreting terms. I conclude then that the Putnamian strategy used in the argument for (C1.1) is non-demonstrative.

IV. Indeterminacy and the Lewis Objection

We now come to the second stage of the model-theoretic argument, where Putnam contends that (C1) or (C1.1), plus realism, entails (C2). But given the above problem with (C1.1), I shall not consider it further. Still, if (C1.1) holds *on the condition* that a theory uses nontrivial predicates in a world, perhaps that would also be enough to show (C2) from realism. This conditionalized version of (C1.1) would be:

(C1.2) For each possible world in which some predicate(s) of a theory are nontrivial, that theory can be given a nonstandard interpretation which conserves truth-values in that world.

Yet whether we assume (C1) or (C1.2), the puzzle now is to figure out *how* (C2) is supposed to follow from this plus realism.

It is agreed that the model-theoretic considerations suggest that no theory can fix the interpretation of its own terms. This is apparently what motivates Putnam to ask the realist if *anything* can fix a theory's interpretation on her view. But the negative answer at (C2) follows only if all other plausible, realist options have been eliminated. Putnam, however, has some sensitivity to this and does touch upon a few alternatives. One considered in his (1981) is the proposal is that, if a speaker's *intends* her use of 'Everest' to denote Everest, then this intention determines that Everest is the unique object of her term. But Putnam objects that there are nonstandard interpretations of her intentions as well, e.g., where the speaker intends to use 'Everest' to denote the number 0. And thus the issue is pushed back onto mental representations.

Perhaps this response is not entirely effective—but whatever we think here, other accounts of reference-fixing still need to be dealt with. In particular, there are two accounts mentioned in Lewis (1984) which are fairly compelling options, viz., the causal theory of reference, as well as a view Lewis calls "causal descriptivism."

As concerns the causal theory: The idea is that the denotation of a term is fixed in virtue of a causal-chain (of a specified sort) leading from the object, into the mind of the speaker, to tokenings of the term. Granted, the details of such an account may be unknown, but many philosophers could agree that *some* kind of causal-condition is what fixes denotation. Note, however, that the causal theorist is not suggesting that the causal *theory* fixes the interpretation of terms. Putnam (1977) rightly emphasizes that this is "just more theory;" after all, the model-theoretic reflections show that *no* theory can fix

denotation, including the causal theory. And so instead, the causal theorist is suggesting that the casual chains *themselves* fix denotation, independently of any theory of them.

But according to Lewis, the causal theory cannot be the whole story about reference (p. 227). Lewis thus favors a view he calls “causal descriptivism” which he defines as “descriptivism, global or local, in which the descriptions are largely couched in causal terms” (p. 226). However, Lewis admits that if such a descriptivism attempts to fix the meaning of nontrivial predicates via description (and nothing else), then the model-theoretic considerations show that such a view fails. Lewis thus adds a further condition on the determination of reference, which he introduces as follows:

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 sameness and difference in nature. Only these elite things and classes are eligible to serve as referents. The world—any world—has the makings of many interpretations that satisfy many theories; but most of these interpretations are disqualified because they employ ineligible referents. (p. 227)

The basic idea is that *the world* helps determine what terms denote, by the fact that the world contains exactly those things (and kinds) which are eligible referents.

Part of this restriction is apparently on which domain of objects will be the standard domain, viz., only objects with “objective sameness and difference in nature.” Yet the previous section revealed that a constant domain does not always safeguard against nonstandard interpretations. However, Lewis also constrains things by limiting the denotation of terms to *real* kinds and particulars, as opposed to “gruesome” or Cambridge kinds or particulars. The metaphysical essentialism in this is actually unpalatable to Putnam (1983, p. xii)—though as Lewis points out, he gives no real argument here. [This is especially strange, since it was Putnam (1975) who did much to

establish essentialism as the current orthodoxy]. And since Lewis' view explicitly embraces essentialism, Putnam would need more than an incredulous stare to write it off.

Still, the restriction on the domain, plus the restriction to real kinds/particulars, does not fix a unique extension of *all* terms in *any* scientific theory. In fact, Lewis himself concedes that a theory may describe two real particulars by the same predicates. (p. 223). In such a case, then, a descriptivist will be unable to fix a name for just one of these objects. Nonetheless, Lewis claims this is a "moderate" kind of indeterminacy, and that "the existence of moderate indeterminacy is not to be denied" (ibid).

Yet moderate indeterminacy would seem contentious indeed. In fact, a causal theorist might provide exactly the kind of opposition which Lewis discounts. For many causal theorists, there would be a fact-of-the-matter as to which real particular is causally linked to the introduction of a term (at a "baptism ceremony"), even if our descriptions are not definite. Of course, whether this is plausible in every case may be disputed. But the point is that moderate indeterminacy could be more of a liability than Lewis lets on.

Nonetheless, if causal descriptivism is just intended to be an alternative account of reference-fixing which enjoys some plausibility, then it seems fair to grant Lewis this much. Or at least, one can point to the causal theory of reference as a popular view of reference-fixing which Putnam should contend with. Either way, Putnam would not yet be warranted in concluding (C2) from the model-theoretic considerations and realism.

At various points, however, Putnam addresses the type of account which concerns Lewis. In his (1977), for instance, he complains that "the world doesn't pick models or interpret languages. *We* interpret our languages or nothing does" (p. 24). But

this remark strikes me as unfair. For the realist does not endow the world with any *agency* with respect to interpretation. According to the causal theory, *we* interpret a term as denoting a particular object in the world, e.g. at a “baptism ceremony.” Whereas, for causal descriptivism, *we* interpret a term by describing real kinds/particulars in the world. Of course, the world is relevant to the process of interpretation—but the point stands that we are the agents in this process.

In Putnam (1981), the causal theory is dealt with in a different manner. He first identifies a class of reference-fixing accounts which conform to the following schema:

(8) *x* refers to *y* iff *x* bears R to *y*

In this, “R is a relation definable in natural science vocabulary without using any semantical notions” (p. 45) Presumably, the causal theory of reference will be an example of an (8)-style account, though Lewis’ causal descriptivism may be as well—if talk of “describing” an object can be naturalized appropriately.

In any case, Putnam’s objection to (8)-style accounts is expressed as follows:

Given that there are many ‘correspondences’ between words and things, even many that satisfy our constraints, what *singles out* one particular correspondence R? Not the empirical correctness constraints [which have nonstandard interpretations]. Not, as we have seen, our intentions (rather R enters into determining what our intentions signify). It seems as if the fact that R *is* reference must be a *metaphysically, unexplainable* fact, a kind of primitive, surd, metaphysical truth. (p. 46)

However, this reply is dubious for two reasons. First, even if certain realist accounts have failed to explain reference-fixing, it does not follow that the reference-fixing is *inexplicable* for the realist (though an explanatory challenge would remain, no doubt). Second, what’s supposed to undermine the causal theory (among others) is that *other* accounts of reference-fixing have failed, e.g., the “intentions” account. But just because

certain other accounts have failed, it of course does not follow that the causal theory, etc., also fail.

Thus the causal accounts seem untouched by Putnam's objections. And so, there remain viable, realist options for reference-fixing which Putnam has not ruled out—meaning that (C2) will not follow from the model-theoretic considerations plus realism.

V. Another Model-Theoretic Argument.

I thus conclude that the model-theoretic argument, properly so called, is unsuccessful. However, there is another, less noted, anti-realist argument in Putnam (1976) and (1977), which also starts from the model-theoretic considerations. This second “model-theoretic argument” is outlined in Putnam (1977) as follows:

The problem with realist semantics—truth-conditional semantics—as Dummett has emphasized, is that if we hold that the understanding of the sentences of, say, set theory consists in our knowledge of their ‘truth conditions’, then how can we possibly say what *that* knowledge in turn consists in? (It cannot... consist in the use of language or ‘mentalese’ under... constraints, be they fixed or evolving, since such constraints are too weak to provide a determinate extension for the terms...) (p. 20).

The target here is truth-conditional semantics (TCS), based on a Tarskian theory of satisfaction for L-terms—a semantics which Putnam also calls “realist semantics.” Such a semantics is usually given under the assumption that understanding L consists in knowing (at least tacitly) a Tarskian theory for L. [N.B., Putnam switches between talking of L-sentences and of L-terms—for simplicity, I’ll just stick to terms.]

My best guess is that the argument above has something like the following, basic structure. For any language L:

(TCS) To understand L is to know (perhaps tacitly) a Tarskian theory of L.¹⁰
[Assume for *reductio*]

(9) A Tarskian theory of L has nonstandard interpretations which conserve truth-values. [from (C1)]¹¹

(10) If (TCS) and (9), then to understand L is not to know (either tacitly or explicitly) the extension of an L-term. [Assumption]

(11) To understand L is not to know (either tacitly or explicitly) the extension of an L-term. [from (TCS), (9), (10)]

However, (11) is seen as an absurdity. So to use Putnam's turn of phrase, how can (TCS) possibly explain what knowing the extension of a term (perhaps tacitly) consists in?

But again, I find Putnam's argument disorienting. The first concern I think can be quelled—though the second and third ones are more persistent.

First Concern: One issue is that falsity of (11) seems disputable. Consider a youngster who is competent to a significant degree in using 'Pollux;' she can competently report, for example, that Pollux is the brightest star in Gemini. But suppose she has never heard of Castor (and suppose she would be unable to distinguish Castor from Pollux, were we to point them out to her). In such a case, it looks like she would not know (not even tacitly) that 'Pollux' does not denote Castor. (She probably does not even have the *concept* of Castor.) If so, then she would not know (not even tacitly) a unique referent for 'Pollux.' 'Pollux' may denote Castor for all she knows.

Ironically, this kind of phenomenon is often cited *against* Tarskian theory—it seems to show that competence does not consist in knowing necessary and sufficient

¹⁰ Tacitness is not mentioned in Putnam; yet this is certainly necessary to ensure Putnam is not attacking a straw-man. Only readers of Tarski explicitly know a Tarskian theory for L, but of course, children can understand L without such explicit knowledge.

¹¹ It may be odd to assume that (C1) applies to a Tarskian theory, since this would assume the Tarskian theory is "scientific." Nonetheless, the argument for (C1) can be made *vis-à-vis* a Tarskian theory; so in that sense, the conditions that suffice for (C1) also suffice for (9). [N.B., (C1.2) would seem to entail (9) as well, assuming that a Tarskian theory has a nontrivial predicate in our world.]

conditions for the application of a term. But in the present context, the point is used against (11) in order to *defend* the Tarskian theory from Putnam's argument.

Regardless, even if (11) is false and Putnam's argument is unsound, he could still make a very similar case against Tarskian semantics. Consider: Putnam's rival would presumably contend that the youngster knows *something* about the object of 'Pollux.' Yet if (TCS) and (9) are true, Putnam can insist that she would know *nothing* about what 'Pollux' denotes. In this, Putnam would be assuming the following revision of (10):

(10.1) If (TCS) and (9), then to understand L is not to know anything (not even tacitly) about what L-terms denote.

So given (TCS) and (9), it would then follow:

(11.1) To understand L is not to know anything (not even tacitly) about what L-terms denote.

Yet (11.1), surely, is absurd. Even our youngster knows *something* about Pollux, e.g., that it is part of Gemini.

Second Concern: But this raises the obvious question as to why (10.1) is true. [I assume that (10) is now superseded by this premise.] Why should the *metaphysical* claim at (9), plus (TCS), entail the *epistemic* claim at (11.1)?

The matter requires us to delve deeper into what (10.1) means. First, as concerns its antecedent, (9) says that a Tarskian theory can be satisfied in many different ways. And in the present context, this means a Tarskian theory is *indeterminate* in what it says, unless something *other than itself* determines the extensions of its terms.

Thus (9) is true because either a Tarskian theory does not say anything—or, a Tarskian theory says something given that *something else* fixes denotation. Now in the former scenario, it is clear why (TCS) and (9) would entail the epistemic claim at (11.1);

that is to say, in such a case, it is clear why (10.1) is true. For if the Tarskian theory does not *say* anything, then merely being acquainted with its meaningless sentences would not suffice for *any* knowledge of L-terms. And that is precisely what (11.1) says. So (10.1) would be on target here, and the absurdity at (11.1) would follow.

Yet consider the case where (9) is true because something *other* than a Tarskian theory fixes denotation. Would (TCS) and (9) then suffice for (11.1)? No. As we just saw, if (TCS) has any chance at surviving the argument, then its talk of “knowing a Tarskian theory” should not concern just an acquaintance with meaningless expressions. Rather, it should concern knowledge of what these expressions *express*. Moreover, in the present case, we can assume that these expressions do express something, since *ex hypothesi* something (other than a Tarskian theory) fixes the denotation of terms. But then if (TCS) is true and read along these lines, it would seem that understanding L *does* mean knowing something about what L-terms denote—even if (9) is true. Knowing what a Tarskian theory for L expresses would appear sufficient for knowing *something* about what L-terms denote, contra (11.1).

Thus, I find the Putnam argument against (TCS) unsound. Nonetheless, there is a closely related, Wittgensteinian problem for (TCS). Suppose with (TCS) that understanding L consists in knowing a Tarskian theory of L. Then, if the Tarskian theory is in L, understanding L would presuppose understanding L—meaning it would be impossible to understand L in the first place. On the other hand, if the Tarskian theory consists of sentences in some other language L*, then understanding L would presuppose understanding this other language. But if understanding L* consists in knowing a Tarskian theory for L*, as (15) implies, then assuming *this* theory is not given

in L or L* (on pain of circularity), understanding L* would require understanding some *third* language. And thus a regress.

A defender of (TCS) might reply that if the Tarskian theory is given in *thought* rather than a natural language, then there is no need to worry about a regress. Yet plausibly, thought counts as a “language” at least in the minimal sense that it is a set of representations where some representations (viz., thoughts) are systematically composed from others (viz., concepts or atomic thoughts). [Note that this does not necessarily imply there is a “language of thought” in the sense of Fodor (1975).] Yet if thought is a language, it cannot function as a regress-stopper if TCS is supposed to account for knowledge of *any* language. Yet TCS is most likely intended to be an account of only spoken and written languages. If so, then thought can indeed stop the regress, though of course TCS would leave open the question of what we know when we know the use of a mental representation. In a loose sense, then, TCS might not count as a complete theory of what we know when we know a “language.”

Third Concern: Yet even if there is a cogent argument against (TCS), it is also unclear why Putnam thinks this affects *realism*. For one, given that he assimilates truth-conditional semantics to “realist semantics,” Putnam seems to think that any realist will accept (TCS). But this is simply incorrect. Realists can follow Wittgenstein instead to the view that linguistic competence consists in having certain abilities, and perhaps connectionist or dynamical models give a clue as to how all this works.^{12, 13}

¹² Apparently, it is also false that *only* realists accept (TCS)—in itself, truth-conditional semantics seems quite metaphysically neutral (see Bar-On et al.).

¹³ A good introduction to connectionism is Rummelhart (1989). Dynamical models are nicely summarized in the Introduction to Port and van Gelder (1998).

Moreover, even if there is a problem of explaining linguistic competence, it doubtful that this is a distinctly *realist* problem. It seems anyone would want an explanation of that. Putnam in fact proffers a “verificationist” semantics which promises to do better explanatory work than the alternative. However, if Putnam’s view is that understanding consists in knowing (at least tacitly) a certain kind of verification-condition, then it seems knowing a language may also consist in knowing (at least tacitly) a certain kind of *theory*, viz., a theory which gives verification-conditions for sentences. But if so, then it the second model-theoretic argument would be just as applicable to Putnam’s view as to (TCS). Of course, Putnam can construe knowledge of verification-conditions in terms of an ability to use sentences in such-and-such ways. But as I mentioned, this move also seems available to the realist—so Putnam has no advantage on this point either.

VI. Pessimism about Semantics.

So even if (TCS) is wrong, it seems that realism would also survive Putnam’s second argument from model theory. But even if Putnam’s arguments are unsound, are the model-theoretic considerations unimportant outside of model theory? Do they indicate anything philosophically significant? As I’ve mentioned, I think in fact they support a certain kind of *pessimistic* claim, viz., that in a fundamental sense (to put it roughly), *there is no answer* to the question of what a term, such as ‘Pollux’, denotes.

To bring this out, let me crudely borrow from Plato the fictional of “Socrates”—who asks us innocently enough “What does ‘Pollux’ denote?”¹⁴ Now at first, Socrates’ question may seem rather contrived. *Of course* we all know what ‘Pollux’ denotes. But

¹⁴ Of course, I do not mean to attribute to Plato or the historical Socrates any of the views discussed here.

Socrates grants that we all “know what ‘Pollux’ denotes,” in the sense that we all have a Wittgensteinian understanding of the term. Yet understanding the term in that sense does not mean we know an *interpretation* of the term. And Socrates is challenging us to give an interpretation which answers his question “What does ‘Pollux’ denote?”

But conjuring up such an interpretation seems very easy. In a Tarskian theory, a base clause such as the following would seem to resolve the matter:

(12) ‘Pollux’ denotes Pollux.

Or, taking a lead from Field (1972):

(13) ‘Pollux’ denotes what it denotes.

But though these are clearly true, they are quite *uninformative*. And so, they are not well-suited to *answer the question* of what ‘Pollux’ denotes. (Henceforth, I use ‘answer’ in a robust sense, so that an answer is the same as an adequate answer.) After all, if you began in ignorance of what ‘Pollux’ denotes, (12) and (13) will not dispel that ignorance.

We can take away a few lessons here however. First, given an interrogative with the normal form “What does ‘Pollux’ denote?,” an answer (if any) would be a *true* sentence having the normal form “‘Pollux’ denotes *y*,” i.e., a true interpretation of ‘Pollux’.¹⁵ As the more important lesson, however, (12) and (13) show that truth is not enough for an interpretation to answer a question; an interpretation must also be informative, to say the least. But of course, the additional informativeness requirement seems easy enough to meet. For instance, a theory might contain the following:

(14) ‘Pollux’ denotes Polydeuces.

¹⁵ An “interrogative” is a *sentence* ending with ‘?’, whereas “questions” are what interrogatives *express*. I assume an interrogative might express different questions in different contexts, depending on what is assumed as background. If this is a contentious view of questions, you can have the term ‘question’ as long as I can coin some term to express the notion here.

(15) ‘Pollux’ denotes the red giant in Gemini.

So again, Socrates seems mistaken if he means to suggest that his question is unanswerable.

However, it depends on the background knowledge or the assumptions in a context, as to whether (14) or (15) answer “What does ‘Pollux’ denote?.” The easiest way to see this is when ‘Polydeuces’ is a completely unfamiliar term, and the only response given to the question is (14). Yet for Socrates, this kind of thing is not going to be an issue. We may assume that all terms are understood in the Wittgensteinian sense (at least to a non-negligible degree).

But even so, (14) or (15) may not answer such a question, since an inquisitor might have the same sort of question for the right-hand terms. Suppose someone asks “What does ‘Pollux’ denote?” for the purpose of locating the object overhead in the night sky.¹⁶ Suppose further that she is familiar the terms ‘Polydeuces’ and ‘the red giant in Gemini’, though she is also unsure of the location of their object as well. In that case, neither (14) nor (15) will answer her question, even though they may tell her *something* new. In contrast, however, if she did previously know the location of “Polydeuces” or “the red giant in Gemini,” then (14) or (15) would presumably answer her question.

A variation on this case is where the inquisitor *antecedently knows* (14) and (15), though is again ignorant about where to look for “Polydeuces” and “the red giant in Gemini.” Here, it is even clearer that (14) or (15) will not resolve her question. When given (14) and (15), she may reply “Sure, I knew *that* already, but I might just as well

¹⁶ This illustrates the *purpose-relative* nature of wh-questions, as discussed by Boër and Lycan (1984). Intuitively, since the inquisitor’s purpose is to locate Pollux overhead, *x* answers her question only if it enables her to do this. However, purpose-relativity is a complication which we can avoid in the skeptic’s case, and so I mention it here only.

ask what ‘Polydeuces’ or ‘the red giant in Gemini’ picks out overhead.” The variant-case is notable, since it shows there can be a question about what ‘Pollux’ denotes, even if some true, informative interpretation(s) of ‘Pollux’ are antecedently known.

The upshot is that order to answer what a term denotes, it is not enough if you offer a true interpretation which is informative in *some* way or other. It also must be informative *in the right way*. This at least means: It must be informative in such a way that the parallel question for the right-hand term does not arise for the inquisitor.¹⁷ And whether the question gets “pushed back” onto the right-hand term is largely determined by what is previously known or assumed in the context of inquiry.¹⁸

Getting back to Socrates, let us first consider (14) as a possible response to his question. Like the stargazer, Socrates might object that (14) merely pushes back the issue onto ‘Polydeuces’, given his background. Perhaps, but what is his background exactly? Well, whatever else he assumes, my Socrates asks his question when a certain model-theoretic result is on his mind. Indeed, what first prompts his question is:

(16) ‘Pollux’ could denote any number of things, without effect to the truth-values of sentences.

This is shown in the same way as (C1): (16) is true since there are a variety of models isomorphic to the standard model¹⁹ where the denotation of ‘Pollux’ differs between these models. The models are constructed by first interpreting the names for Pollux as

¹⁷ The right-hand term in (14) may be an issue for the stargazer even if she does not ask a follow-up question with normal form “What does ‘Polydeuces’ denote?” She may instead ask “Where is Polydeuces?” But the metalinguistic question is still an issue for her, given that an answer to that question would obviate the need for the other, non-metalinguistic question.

¹⁸ In practice questions of course do not get pushed back *ad infinitum*; on this point, Boër and Lycan (1984) argue forcefully that a term stops the regress if it counts as “important” in a context. (Even so, the present section can be seen as arguing that nothing stops the regress in certain model-theoretic inquiries.)

¹⁹ N.B., the standard model is trivially isomorphic to itself. Thus, when I speak of models isomorphic to the standard model, this includes standard model.

names for an object y , and then stipulating that y satisfies an arbitrary predicate ‘F’ in the constructed model iff Pollux satisfies ‘F’ in the standard model.²⁰

Thus Socrates begins in the recognition that there are several possibilities for what ‘Pollux’ denotes, truth-values being what they are. This then prompts the question, roughly, as to which possibility is actual. Now Socrates grants that a question about what ‘Pollux’ denotes *might* be answered by (14). But in those cases, he worries that the question was not a very good one. Suppose, for example, that Socrates antecedently knows (14). Further, suppose he knows that in each isomorphism under consideration:

(17) What ‘Pollux’ denotes = what ‘Polydeuces’ denotes.

Socrates knows (17) must be true, based on his knowledge that (14) is true, and indeed, true in any isomorphism. For the truth of (14) in an isomorphism indicates that its terms co-refer.²¹ And such co-reference is exactly what (17) expresses.

Now given (16) and (17), Socrates will also be assuming that:

(18) ‘Pollux’ and ‘Polydeuces’ could co-denote any number of things, without effect to the truth-values of sentences.

After all, the various isomorphisms show that what ‘Pollux’ denotes could be any number of things, without a change in truth-values. But *per* (17), ‘Polydeuces’ will denote whatever ‘Pollux’ denotes in each of these isomorphisms. And so, what these terms co-denote may be any number of things, without a change in truth-values.

²⁰ This differs somewhat from Putnam’s method for constructing isomorphisms, since he operates at the “object level” by first mapping the *object* of ‘Pollux’ onto some y . Yet the upshot is the same since if we instead begin by mapping *names* for Pollux onto y , we will end up with the same isomorphism.

²¹ This reveals that Socrates is only considering isomorphisms where quote-names and the ‘denotes’ predicate are constant in meaning. Socrates could enlarge the set of isomorphisms so that this is not the case, but it suffices to . After all, if no candidate-answer “singles out” one possibility in the smaller set, the same will be true vis-à-vis the superset.

But given his assumption at (17), it is important to note something that Socrates will *not* be assuming. Among the possible objects for ‘Pollux’, Socrates of course does not presuppose that any particular possibility is actually what ‘Pollux’ denotes, on pain of begging his own question. Yet since he is antecedently aware of (17), he equally does not presuppose that any particular possibility is actually what ‘Polydeuces’ denotes. By his lights, to do so would *also* beg the question.

So although Socrates is asking a question about ‘Pollux’, he is really putting ‘Polydeuces’ in question as well. Yet that means (14) *does not answer his question*. Since the question about ‘Polydeuces’ is equally an issue, (14) only pushes her question about ‘Pollux’ back onto ‘Polydeuces’. He still accepts truth of (14)—but with all the isomorphisms in play, this is only to grant that its terms co-refer. [See Figure 1.]

The ineffectiveness of (14) might have been obvious enough; *ex hypothesi*, my Socrates had foreknowledge of its truth after all. But it is important to appreciate the reasons as to *why* his question arises, despite this foreknowledge. These reasons are important since, if Socrates had *not* antecedently known (14), he nonetheless would have had similar reasons to be dissatisfied.

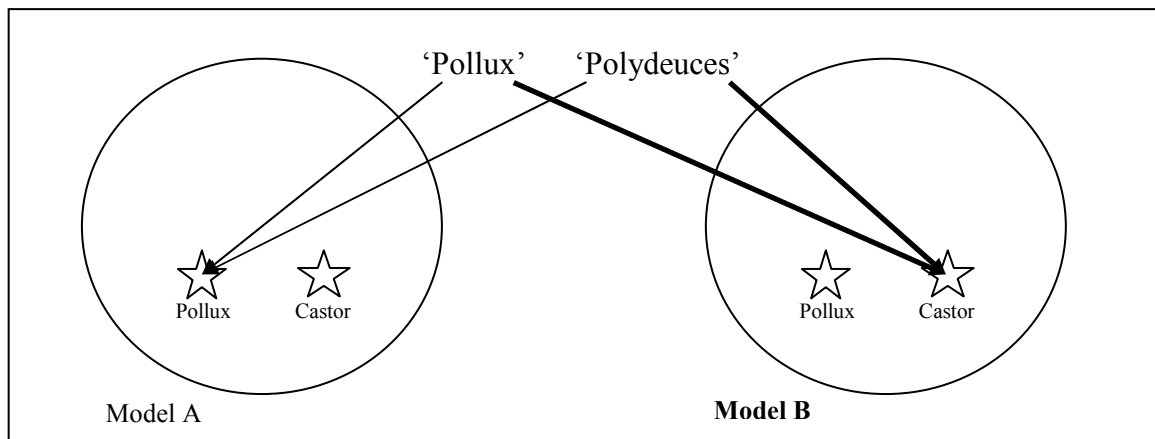


Figure 1: Two Models for (14). In both models depicted, (14) will be evaluated as true (given the co-reference of the two names), even though the terms are mapped to Castor in one case. Bold and unbold arrows show what the terms are mapped onto by the two different interpretation-functions.

So assume for the sake of argument that, although Socrates is familiar with the terms ‘Pollux’ and ‘Polydeuces’, he is unaware that they denote the same thing. (Cf. Puzzled Pierre’s competence with ‘London’ and ‘Londres’.) Thus, we may imagine he is antecedently unaware that (14) is true. Suppose then that he asks “What does ‘Pollux’ denote?” with (16) on his mind. Will (14) then answer his question? Well, let’s suppose that instead of (17), Socrates starts with the more general premise that:

(17.1) If ‘Pollux’ and a name ϕ co-refer, they do so in each isomorphic model under consideration.

(17.1) is true for a similar reason as (17): The relevant isomorphisms are simply defined as ones where *all* names for Pollux are interpreted as names for some object y .

Accordingly, given (16) and (17.1), he also takes as background that for any ϕ in the set N of names which co-refer with ‘Pollux’:

(18.1) ‘Pollux’ and ϕ could co-denote any number of things, without effect to the truth-values of sentences.

The argument for (18.1) is the same as the argument for (18), *mutatis mutandis*. Now it is also important that in the present case, Socrates makes a further assumption that:

(19) If ϕ is a name and \ulcorner ‘Pollux’ denotes ϕ \urcorner is true, then $\phi \in N$.

Though this is fairly trivial, it has a notable consequence when coupled with (18.1).

(20) If ϕ is a name and \ulcorner ‘Pollux’ denotes ϕ \urcorner is true, then ‘Pollux’ and ϕ could co-denote any number of things, without effect to the truth-values of sentences.

The argument for (20) and (18.1) are almost the same; we only need to add *per* (19) that if \ulcorner ‘Pollux’ denotes ϕ \urcorner is true, ϕ co-refers with ‘Pollux’.

Now as before, given Socrates’ background in this case, there are a few important things he will *not* be assuming. First: Just as he does not presuppose that some

particular object is what ‘Pollux’ denotes, he also does not presuppose that some particular object is what a co-referring name denotes. For he is well-aware by (17.1) that such an object *just is* what ‘Pollux’ denotes—a co-referring name is merely another label for the same thing. So by his lights, to presuppose that some particular object is what a co-referring name denotes would just beg the question he is asking. Second, and as an instance of this: He will not make any presupposition as to what is denoted by a name on the right-hand side of a true ‘Pollux’-interpretation. For the name will be co-referring, *per* (19). Generally, as far as Socrates is concerned, the identity-relation between the object of ‘Pollux’ and the object of an equivalent label is just too obvious to presuppose knowledge of one while asking about the other.

So again, though his question is about ‘Pollux’, Socrates takes the attitude that his question really indicts co-referring names as well, including any names on the right-side of true ‘Pollux’-interpretations. And so, if this is the background, she will be dissatisfied with (14), *even if (14) teaches him something new*. Once given (14), he can accept it is true, and thereby learn that ‘Pollux’ denotes Polydeuces (in the way that Puzzled Pierre might learn that ‘Londres’ denotes London). Nonetheless, as with the neophyte stargazer, Socrates may still complain that it merely pushes back his question about ‘Pollux’ onto ‘Polydeuces’. For given his background, he will hold ‘Polydeuces’ in question just as much as ‘Pollux’. After all, Socrates assumed all along that ‘Pollux’ and an arbitrary co-referring name will co-denote any number of things between the isomorphisms. But he is not simply going to assume that some particular object is what a co-referring name denotes, on pain of begging the question.

I have perhaps belabored the inadequacy of (14). But this is only because, if we also add a more descriptive interpretation such as (15), it turns out that *Socrates can make the same kind of complaint*. The kind of inquisitor which ultimately concerns me is one who accepts a variant of (17.1) which covers not just names, but *any* term ψ including the descriptor used in (15). Thus, assume that Socrates knows :

(17.2) If ‘Pollux’ and ψ co-refer, they do so in each isomorphic model under consideration.

Now when ψ is a descriptor, it may not be as obvious why this version of (17) is true. We know that in each isomorphism under consideration, once the names for Pollux have been interpreted as names for some y , it is then stipulated that y satisfies a predicate ‘F’ in the isomorphism iff Pollux satisfies ‘F’ in the standard model. Thus, all and only the same predicates which Pollux satisfies in the standard model are satisfied by this y . Moreover, if a descriptor is satisfied by Pollux in the standard model, it will be satisfied by this y in the isomorphism. After all, whether a descriptor is satisfied by an object is determined by which predicates it satisfies. Yet since the same predicates are satisfied in either model, the descriptor will be satisfied in either case. Thus, a descriptor will be satisfied by y in the isomorphism iff it is satisfied by Pollux in the standard model.

As the reader may expect, then, under (17.2) my Socrates is also aware that:

(18.2) If ‘Pollux’ and a term ψ co-refer, then they could co-denote any number of things, without effect to the truth-values of sentences.

In addition, he will presuppose the truism:

(19.1) Given an arbitrary term ψ , if \ulcorner ‘Pollux’ denotes $\psi \urcorner$ is true, then ‘Pollux’ and ψ co-refer.

And so parallel to the case of names, Socrates antecedently makes the inference from (18†) and (19.1) to:

(20.1) Given a term ψ , if \ulcorner ‘Pollux’ denotes ψ \urcorner is true, then ‘Pollux’ and ψ could co-denote any number of things, without effect to the truth-values of sentences.

But also, like the case of names, Socrates will *not* assume that any particular object is actually what a co-referring term denotes. And in particular, he will not assume that any particular object is denoted by a term on the right-side of a true ‘Pollux’-interpretation. Again, it seems rather obvious to him that, once we sufficiently comprehend the isomorphic possibilities being raised, it begs the question to assume such a thing.

So in this set-up, Socrates views his question about ‘Pollux’ as a question about its co-referring terms as well—including co-referring terms on the right-side of true ‘Pollux’-interpretations. This means he will be dissatisfied with (15) as well as (14), even if both teach him *something* new, e.g., that ‘Pollux’ and ‘Polydeuces’ co-refer. And as with (14), he can still grant that (15) is true, since relative to his assumptions, this only grants the co-reference of its terms. [See Figure II]. Even so, (15) merely pushes back the question about ‘Pollux’ onto the descriptor ‘the red giant in Gemini’. For he might have known all along that some true interpretations would use a co-referring

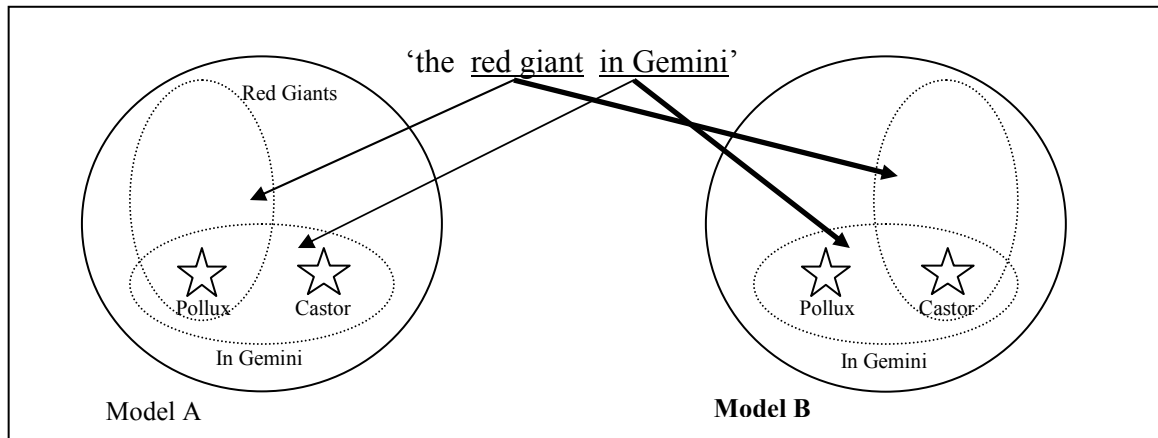


Figure 2: (15) in Model A and Model B. If we take models A and B from Figure I, we can also work it so that (15) is true in both models. The trick is in Model B: Though ‘in Gemini’ still denotes things in Gemini, ‘red giant’ does not denote red giants, but rather objects in a set X , where Castor is uniquely in X and in Gemini. (X could be, e.g., the set of things that are such-and-such-years from the sun.)

descriptor—yet he does not assume that any particular object is what such a descriptor denotes. Again, that object obviously just is what ‘Pollux’ denotes, meaning that to assume such a thing would be to beg the question.

We can imagine Socrates summing up the situation as follows:

Since there is an isomorphism where ‘Pollux’ denotes, e.g., Castor, my inquiry starts from the assumption that ‘Pollux’ might denote Castor (truth-values being what they are). Yet since I acknowledge (18.2), I am equally starting from the assumption that any co-referring term might denote Castor. That’s because I accept (17.2), which says that in an isomorphism where ‘Pollux’ denotes Castor, its co-referring terms all denote Castor as well. And so, since I suspend judgment about ‘Pollux’, I equally suspend judgment about its co-referring terms.

In particular, then, she suspends judgment about the right-hand terms in true ‘Pollux’-interpretations. And so, the right-hand term in (15) is just as much as in question as ‘Pollux’. Thus (15) will merely push the question about ‘Pollux’ back onto its descriptor.

Socrates even takes this attitude toward certain *demonstratives*, occurring in sentences like ‘ ‘Pollux’ denotes *that*’. When accompanied by the appropriate ostension, such a sentence may seem to provide a “information link” or “directly acquaint”

Socrates with the object in a way that would dispel his question. [Cf. Evans (1982).] But again, in a nonstandard isomorphism where ‘Pollux’ denotes Castor, the demonstrative will also denote Castor, if the truth of the sentence is to remain constant.²² Hence, since Socrates is aware of such things, the ostensive definition merely pushes the question back onto the demonstrative.

As the reader may have anticipated, if (16), (17.2), (18.2), (19.1), and (20.1) characterize Socrates’ background, it is not merely (15) which is inadequate to answer her question. Rather, it seems that *every* true ‘Pollux’-interpretation will be inadequate.

²² One may at least re-interpret the act of *pointing* so that the demonstrative denotes Castor. After all, a demonstrative itself has no “standard denotation,” unless it is trivially “whatever is ostended in a context.”

That's because any true interpretation of 'Pollux' will only push the question back onto the right-hand term, since *per* Socrates' background, the parallel question for the right-hand term will equally be an issue. So in this sense, these interpretations will be *just more theory* that is indicted by his question. But here is the rub: *Ex hypothesi*, an answer to Socrates' question, if any, *must* be a true interpretation of 'Pollux'. So this apparently means that *there is no answer to Socrates' question*.

The argument here may be seen as having the following, basic structure:

- (21) If there is an answer to "What does 'Pollux' denote?" relative to a set of background assumptions, then it is a true sentence with the normal form " 'Pollux' denotes *x*", i.e., it is a true interpretation of 'Pollux'. [Assumption]
- (22) In light of (16), Socrates asks "What does 'Pollux' denote?" while assuming that the parallel question for its co-referring terms is equally an issue. [Stipulation]
- (23) If (22), then no true interpretation of 'Pollux' answers "What does 'Pollux' denote?" relative to Socrates' assumptions. [Assumption]
- (24) No true interpretation of 'Pollux' answers "What does 'Pollux' denote?" relative to Socrates' assumptions. [from (22), (23)]
- (25) There is no answer to "What does 'Pollux' denote?" relative to Socrates' assumptions. [from (21), (24)]

And (25) just is the "pessimistic," epistemic-informational claim about semantic questions which I have been advertising. Specifically: Relative to certain assumptions, some non-trivial questions about denotation lack suitable answers.

First Concern: One concern with this argument is that the conclusion is unsurprising; it is just making the rather obvious point that you cannot put all of your terms in question at once. *Of course* Socrates must assume knowledge of what some term denotes, if he wants an answer to his question. However, it is important that Socrates is able to ask his question about 'Pollux', even when taking *most other terms*

for granted, e.g. ‘Nixon’, ‘water’, ‘sofa’, ‘the fastest antelope in 2008’, etc. For his assumptions only call into question ‘Pollux’ and its co-referring terms. Yet this is enough to make his question unanswerable.²³ Of course, when a co-referring term is a descriptor, there must be certain predicates whose denotation is a bit shifty between the isomorphisms.²⁴ But for each nonstandard isomorphism, such a predicate need only change whether the y denoted by ‘Pollux’ is in its extension. Beyond that, the extension need not be any different from what it is in the standard model.²⁵

Second Concern: Even so, it is one thing to put ‘Pollux’ in question, and it is another to put its co-referring terms in question. It may be true enough that *if* both are in question, then nothing can answer Socrates. But that may just show it’s a silly question.

However, it is a contingent matter as to what the sign-design ‘Pollux’ denotes, and so, inquiring into the matter *per se* is not obviously illegitimate. The issue, then, would be whether Socrates’ antecedent assumptions *make* it illegitimate. Socrates will contend, however, that his assumptions are merely the result of reflecting carefully on the possibilities. Again, the inquiry was prompted by the fact that ‘Pollux’ could denote any number of things, even if the truth-values of sentences are a given. So his assumption at (16) just expresses a fact of model theory. In addition, Socrates then recognized that its co-referring terms will denote whatever ‘Pollux’ denotes. Since this was deduced from facts about the isomorphisms under consideration, (17.2) also seems

²³ Socrates does take ‘Pollux’ to be an arbitrary case however. So though the problem can be somewhat localized to ‘Pollux’ and co-referring terms, it may arise for any grouping of co-referring terms.

²⁴ Note, however, that one need not regard every predicate in a given descriptor as shifty. Figure 2 indicates in the case of ‘the red giant in Gemini’, Socrates can keep ‘in Gemini’ constant in meaning while raising multiple possibilities for what the descriptor denotes.

²⁵ That is so, unless our theory makes some claim about the cardinality of the extension. Then, the extension would also need to exclude exactly one other z , and then semantic adjustments would need to be made in relation to z as well.

unproblematic as a starting assumption. (18.2) was then deduced, and given the truism at (19.1), the claim at (20.1) also was inferred.

So if Socrates' assumptions are illegitimate, it is hard to see why. Sure, it is possible not to follow Socrates in doubting all co-referring terms when asking about 'Pollux'. Yet then my Socrates will accuse you of being insufficiently reflective. According to him, if you are considering an isomorphic possibility where 'Pollux' denotes something nonstandard, you should realize that in such a model certain *other* terms have nonstandard denotations, co-referring terms at the least. Thus, the isomorphic possibilities put the co-referring terms up for grabs as much as 'Pollux'.

Third Concern: There is a different worry to the effect that Socrates' inquiry is rather silly. His question arises in light of certain isomorphisms—but when Socrates described these isomorphisms, he said *inter alia* that:

(26) In an isomorphism, the object y assigned to 'Pollux' satisfies a predicate iff: in the standard model, Pollux satisfies the predicate.

But (26) apparently presupposes that 'Pollux' denotes *Pollux* in the standard model.

Even so, this is an uninformative, homophonic interpretation; hence, it is not as if Socrates is presupposing an *answer* to her own question. Perhaps this point is clearer if, instead of (26), Socrates constructs the isomorphisms via the following:

(26.1) In an isomorphism, the object y assigned to 'Pollux' satisfies a predicate iff: In the standard model, whatever 'Pollux' denotes satisfies the predicate.

So it seems Socrates can construct the isomorphisms without prejudging his question.

Yet, even though (26.1) and (26) may be quite non-committal, they may still provide the start of an answer to Socrates' question. Specifically, it seems we could add

various empirical facts about Pollux to “thicken” the conception of what ‘Pollux’ denotes, such as:

(27) Whatever ‘Pollux’ denotes = Pollux = *the red giant in Gemini*.

So here, it may be that Socrates himself provides the basis for an answer to his question.

However, Socrates will regard the descriptor in (27) as part of what’s in question. Hence, although (27) may be (at least) informative of the co-reference between its terms, the right-most term cannot be used to resolve her question about the object referred to—and so, his question remains an open one.

Basically, the reason why Socrates’ question is not self-undermining is that he can describe the isomorphisms using an idiom that only allows for an uninformative response to his question. However, there is something in the area that Socrates should concede. Suppose that we use italicized English terms as terms of a distinct metalanguage. Then, instead of the English statement at (26) or (26.1), Socrates might specify the relevant isomorphisms in the metalanguage as follows:

(26.2) *In an isomorphism, the object y assigned to ‘Pollux’ satisfies a predicate iff: in the standard model, Pollux satisfies the predicate.*

In assenting to this thought, Socrates would apparently be presupposing the truth of:

(28) *In the standard model, ‘Pollux’ denotes Pollux.*

Although (28) resembles an uninformative, homophonic interpretation, strictly speaking it would not be. For ‘Pollux’ and ‘*Pollux*’ are now assumed to be terms of different languages; hence, (28) at least informs us of how to translate ‘Pollux’ in the metalanguage. (And since our metalanguage need not be “metaEnglish”, as it were, this translation is not insignificant.)

But if (28) is informative, this makes a significant difference. Unlike the case where the metalanguage is English itself, it seems Socrates' question here would be raised while an answer *is* presupposed. For (29) is a true, informative interpretation which uses a term not indicted by his question. That is so, assuming Socrates' question only concerns co-referring terms in English (vs. metaEnglish).²⁶

I think Socrates should concede this objection, so far as it goes. There is no reason to ask what the English term 'Pollux' denotes under Socrates' current assumptions, if the isomorphisms are constructed via (28). For in representing those isomorphisms, Socrates would in effect presuppose the truth of a sentence which settles that question before it arises, such as (28).

Nonetheless, Socrates may think that the success of (28) is rather superficial. Though his question concerned English terms, strictly speaking, his question obviously extends to metaEnglish terms as well. Thus, even if (28) answers his initial question, it will nonetheless be all too apparent to him that:

(17.3) If 'Pollux' and '*Pollux*' co-refer, they do so in each isomorphic model under consideration.

[The argument for (17.3) is virtually the same as for (17.2); that argument remains sound if it extends to '*Pollux*' in addition to non-italicized terms.] So under (17.3), the isomorphisms indicate that 'Pollux' and '*Pollux*' could co-denote any number of things. Thus, Socrates might have no antecedent assumption as to what either actually denote, meaning that under (17.3), (28) will not answer the question. For (28) uses a term which now indicted by that question.

²⁶ A similar point holds if the metalanguage is a different natural language instead, e.g., French, Urdu, etc.

Still, we could answer the extended question as well, provided there is a term μ which is not yet indicted by the inquiry. But even if such a μ remains, Socrates can then extend her question to cover μ as well—and indeed, he can continue to extend the question until all co-referring terms are indicted, in every language. In this, Socrates is not necessarily trying to be uncooperative. He is rather trying to get us to admit the rather obvious fact that the object of ‘Pollux’ *just is* the object of any co-referring term, in all the isomorphisms. Once we are clear on that, it is just idle to use such a term in an attempt to answer Socrates’ question.

So even though his question may have an answer in certain cases, Socrates thinks this happens only under a certain feigned naivety. After all, it is obvious enough that his question can cover any co-referring term. Furthermore, if a co-referring term is not indicted by the question, he maintains this is *not* because the issue has been antecedently resolved in case of that term. Rather, he might hold that we simply agree (perhaps tacitly) to use the term “blindly,” as it were. It will be a representation which we understand in the Wittgenstein sense, but our understanding will occur absent an answer to the question of what it denotes.

In Quine’s (1969) terminology, such a term will be part of a “background language”—a language where semantic questions are answered by terms not indicted by those questions. However, even if this language affords us answers to certain questions about denotation, Socrates has pointed out two undesirable features of those answers: (i) they function as answers only if we (perhaps deliberately) ignore the question in relation to the terms used in those answers—and as a consequence, (ii) they employ terms that we use “blindly,” or terms that we merely “take at face value” (to use Quine’s phrase).

Thus, the only questions about denotation we can answer are relatively naïve ones, and the answers we give will just be idle when seen in a more honest light. As Jay Rosenberg once said, “To terminate a regress of background languages in our own “taken at face value” is to cash a forged check with counterfeit currency” (1974, p. 56).

In subsequent chapters, I shall call “semantic pessimism” the view which (at least) acknowledges that some questions about denotation have no answers, except in a language whose terms are used unquestioningly. Thus, semantic pessimism is not merely the truism that (non-circular) interpretation has to stop somewhere. It is also an admission that some questions about denotation will inevitably go unanswered.

VII. Contrast with Cartesian and Quinean Scepticisms.

It is important to reiterate, however, that the pessimistic view developed here is different from Quine’s (1960), (1969) semantic nihilism. Unlike Quine, I am not saying there is *no fact-of-the-matter* as to the semantics of terms. I have not been concerned to show such a metaphysical claim, but rather just the *epistemic* claim that certain semantic questions have no informationally adequate answers. And one may grant there is no answer to the Socrates’ question, without granting there is no absolute fact-of-the-matter about what terms denote.²⁷ It might be a very robust fact that ‘Pollux’ denotes a certain object, even if no sentence answers Socrates’ question on the matter.

There are two other noteworthy differences between semantic pessimism and Quinean indeterminism. For one, Quine’s argument for indeterminacy begins in

²⁷ Actually, the epistemic claim would imply the metaphysical claim *if* the questions were unanswerable because there are no *truths* on the matter. Yet this is not pertinent—I have assumed all along that there are true interpretations of ‘Pollux’. Instead, the argument has been that any response to Lucy’s question ends up being question-begging, or at least, accepted under a contrived naivety.

behaviorist commitments, whereas the pessimistic argument at (21)-(25) does not. Nor is behaviorism entailed by that argument; the premises make no mention of behavior at all. Second, Quine (*op. cit.*) apparently thought that questions of denotation might be settled by ostension. “In practice of course we end the regress...by something like pointing” (p. 49). Ostension would naturally answer our stargazer’s earlier question about what ‘Pollux’ denotes. But as concerns Socrates’ question, the ostensive definition can only function (at best) as a sort of “counterfeit” answer.²⁸

It is also important that Socrates’ question is not just a skeptical question about what a term denotes. Recall that the Cartesian skeptic presses the question of whether our beliefs are *completely false*, thanks to an evil demon. But Socrates can *take for granted* the truth of what we assert; he can accept the truth of ‘ ‘Pollux’ denotes Pollux,’ ‘ ‘Pollux’ denotes Polydeuces,’ ‘ ‘Pollux’ denotes *that*’, etc. Even so, there remains a question about what ‘Pollux’ denotes, as illustrated by the various isomorphisms.

Still, Socrates’ question seems to take seriously skeptical possibilities where ‘Pollux’ denotes something other than what believe it denotes. So perhaps Socrates’ question is a skeptical question after all. But, assuming our commonsense belief would be expressed by an interpretation, Socrates can *grant* the truth of that belief, since he can grant the truth of *all* interpretations of ‘Pollux’. Again, that’s because the truth of all these interpretations does not settle his particular question of what ‘Pollux’ denotes.

Relatedly, one can in fact show that Cartesian skepticism is independent of semantic pessimism. First, Descartes provides an example of entertaining Cartesian

²⁸ Eventually, Quine also criticizes the ostensive definition, via his thesis about the inscrutability of reference (applied to demonstratives). Yet again, my criticism is not a metaphysical point about there being no fact-of-the-matter about what is ostended—it is an epistemic point that the ostensive definition is informationally inadequate in the context of Socrates’ question.

skepticism while using your terms unquestioningly. (In Descartes, your assertion is assumed to be true or false with respect to its standard truth-condition.) But just as you can hold the *content* of an assertion fixed while questioning its truth, so too can you hold its *truth* fixed while questioning its content.²⁹ Thus, Socrates can assume our interpretations of ‘Pollux’ are all true (so that there is no possibility of massive error), though the worldly-conditions which *make* them true are put into question.

Still, both the skeptic and my Socrates are raising multiple possibilities for what the world “in itself” is like. The Cartesian skeptic suggests that an assertion about Pollux might not be veridical, whereas Socrates suggests (in a background language) that its veridicality might depend on something other than Pollux. Of course, it could be that *both* the assertion is non-veridical, *and* the assertion is about something other than Pollux. But the point is that each character individually calls into question what kind of world the assertion occurs in. That is so, even if one questions the veridicality of the assertion, whereas the other questions what its veridicality would mean.

However, even if semantic pessimism is not Cartesian skepticism, it may seem to be nothing new—for Socrates seems just to present a case where “relevant alternatives” block knowledge. In this light, my Socrates would be presenting the same kind of problem as in Goldman’s fake barns case. In Goldman (1976), we are to imagine Henry who drives through a countryside populated with many barn-facades, unaware of this strange situation. We then suppose Henry happens to believe “there is a barn” in the presence of the one real barn in the area. Goldman’s intuition is that Henry does not

²⁹ I am reminded here of Davidson (1974): “[W]e cannot infer the belief without knowing the meaning, and have no chance of inferring the meaning without the belief” (pp. 310-11). However, the upshot of the present discussion is (roughly) that holding one of these fixed is insufficient for knowing the other. The skeptic thinks you cannot learn truths about the world, even if meanings are used unquestioningly—whereas my Socrates thinks you could not discern meaning, even granting the truth of your beliefs.

know there is a genuine barn before him, given that Henry would not discriminate between the real barn and certain “relevant alternatives”, viz., the fakes. Yet with Socrates, relevant alternatives seem to be exactly what prevents him from knowing: Socrates does not know what ‘Pollux’ denotes, since other objects are “relevant alternatives” that are not discriminated from its actual object.

I would agree that Socrates’ problem about denotation resembles Henry’s problem to a degree. In both cases, a person accepts certain truths, yet lacks knowledge, thanks to “relevant alternatives.” Nonetheless, there is at least one crucial difference. Unlike Henry’s problem, Socrates’ problem persists even if one is *de facto* able to discriminate between the object of ‘Pollux’ and all relevant alternatives. Knowing that ‘Pollux’ denotes *that* [with the appropriate ostension] may be enough for you to discriminate the object of ‘Pollux’ from any other object. After all, knowing the ostensive definition would at least allow you to distinguish the object of ‘Pollux’ from other objects by its *actual location*. Nonetheless, we saw before that the ostensive definition would not settle Socrates’ question of what ‘Pollux’ denotes, when he antecedently recognizes the co-reference between ‘Pollux’ and the right-hand term of any interpretation. Thus, although he is *de facto* able to discriminate what ‘Pollux’ denotes from other objects, his means of discriminating it may be question-begging, when it comes to answering his question.

VIII. Semantic Pessimism and Anti-Realism.

Importantly, the problem with denotation is not specific to realism, as Putman might have hoped. After all, if the challenge is (very roughly) just to *say* which objects

our terms denote in the standard model, the metaphysics of these objects seems irrelevant. In particular, the argument would seem to work even if denoted objects are theory-*dependent*, in the way that Putnam's (C3) suggests. Suppose, for example, that the ideal theory contains the sentence:

“ ‘Pollux’ denotes Polydeuces. ”

Since it is part of an epistemically ideal theory, we can assume that this sentence has ideal justification. Nonetheless, the semantic point would remain that such sentences have nonstandard interpretations by (C1). Thus, even here there is an entrée for Socrates' question—and that is apparently enough to get the argument going. This remains the case even if the denotation of ‘Pollux’ is *fixed* by what the ideal theory says it denotes. For Socrates' question concerns what the theory says in the first place.

Indeed, the point would also apply if the denotation of a term is even more radically dependent on a theory. Consider the following:

(30) ‘This very sentence’ denotes this very sentence.

In English, ‘this very sentence’ above would denote its own interpretation at (30). So the denotation of ‘this very sentence’ is utterly dependent on this little “theory” of what it denotes. Nonetheless, (30) is an uninformative *homophonic* interpretation, so it is fairly clear that cannot answer any questions about what ‘this very sentence’ denotes. And this just means the skeptic's question about ‘this very sentence’ is viable here as well.

In this chapter, I first argued that the model-theoretic considerations do not entail anti-realism, as Putnam claims. However, they do indicate something philosophically significant nonetheless. Namely, it seems both realists and anti-realists *cannot answer*

Socrates' question about the denotation of 'Pollux', except perhaps by giving a "counterfeit" answer, articulated in a language used unquestioningly.

IX. The Significance of Pessimism

Although semantic pessimism does not imply anything as metaphysical as anti-realism, it does have philosophical import. One notable consequence, pointed out to me by Dorit Bar-On, is that the argument demonstrates that on a certain reading, it is false to say that a formal semantics "gives the meaning" of the object-language expressions. In particular, if a theory "gives the meaning" only if it answers what a term denotes in the *standard* model (as opposed to some non-standard isomorphism), then formal semantics *cannot* give the meaning of an object-language term. There are other ways of reading "gives the meaning," however, where semantics succeeds in this task, e.g., when the metalanguage *translates* a different natural language. But in a more theoretically weighty sense, semantics cannot give the meaning of the target terms.

More importantly, however, the result of this Chapter checks the default optimism with which we approach semantics, viz., the optimism that any well-formed question about meaning is assumed to be answerable. Also, as I mentioned, the argument here vindicates Quine's thesis of "ontological relativity," where questions about denotation are, at best, answerable only relative to a "background language" whose terms are used unquestioningly.

Interestingly, Quine himself hints how this thesis has greater philosophical significance. Note the closing of "Ontological Relativity":

[B]oth ontology and satisfaction are matters of reference. In their elusiveness, at any rate in their emptiness now and again except relative to a broader background—both truth and ontology may in a suddenly rather clear and even tolerant sense be said to belong to transcendental metaphysics. (p. 212)

Quine's reference to "transcendental metaphysics" invites thoughts of Kant, who held that the "things-in-themselves," i.e., the objects of transcendental metaphysics, are unknowable. I do not wish to commit Quine or myself to such a sweeping claim—nonetheless, Quine seems to see some important relation between the thesis of ontological relativity and the Kantian view. In fact, if questions about ontology are unanswerable except relative to a background language, this could be *one* manner in which the "things-in-themselves" are unknowable. (Though for Kant to say such objects are *completely* unknowable is probably to exaggerate the point.)

CHAPTER TWO: SEMANTIC PESSIMISM AND ONTIC-IDIOMS

In the last chapter, we saw that there was no answer to the question “What does ‘Pollux’ denote?” in certain model-theoretic contexts. This, I think, indicates a limit on what semantic theory may teach us, regarding the *extension* of our terms. In this chapter, however, I want to discuss a different limit concerning the *intension* of certain expressions. In particular, in this chapter I will argue that a semantic theory cannot answer unequivocally what the intension is of ontologically significant idioms [a.k.a. “ontic-idioms”] such as ‘exist’, ‘there is’, ‘actual’, ‘real’, and the like. I shall then draw out a philosophical consequence of this result, viz., that there is *no unequivocal statement of Realism* about x , for any x , in our language.

The result here constitutes a further limit on the “default stance” in semantics, mentioned in the last chapter. That is, the inability to answer certain questions about ontic-idioms tempers the “optimistic” view that any well-formed question in semantic theory has an adequate answer (barring cases of presupposition-failure and the like). The present chapter, then, can be seen as a further development of the more “pessimistic” stance I take toward semantic theory. But as before, I adopt this stance not to belittle semantic theory and its many important accomplishments. Again, semantic pessimism is rather an acknowledgement of the limits of semantic theorizing, where this is meant to service our understanding of the epistemic position of human beings.

I. Five-and-a-Half Ways to be Innocent.

Ontic-idioms, of course, are paradigmatically used to assert the existence of various entities, as in sentences such as ‘I exist’, ‘There are black swans,’ ‘Santa Claus is real,’ and so forth. My primary concern here, however, is the fact that ontic-idioms have ontologically “innocent” or non-commissive uses (besides their usual “loaded” or commissive uses).¹ That is, there are containing sentences which do not entail the existence of the object, to which the idiom applies. (I shall give examples shortly). Normally of course, no special philosophical problem arises when an idiom has more than one use. But in the case of ontic-idioms, things are different: I will argue that the multiple uses make it *impossible to define* an ontic-idiom unequivocally. And that will signify that a semantic theory cannot adequately articulate what the intension is of an ontic-idiom (on an occasion of use).

How can an ontic-idiom be used innocently? First, it is notable that an ontic-idiom can sometimes be used without any ontological meaning whatsoever.² For instance, ‘really’ can be used as an “intensifier,” as in the following:

(1) The Fonz is really cool.

Here, ‘really’ serves to indicate a high degree of “coolness,” akin to the word ‘very’.

‘Really’ is not used to indicate the *bona fide* existence of anything.

¹ Azzouni (2007) in fact argues that *every* use of an ontic-idiom is non-commissive. (Thus, whether ‘I exists’ ontologically commits you to a “self” is not a semantic affair, but an *illocutionary* one). I am sympathetic to his view, especially since it also suggests that a criterion of ontological commitment is impossible. But I shall not pursue this topic here. My aim is rather to illustrate the various non-commissive uses—and to show that *even if* there are commissive uses, a criterion of ontological commitment is *still* impossible.

² I won’t discuss the use of ‘actually’ where it is merely a pragmatic device to indicate something unexpected: ‘Actually, artificial creamer is worse for you than coffee.’ On its natural reading, the sentence is equivalent to the same sentence minus ‘actually’; so there’s nothing semantic to discuss here.

A second kind of case, where an ontic-idiom has no distinctly ontological meaning, may be seen in the following:

(2) Max bought canned fruit instead of real fruit.

In this, ‘real’ is not used to distinguish fruit with one ontological status from those of a different status. Instead, ‘real’ is used to distinguish fruit by some other category.³ So in a vacuous way, ‘real’ is used in an ontologically “innocent” way, since here it lacks an ontological meaning *per se*.⁴

Since these examples feature ‘real’ *sans* ontological meaning, the idiom is apparently ambiguous in the most robust sense. But even when an ontic-idiom is used with a distinctly ontological sense, we shall see that these idioms can still be read in more than one way. However, whether these additional readings show further *ambiguity*, as opposed to *polysemy* or mere *context-sensitivity*, is not something I shall address. I shall rather say in a rather neutral way that these ontic-idioms are *equivocal*.⁵

When an ontic-idiom is used with an ontological meaning, we can identify at least three other ways the idiom can be non-commisive. A familiar example can be illustrated by the occurrence of ‘exists’ in the following:

(3) Napoleon exists.

Note that ‘exists’ can be read so that the sentence implies that Napoleon exists *now*. And if people do not survive their deaths, the sentence is false on that reading. But (3) also has

³ As a special case, ‘real’ might be used to distinguish fruit from fake (plastic) fruit. But since fake fruit exists just as much as real fruit, I assume this would also be a use without a “distinctly ontological” meaning. (If anything, ‘real’ in this case would have *epistemic* significance, where it contrasts fruit from objects that merely *look like* fruit.)

⁴ One might object that the sentence nonetheless implies that the uncanned vegetables exist, so we cannot yet conclude that ‘real’ is devoid of ontological import. But in fact, (1) does not have this implication, since for all I’ve said, Max and his vegetables are fictitious.

⁵ Here I follow Fodor (2006). See p. 99 n. 20.

an uncontroversially true reading, where Napoleon is contrasted with a fictional person—where the sentence just puts Napoleon in a different ontological category than, say, Darth Vader. On that reading, then, the ontic-idiom would be used in a non-commisive manner, at least in the sense that it would not entail the existence of Napoleon at the *present* time.

Another familiar example of non-commisive use is when an ontic-idiom is *relativized to a non-actual world*. Lewis (1986), in particular, taught us that ‘exist’ is world-relative, which implies that a containing sentence can be true even if the ontic-idiom applies to a non-actual object. So for example, if we are talking about a world described by Greek Myth, in that context the following sentence will be true:

(4) Pegasus exists.

Yet there is no mystery here, since the English sentence in this context is plausibly shorthand for something more complex, such as “Pegasus exists *in the world described by Greek Myth*.” Nonetheless, since the bit in italics is omitted in (4), the result is that (4) can illustrate a non-commisive use of the ontic-idiom: In a context where it is true, (4) does not entail that Pegasus is a denizen of the *actual* world.⁶

An importantly different non-commisive use can also be illustrated by (4). We just read the ontic-idiom in (4) as applying to something in a *non-actual* world, but we can also read the idiom as applying to an *actual* imaginary object, an object that is created in the minds of *actual* people. This kind of reading is, at least, the most natural one if we are to read the following sentence as true:

(5) Pegasus exists as a creation of the imagination.

⁶ The world-relative use of ontic-idioms, I take it, would subsume those uses where an ontic-idiom is relativized to a *fiction*. So there is no need to mention “fiction-relative” uses as a separate category.

In (5), ‘exists’ applies to an object of the imagination, Pegasus, yet the sentence does not entail that Pegasus exists in the way that real, live horses do. Quite the contrary, the truth of the sentence implies that Pegasus is only imaginary. So in that sense, ‘exists’ is non-commissive. And that is so, even though it is relativized neither to a time, *nor to some other possible world*. The ontic-idiom in (5) is non-commissive because of what Pegasus is in the *actual* world, viz., a mere object of the imagination.

Some will balk at the suggestion that Pegasus exists in the actual world, since it is a truism that Pegasus does not exist. But I agree that Pegasus does not **exist** (where boldface type indicates the commissive use of an ontic-idiom). Yet if (5) has a reading where it is true, then Pegasus indeed “exists” as something imaginary, where “exist” here ranges over **existing** objects as well as the objects that actual people imagine. Still, some may demand further explanation concerning the “existence” of Pegasus in this sense, and I offer additional explanation in the next chapter. But my present aim is just to document the different non-commissive uses of ontic-idioms in the language. Suffice it to say, then, that (5) indeed uses its ontic-idiom in a non-commissive manner, even though the idiom is not relativized to some other world. That linguistic point stands, even though the metaphysics behind the point may be somewhat puzzling.

My section-header indicated five-and-a-half ways for an ontic-idiom to be innocent, and the “half” concerns the use of ontic-idioms in mathematical assertions:

(6) There is an even prime.

If (6) is asserted in an ordinary high-school math class, for instance, then the sentence will not commit the speaker to Platonism about the number two. (That is so even if the speaker happens to be a Platonist.) So is this a further type of non-commissive use? Well,

that depends on your view of numbers. If numbers are creations of the imagination (as Kant held?), then the use of ‘there is’ in (6) may be of the same kind as the use of ‘exists’ in (5). Or, if numbers are found in other possible worlds (but not the actual world), then (6) may use ‘there is’ in the same way that (4) uses ‘exist’.

But to take a different angle, suppose Platonism is true. Then, (6) would materially imply⁷ the existence of a number—and so in that sense, ‘there is’ in (6) would be *commissive*. Yet assuming Platonism would not be a necessary truth,⁸ (6) would not *strictly* imply the existence of two. And so in that sense, ‘there is’ would still be “non-commissive.” But notably, this non-commissive use of ‘exist’ would differ from the non-commissive uses mentioned previously. For in the other cases, the containing sentences did not even materially imply the existence of the relevant objects. Whereas, if Platonism is true (but not necessary), the truth of (6) would materially (though not strictly) imply the existence of two. So depending on how things work out with Platonism, (6) might indeed illustrate a further, non-commissive use of an ontic-idiom.

II. Defining the Commissive Use.

Now all this would be well and good, except there is a problem in separating these non-commissive uses from the commissive uses. As a first pass, we may say that:

- (*) An ontic-idiom is used commissively just in case a containing sentence entails the existence of the object to which the idiom applies.

⁷ To be clear, p materially implies q iff $\sim(p \ \& \ \sim q)$

⁸ Incidentally, things get interesting if the Platonist holds that Platonism is necessary. For then, ‘exist’ in (6) would have to be commissive. So in that case, if the Platonist allows that one can assert (6) without committing to Platonism (as she should), then the Platonist would have to deny that ontological commitment is a *semantic* affair. It might instead be an illocutionary affair, as Azzouni (2007) suggests. (See n. 1).

But the problem is that ‘existence’ in the definition is precisely one of the equivocal idioms in question—meaning that the definition is also equivocal.

The equivocation is bothersome since, on some readings of (*), the ontic-idioms in (4) and (5) will count as “commissive” when clearly they are not. After all, (4) and (5) entail the “existence” of Pegasus in some sense (though each sentence would imply his “existence” in differing senses). Yet the sorts of “existence” in (4) and (5) were not intended when I defined the *commissive* uses of ontic-idioms. The sense of ‘existence’ intended was **existence**, the kind of existence that Obama, Seattle, and the rock of Gibraltar have—and that Pegasus does not.

We could of course define the commissive uses of ontic-idioms using the term ‘**existence**.’ But ‘**existence**’ merely indicates the commissive use of ‘existence,’ and the “commissive” use of an ontic-idiom was exactly what we want to define in the first place. Of course, a circular definition would be quite correct (“An ontic-idiom is used commissively just in case the containing sentence entails the existence of the object, where ‘existence’ is used commissively.”) But the circular definition is rather uninformative. We thus would prefer some independent way of defining ‘existence’ on its commissive reading, as it occurs in (*).

It is worth noting, however, that the issue is not merely of interest to making the commissive/non-commissive distinction. In practice, philosophers routinely distinguish between ontologically committing assertions, and assertions which are ontologically innocent. So for instance, ‘There is a possible world where Madonna was never born’ has a reading where it entails *modal realism* about the relevant world. But of course, one can read it in an innocent way as well. But if we want to “regiment” the distinction here, one

would need to specify (in unequivocal terms) the meaning of ‘There is’ on its commissive or loaded use, versus its innocent use. And that requires us to define the commissive use of ‘There is’ in unequivocal terms, for use in our regimentation.

So generally philosophers have a reason to define the commissive use of ‘exist’—or alternatively, the commissive use of ‘There is’ or some other ontic-idiom. Now one proposal here might be to define ‘**exist**’ as follows:

(7) x satisfies ‘**exist**’ iff x is material.

(8) x satisfies ‘**exist**’ iff x is in space and time.

Yet the problem is that, in some possible worlds, certain non-**existing** objects have the status of being “material” and “in space and time.” In a world described by Greek Myth, for instance, Pegasus is a material object that is in space and time—yet still he does not **exist** (even if he exists-in-*that*-world). But this means (7) and (8) are extensionally inadequate. For when we go across possible worlds, (7) and (8) identify certain non-**existing** objects as things which **exist**.⁹ In response, one might specify that, e.g., in (7), we mean **existing** material objects. But this would re-establish the circularity.¹⁰

Perhaps the mistake was in trying to describe objects of a certain *ontological* status by their *nature*, e.g., by describing them as material. Instead, we might try to describe this ontological status directly. But of course, it will not suffice just to define ‘**exist**’ by another equivocal ontic-idiom, as in the following:

(9) x satisfies ‘**exist**’ iff x is real.

⁹ Of course, (8) also is problematic in implying that the mind does not **exist**, even though it refers to it.

¹⁰ Circularity here would not result in completely uninformative definitions; the definitions would teach you, e.g., that **existing** objects are physical. But ‘**existing**’ is doing the real defining work, since it alone is sufficient to define ‘**exist**,’ whereas ‘physical’ and ‘mind-independent’ are not.

After all, ‘real’ on the right-hand side is also equivocal between a commissive use and five-and-a-half non-commissive uses. A similar point apparently holds for other ontic-idioms when they are used to defined ‘**exist**,’ such as ‘actual’, ‘nonfictional’, etc. But in fact, not every ontic-idiom is susceptible to *all* five-and-a-half non-commissive readings [e.g., ‘nonfictionally’ is not used as an intensifier in standard English)]. But it is enough for equivocation if there is more than one way to read the idiom. Accordingly, I submit the following proposed definitions of ‘**exist**’ would also be equivocal:

(10) x satisfies ‘**exist**’ iff x is nonfictional.

(11) x satisfies ‘**exist**’ iff there is a $y = x$.

(12) x satisfies ‘**exist**’ iff x is actual.

For instance, all the idioms on the right-hand side can be relativized to non-actual worlds. E.g., ‘nonfictional’ in (10) has a reading where it is relativized to a world of Greek Myth—and if the idiom is read in that way, then the following is true:

(13) Pegasus is nonfictional.

So Pegasus would satisfy ‘**exist**’ according to (10), since Pegasus *is* nonfictional in a world of Greek Myth. Yet Pegasus does not **exist**, not even at that world. After all, ‘**exist**’ is *commissive*, i.e., ‘**exist**’ is satisfied by **existing** objects and **existing** objects alone.

In the same manner, if the conversation concerns a world of Greek Myth, the following sentences are true:

(14) There is a $y =$ Pegasus.

(15) Pegasus is actual.

And so without further qualification, none of these ontic-idioms can define ‘**exist**’, in the manner suggested by (11) and (12).

III. Lewisian Proposals.

Contemporary metaphysicians are unaccustomed to non-commissive uses of ‘actual’, but it seems that English indeed permits these uses. Even David Lewis (1986) agrees that ‘actual’ can be used non-commissively, i.e., as a “blanket” term for what exists in various possible worlds. But this is of course not how he uses the term:

I myself do not use ‘actual’ as a blanket term...I use it to mean ‘this-worldly’: It is an indexical, relative term, and as used by us it distinguishes our world and our worldmates from all the other worlds and their inhabitants. (p. 99)

So Lewis uses ‘actual’ commissively, but he admits it is a matter of “terminology” (p. 100) and that ‘actual’ has a use as a blanket term.

But in this passage, Lewis may show us a different way to specify the commissive use of ‘actual’ (which would be equivalent to ‘**exist**’). The proposal would be:

(16) x satisfies ‘**actual**’ iff x is in *this* world.

I do not doubt that we understand (16) univocally in most contexts. Even so, (16) itself does not make explicit which world the demonstrative is fixed on. And for that reason, (16) as formulated has multiple readings in English, some of which concern the **actual** world, and some of which do not. [One can imagine a context where a world of Greek Myth is the topic, and (16) introduces a term-of-art for denizens of *that* world.] Now again, none of this shows that (16) is inevitably equivocal *in context*. But it indeed shows that (16) fails to describe in unequivocal terms what ‘**actual**’ means.¹¹

It is notable, moreover, that in some contexts the ostensive definition does *not* suffice for a univocal understanding of ‘actual’, even by English speakers who have some understanding of what the demonstratum is. Indeed, one such case occurs in Lewis’ own

¹¹ The point also holds if we rely on pronominal reference, as in ‘**actual** objects are objects in *our* world,’ since ‘our’ is a contextually shifty term.

discussion of modal realism. At one point, Lewis considers Lycan's (1979) objection that modal realism implies the absurdity that "all possibilities are actual." This would follow from his modal realism if 'real' means "actual." Yet it is here that Lewis says 'actual' is not used as a "blanket term" to cover all possibilities, even though he calls all possibilities "real." He takes it even further and says "Nobody could have thought I *meant* to use either 'actual' or 'world' as a blanket term" (p. 99).

But first, note it is unclear what Lewis means by 'real'. If 'real' is used non-commisively, then it would be *trivial* to say that possible worlds are "real." Yet since modal realism is not supposed to be trivial, it must be that 'real' is used *commisively*, just like how Lewis uses 'actual'. Thus, since modal realism says that possible worlds are "real," it would follow that all possible worlds are actual. So it is understandable if Lewis' critics failed to grasp his use of 'actual'. In order to avoid triviality, it seems his possibilities are meant to be "real" in a *commisive* sense, that is to say, *actual*, though he denies this.¹²

In any case, for our purposes it is important that the confusion with 'actual' arises, even though Lewis defines 'actual' ostensively, akin to (16). For even though the ostensive definition tells us that the "actual" objects are objects in *this* world, and we

¹² In fact, there are non-commisive readings of 'real' which would make modal "realism" non-trivial, e.g., the reading where 'real' is equivalent to 'exists' in (5). For that would amount to saying that possibilities are objects *created in the imagination* of actual people. But of course, Lewis is no fictionalist.

Yet perhaps there is another use of 'real,' which makes modal realism non-trivial. Lewis apparently believes in a sort of "pluriverse" where possible worlds exist as spatio-temporally distinct entities. Thus, the actual world is just as *real* as any world—though of course only the actual world is *actual*. Unfortunately, I cannot properly address this "pluriverse" idea without getting far off topic. But I suspect the idea is intelligible only by the *metaphor* of restricting our discourse to a particular spatio-temporal location. (Cf. Lewis' beer-in-the-fridge metaphor.) But the problem, then, is that possible worlds are not *located*, not even relative to other worlds, since a "common space" for worlds would itself individuate a world on Lewis' view (an "überworld").

In any case, this worry about Lewis is inessential for present purposes. My take-home message here is only that there are some contexts where an ostensive definition for the commisive 'actual' fails to establish a univocal understanding of the term, as seen in the debate between Lewis and Lycan.

have *some* understanding of what Lewis means here, there still arises an issue about what objects are in *this* world. And hence, there arises an issue about which world *this* world is. So in the relevant context, the equivocal phrase ‘in *this* world’ gets the better of us: The ostensive definition thus fails to transmit a univocal understanding of what Lewis means by ‘actual’ in that context. So the ostensive definition of ‘actual’ does not always suffice for a univocal understanding of the idiom.

IV. Another Proposal from Lewis.

I have made this side-trip through Lewis scholarship to bring out that the equivocal nature of the ostensive definition is not an idle concern. One might have thought that *strictly speaking* (16) is equivocal, but within a conversational context it is not. Yet there *are* contexts where it remains equivocal, and philosophically important contexts at that. So continuing the search for an unequivocal definition is not an idle one.

Notably, however, Lewis offers a second way to describe the semantics of a commissive ontic-idiom. In his Preface, he writes:

[S]ome things exist here at our world, other things exist at other worlds...You might say that strictly speaking, only this-worldly things *really exist*; and I am ready enough to agree; but on my view this ‘strict’ speaking is *restricted speaking*, on a par with saying that all the beer is in the fridge and ignoring most of all the beer there is...If I am right, other-worldly things exist...though often it is very sensible to ignore them and quantify restrictedly over our worldmates. (p. 3)

The idea here, then, would be to define a commissive ontic-idiom by construing it as a restricted quantifier—where the quantifier does not range over all possibilities but just over this-worldly things. But as the phrase ‘this-worldly things’ makes clear, restricting the quantifier uses the same strategy that (16) uses in defining ‘**actual**.’ Accordingly, the difficulty with the strategy in (16) carries over to the present case.

However, a second strategy in this passage is Lewis' *iteration* of ontic-idioms in succession. That is, instead of using 'exist' alone, Lewis goes commissive by using two ontic-idioms at once to obtain 'really exist'. This iteration may seem to signal that the resulting expression is univocally commissive.

But in fact, it does not. For the phrase 'really exist' and (similar ontic-phrases) can be relativized to a non-actual world just as much as any one ontic-idiom in isolation. Thus, the ontic-phrases in the following sentences can be relativized to a world of Greek Myth, so that in certain contexts, the following sentences will be true:

(18) Pegasus actually exists.

(19) Pegasus is really nonfictional.

(20) In reality, Pegasus actually is a nonfictional entity that truly exists.

So again, when the topic is Greek Myth, (18) for example will be shorthand for "Pegasus really exists *in a world of Greek Myth*." So we have not yet unearthed an ontic-expression in the language which is unequivocally commissive.¹³

IV. The Significance of the Preceding.

However, Lewis may not be to blame for these circumstances. We have been reviewing a few options for defining the commissive use of an ontic-idiom, in an extensionally adequate, unequivocal, and non-circular manner. But in fact, there is a general argument to the effect that *such a definition is impossible*. First, I think it can be said that:

¹³ Someone might suggest there would be no reason for a speaker to use an iterative construction, other than to identify an ontic-phrase as commissive. So we *should* see these phrases as only having a commissive use. But the fact is that redundancy is always possible in English, even if there is no reason for a speaker to do it. And that means iterating ontic-idioms does not guarantee an ontic-phrase that is univocally commissive.

(21) To define an ontic-idiom unequivocally, one must employ an unequivocal ontic-idiom.

However, if (21) is established, we can then observe in addition that *every* ontic-idiom seems equivocal.¹⁴ For it always seems possible for context to relativize the use of an ontic-idiom to a fiction, as in the Greek Myth examples. After all, in these fiction-centered contexts, it is not as if we are *forbidden* from using some of these idioms. Yet if all ontic-idioms are thus equivocal, then given (21), it would follow that an unequivocal definition of an ontic-idiom (used commissively) is *impossible*.

The argument for (21) is this: Let us say that OI is an ontic-idiom which applies to an object *a*. (Notation: *a*_OI.) Suppose, moreover, that *Z* is an idiom which interprets OI in a univocally commissive manner, in the interpretation ‘‘OI’ applies to *a* iff *a*_Z.’ Now since ‘*a*_Z’ is supposed to be equivalent to ‘*a*_OI’ where ‘OI’ is commissive, this interpretation will be adequate only if ‘*a*_Z’ entails that *a* **exists**. But if ‘*a*_Z’ entails that *a* **exists**, then by definition ‘Z’ is a commissive ontic-idiom (and univocally so).

The impossibility of defining a commissive idiom has, I think, important philosophical consequences. Some of these consequences will be discussed in the next chapter, where I discuss further the non-commissive use demonstrated by (5). But the immediate consequence is that, in fact, there is no way to make explicit the commissive/non-commissive distinction in unequivocal terms. Be that as it may, we of course *understand* this distinction implicitly—it is what we understand when we distinguish the controversial and the uncontroversial reading of ‘Napoleon exists.’¹⁵

¹⁴ Azzouni (in progress) also makes this point.

¹⁵ Azzouni (2007) also emphasizes that terms can be used even if they lack precise definitions (e.g., color terms).

A second consequence is that no sentence of the language would univocally express *Realism* about an entity, property, relation, or what have you. For even a sentence like ‘God really and truly exists’ can be properly asserted by an atheist, if the context is (say) a heated dispute about whether God exists in *Star Trek V*. Conversely, no sentence of the language univocally expresses Anti-Realism about some *x*. Even a Berkeleyan claim like ‘The world is nothing but ideas in the mind of God’ could be asserted by a Realist, if the context relativizes ontic-idioms to a Berkeleyan world.

But again, none of this shows that we cannot *understand* an ontic-idiom univocally, in some contexts. Yet if there is no defining what it is we understand, then our understanding cannot be a matter of knowing an unequivocal definition or interpretation. So the rule for commissively using an ontic-idiom is, apparently, a perfect illustration of Wittgenstein’s (1953) dictum “there is a way of grasping a rule which is *not an interpretation*” (§201). But in the present case, this dictum is not shown by considering skeptical hypotheses for the meaning of unequivocal terms (‘plus’ vs. ‘quus’). For in the case of ontic-idioms, the alternatives are *relevant* alternatives, since the commissive and non-commissive uses actually occur in the language. So *inter alia*, I take the present chapter to provide a more robust argument for the Wittgensteinian dictum.

But the most interesting lesson, I think, is the limitation this reveals on semantic theory. Given an ontic-idiom OI, if the question “What does OI mean on its commissive use?” is raised during the process of “regimentation,” there will be no adequate answer. For any answer will use an equivocal ontic-idiom, as I’ve shown. Some may be willing to overlook this as long as the answer is *understood* unequivocally, yet “regimentation” in

semantics typically aims for unequivocal answers about meaning. And so, in this kind of context, “optimism” about an adequate answer is misplaced. It is another instance of how semantic theory is limited in what it can accomplish.

CHAPTER THREE: SEMANTIC PESSIMISM AND ONTOLOGY

The preceding chapters argued for a view called “semantic pessimism” where one cannot appropriately answer certain questions about a language’s ontology. In the present chapter, however, I want to deal with “ontology” in a narrower sense, since (for reasons I shall present) the domain of the language must contain objects which do not, strictly speaking, exist. Thus somewhat paradoxically, the ontology of a language does not just contain entities we are ontological committed to. But then the question arises as to which entities in the domain we *are* ontologically committed to. What makes an object part of our *ontology*, among the objects in the language’s “ontology”?

Here, I answer this question by offering a novel criterion of ontological commitment. This criterion shall derive from an independently-motivated semantics for negative existentials, i.e., sentences such as:

(P) Pegasus does not exist.

Naturally, since the semantics of negative existentials is no trifling matter, much of the chapter is devoted to this topic. Of course, philosophers like Frege and Russell have addressed this topic before; however these accounts and their difficulties are well-known, and I shall not review them here.¹ Instead, I shall try to think through the problem anew, and develop a different account for a sentence like (P).

¹ For these and related matters, the reader is directed to standard references such as Lycan (2008) and Soames (2003).

Along the way, some limitations are discovered on any semantics for ‘exist’ and other “ontologically loaded” idioms like ‘actual.’ In particular, it is shown that *any* definition here will engage in circularity—and thus, they are doomed to be rather uninformative. Further, I argue that this circularity guarantees that the definitions will be *ambiguous*.² Yet since we obviously can understand ‘exist’ and ‘actual’ unambiguously, I conclude in the spirit of Wittgenstein that there is a way of understanding such terms which does not consist in knowing an interpretation. The Wittgensteinian view then forms the basis for a new criterion of ontological commitment.

I. The Problem and a Proposal.

Historically, the problem with negative existentials is given in something like the following manner. A subject-predicate sentence is true only if the subject-term is meaningful—and a subject-term is meaningful only if it refers to something. But the subject-term (P) does not refer to anything. Thus, ‘Pegasus’ is not meaningful in (P)—and so (P) is not even truth-apt, much less true. But apparently, (P) *is* true—if you search the world high-and-low, you won’t find Pegasus anywhere.

The historical formulation of the problem may be objectionable, however, since it assumes that a subject-term can be meaningful only if it refers to some *existing* thing. Yet ‘Pegasus’ would be precisely the sort of term which falsifies this, since it apparently refers to a nonexistent thing. For this reason, I prefer to put the problem in a somewhat different manner, where the problem is given in an inconsistent triad:

² Here, I use the term ‘ambiguous’ like a logician, rather than a linguist. The logician’s use of ‘ambiguous’ indicates that the reference- or satisfaction-condition of an expression varies from use-to-use, whereas the linguist uses it to indicate that the *sense* or *meaning* of an expression varies. (Of course, if a term is linguist-ambiguous, it is logician-ambiguous, but the converse does not hold.)

- (1) If a sentence of the form $\lceil \Phi(\alpha) \rceil$ is true in the actual world, then the open formula $\lceil \Phi(x//\alpha) \rceil$ is satisfied by some actual object α .
- (2) (P) is a sentence of the form $\lceil \sim\Psi(\text{pegasus}) \rceil$ and is true in the actual world.³
- (3) The formula $\lceil \sim\Psi(x//\text{pegasus}) \rceil$ is not satisfied by some actual object Pegasus.

Individually, (1)-(3) are plausible—but jointly they are inconsistent. And here there is no dubious condition on meaningfulness; rather, the inconsistency turns on the fact that sentences which are *actually* true necessitate that certain formulae are *actually* satisfied.

Toward resolving the inconsistency, I first want to present (what I see as) a crucial observation about the metaphysics of nonexistents. Consider the following:

- (4) Pegasus is a figment of the imagination.

I take it that (4) is actually true. Yet this means, by (1), that there is an actual object o satisfying $\lceil \text{Figment}(x) \rceil$, where $\lceil \text{Figment}(x) \rceil$ is the predicate of (4) (at the level of logical form).⁴ But if this is correct, then what kind of object is o ? Well, since o satisfies $\lceil \text{Figment}(x) \rceil$, then of course o must be a figment of the imagination. Or more broadly, we might say that o is a *merely intentional object*,⁵ or “MIO” for short. Yet since o is actual *ex hypothesi*, o is an actual yet a merely intentional object that satisfies the predicate of (4).

Nonetheless, we should like to say that Pegasus is what satisfies the predicate, as it occurs in (4). If so, then Pegasus = o , since only one object satisfies the predicate, as it

³ Traditionally, ‘exist’ has been viewed as a quantifier, in such a way that (P) is assigned the logical form “ $\sim(\exists y) y = \text{Pegasus}$.” Yet the view that ‘exist’ is a predicate in logical form has re-emerged with Azzouni (2004), where (P) is assigned the form “ $\sim\text{Exist}(\text{pegasus})$.” Thus, I have been attempted to be neutral as to the logical form of (P), leaving open whether $\lceil \sim\Psi(x) \rceil$ is the negation of a quantified or unquantified formula.

⁴ I do not mean to be wedded to ‘figment’ in particular appearing at the level of logical form. You can suppose any predicate you like to denote objects of the imagination.

⁵ The term derives from Brentano (1874), though I am told it originates with the medieval scholastics. Such an object is “merely” intentional as opposed to being, in addition, a mind-independent object.

occurs in (4). Yet since *o* is an actual object, that means Pegasus *is an actual object* which seems absurd. However, the absurdity lessens if we keep in mind that Pegasus is not a *mind-independent* object in our world; rather, he is *merely intentional*. So Pegasus is indeed an actual object albeit a merely intentional one.

Still, something looks amiss. Even if we grant that Pegasus is an actual MIO, it also is true that Pegasus is *not* actual. Is he both actual and nonactual then? No—I submit that the term ‘actual’ is being used in two different ways here. Specifically: It seems that when we say that Pegasus is not actual, the extension of ‘actual’ consists of objects in our world, *though not including* MIOs. Whereas, if we say (truly) that Pegasus is an actual MIO, the extension of ‘actual’ obviously does include MIOs, in addition to other objects in our world. Hereafter, I shall use ‘actual+’ to indicate the latter use of ‘actual,’ whereas I shall put ‘actual’ in bold (“**actual**”) when it is used in the more restricted sense. (There is a hidden circularity here, but as I shall argue in section IV, this is pretty much unavoidable.)

This distinction in the uses of ‘actual’ seems genuine, since in general, it is ambiguous whether MIOs are “actual.” Consider:

- (5) Mirages are actual.
- (6) Hallucinated objects are actual.
- (7) Imaginary objects are actual.

Apparently, each of (5)-(7) can be read as either true or as false. On the true reading, for example, (7) says that figments of the imagination are part of our world (given that we have an imaginative faculty). But on the false reading, it says that these imagined objects are *not* merely intentional in our world. The ambiguity also seems to occur in (5) and (6)

as well. And in general, it seems that MIOs are “actual” in one sense of ‘actual,’ but not in another sense.

If we suppose ‘actual’ is equivocal along these lines, we can say that ‘Pegasus is actual+’ and that ‘Pegasus is not **actual**’ without contradiction. He is “actually nonactual” by which we signify that Pegasus is actually+ a MIO. Accordingly, we can resolve the inconsistent triad by refining (1)-(3) in the following manner:

- (1*) If a sentence of the form $\lceil \Phi(\alpha) \rceil$ is true in the actual+ world, then the open formula $\lceil \Phi(x//\alpha) \rceil$ is satisfied by some actual+ object α
- (2*) (P) is a sentence of the form $\lceil \sim\Psi(\text{pegasus}) \rceil$ and is true in the actual+ world.
- (3*) The formula $\lceil \sim\Psi(x//\text{pegasus}) \rceil$ is not satisfied by some **actual** object Pegasus.

But to be clear:

- (8) $\lceil \sim\Psi(x//\text{pegasus}) \rceil$ is satisfied by an actual+ object Pegasus.

Thus, we can say that (P) is true in the actual+ world, in virtue of an actual+ object satisfying the predicate—without committing to the **actuality** of Pegasus.

Accordingly, none of this vindicates (P) as **actually** true. But this is how it should be: If we lived in a world containing only **actuals**, a term which denotes nothing **actual** would be meaningless to us. In that case, (P) would be like the sentence ‘Blurgaflurg does not exist,’ since both sentences would use a term which does not even refer to a MIO in that world. Of course, in *our* world (P) is true, and that means that *our* world (somewhat paradoxically) is not the **actual** world, strictly speaking. Yet by this, I just mean that our world also contains MIOs besides **actuals**—our world is home to all *actual+* objects, and not just the **actual** ones.

II. The Metaphysics of MIOs

Admittedly, however, MIOs are strange kinds of objects. But they do not have to be *that* strange. We can think of the situation like this: In creating fictions, we consider a set of descriptors, which are not jointly satisfied in the **actual** world. Nonetheless, we can *imagine* some object satisfying the descriptors, and the result is an actual+ MIO. It may be very mysterious how the mind is capable of *imagining* an object, since it almost as if the mind can create certain objects at will. Regardless, it is clear that we *do* imagine objects, , even though it may be mysterious what that consists in.

Some may protest that my solution to the inconsistency smacks of Quine's (1948) nemesis McX, who thinks that the object of 'Pegasus' is just the *idea of Pegasus*. But I do *not* mean to say that Pegasus, the MIO, is identical to the idea of Pegasus. Rather, I meant to say that the MIO is the object of the term 'Pegasus' *as well as* the object of the idea.

Yet since Pegasus is merely an artifact of the imagination, there is a sense in which Pegasus is "of the mind." But it is clear he is not an *idea* since it is not as if Pegasus *represents* an object in the way that an idea does. Of course, in the vernacular a person might say that Pegasus is "just an idea." Yet this is misleading: Not only does it encourage us to think of Pegasus himself as representing an object, but it encourages the McXian idea that Pegasus is an idea representing *Pegasus* in particular.⁶

But if Pegasus is not an idea strictly speaking, what is he? Granted, he is a product of the imagination, and so, is "of the mind" in that sense. But more on the metaphysics of MIOs is desirable—and indeed, I wish I could say more. However, my primary aim was just to resolve the inconsistent triad at (1)-(3). And to this end, it suffices to note that

⁶ Thanks to William Lycan for bringing this point to my attention.

there really are MIOs in our world, figments of the imagination being a prime example. After all, who would deny that Pegasus is a figment of our imaginations?

Still, it is important what has *not* been said about Pegasus. Specifically, in saying ‘Pegasus is actual+,’ I am not committed to some kind of Meinongian worldview. For in asserting ‘Pegasus is actual+,’ I do not mean to suggest that there is a Platonic third-realm where Pegasus lives as an *abstract, mind-independent* object. Rather, I mean that Pegasus lives in the mind, as it were, as an actual+ MIO. And once we are clear on this, it seems quite plausible to say that Pegasus is actual+ (as a merely intentional object).

Of course, eliminative materialists will not go along with MIOs, but they would reject a discussion of semantics in the first place. Yet other naturalists may also be troubled, since especially now I might sound like some kind of dualist. Lately, however, I have become wary of the ‘dualism’ label. In one sense, everyone is (or should be) a “dualist” since there is a distinction between mental and nonmental phenomena. But this is a trivial kind of dualism. It is somewhat like being a dualist about elephants, on the grounds that there are African and Asiatic elephants.

But if dualism proper is a claim about what exists (where ‘exists’ can be exchanged with ‘is actual’), then the dualist position is an *ambiguous* one. I am *not* a dualist, if dualism is the claim that the **actual** world consists of physical objects and MIOs. But if the actual+ world is the concern, I *am* a dualist—dualism here would be true *by the definition* of ‘actual+’. One then might reply that there is only one sense of ‘actual’, but this forgets the linguistic evidence which suggests otherwise.

Alternatively, naturalists may be mystified by how ‘Pegasus’ comes to *denote* a MIO. But the project of naturalizing denotation presents many difficulties for nonempty

names as well. And what's important, for my purposes, is that the name 'Pegasus' *does* denote a MIO, even if it is far from clear *how* the name does this.

Still, if Pegasus is a MIO, it seems we also want to say that:

(9) Pegasus is a horse with wings.

But, an objector may say, Pegasus could not be *both* a MIO *and* a horse with wings. For MIOs are not horses; MIOs are "of the mind" whereas horses are not.

However, (9) is plausibly shorthand for:

(10) Pegasus is a **nonactual** horse with wings.

After all, in asserting (9), we do not mean to say:

(11) Pegasus is an **actual** horse with wings.

Though, it would be appropriate to say:

(12) Pegasus is actually+ a non-**actual** horse with wings.

Thus, I deny (9) if 'horse' means "**actual** horse;" yet I do want to say that Pegasus is a non-**actual** horse—where this horse is an object of the imagination. And if Pegasus is a non-**actual** horse, then in this sense he *is* both a MIO and a horse.⁷

Yet does this mean both **actual** and non-**actual** objects are in the extension of 'is a horse'? That may seem odd—we usually think that this extension only contains **actual** horses. Of course, one can also ask whether Pegasus is even a *horse*, since it is unclear whether he is of the biological kind *equus caballus*—but let that pass. The main question is whether non-**actuals** are in the extension of nonempty predicates like 'is a horse.'

⁷ Kripke (1972) also observes that 'Pegasus' names a fictional entity (though it does not actually name anything). In our terminology, this would be the observation that 'Pegasus' names an actual+ entity but no **actual** entity. But because names are "rigid" on his view, Kripke thinks 'Pegasus' (in our language) *necessarily* names a non**actual** object. However, even on Kripke's view, this follows only if 'Pegasus' names an object that is *essentially* non**actual**. And it is not obvious that this is true of what 'Pegasus' names. We can follow Kripke here if preferred, but the view does not seem compulsory. One might instead say that 'Pegasus' names a *winged horse* essentially (though not necessarily a fictional one).

In response, I think it depends on our domain-of-discourse, as to whether non-**actuals** figure into these extensions. Naturally, if this domain is the set of **actual** objects, then only **actual** horses will be in the extension of ‘is a horse.’ But in such a domain, (9) has no proper interpretation, thanks to the name ‘Pegasus.’ So if we want to interpret the sentence, we will need to go to a domain which has Pegasus as a member. And in this domain, Pegasus can indeed be in the extension of ‘is a horse.’ (If this still seems odd, it may just reflect that fictional discourse is not our paradigm of discourse in general.)

Nonetheless, if you think Pegasus should *never* be in the extension of ‘is a horse,’ then you can define ‘is a horse’ accordingly, even in a Pegasus-domain. But there, (9) will come out *false*. Indeed, I think that there is a reading of (9) where it is false [assuming that (11) articulates one way of reading (9)]. But the point is that one cannot insist on the truth of (9) *while excluding* Pegasus from the extension of ‘is a horse.’

In sum, if we hold that Pegasus is a horse *and* a figment of the imagination, as I think we should, it follows that some figments of the imagination are horses. This may sound odd in one sense, but we must remember that these horses are not **actual** horses but rather non**actual** horses, where non**actual** horses are actually+ MIOs.

III. Modality and Negative Existentials.

Although ‘**actual**’ and ‘actual+’ are satisfied only by objects in *our* world, we can also acknowledge “world-relative” uses of ‘actual,’ akin to Lewis (1986), though ‘**actual**’ and ‘actual+’ would *each* have world-relative uses. This would help us understand the following as true when uttered by Macbeth:

(13) The dagger I hallucinated was not actual, though the murder weapon was.

Consider: If ‘actual’ in (13) is indexed to Macbeth’s world, then we can understand it as asserting (correctly) that only the murder weapon is a *bona fide* part of Macbeth’s world.

But there is a sense in which the hallucinated dagger is part of that world as well. After all, its presence makes Macbeth’s world different from a world in which Macbeth never hallucinates. Thus it appears we have two uses of ‘actual’ in *Macbeth’s* world as well, corresponding to the **actual**/actual+ distinction in our world. In his world, as well as ours, the distinction turns on whether an object is a MIO in that world.

Indeed, we can characterize all these uses of ‘actual’ just by distinguishing the notion of “**actual**-in-a-world” from “actual+-in-a-world.” In this, being **actual** would be a special case of being “**actual**-in-a-world,” where the world in question is ours. On the other hand, the hallucinated dagger is “actual+-in-Macbeth’s-world,” though it is non**actual**-in-his-world. Though interestingly, the hallucinated dagger is also actual+-in-*our*-world, since it is something created in our imaginations—though it is not something we ourselves *hallucinate*.⁸

With this in place, we can consider negative existentials which are true in worlds *other than ours*. Consider, for example, the fortunate world where the following is true:

(14) George W. Bush does not exist.

⁸ Many problems are avoided by the fact that one can imagine the very same MIO which another person hallucinates. An example by Thomas Hofweber: Suppose you are told that a certain drug causes hallucinations of a mythical creature called “Tedasus.” If you are uninitiated in the drug, you will still understand this report, even if you lack first-hand experience of this vision. Yet on the present view, such experience seems required if ‘Tedasus’ is to avoid the fate of ‘Blurgaflurg.’ However, in this case we must distinguish our *imagining* from our *hallucinating* Tedasus. To be sure, the two experiences may be different, since only the latter may involve a visual image of the creature. But even so, the object imagined can be the same as the object hallucinated. For the creature is not the *image* experienced (if any) but rather the *object* of these images.

If (14) is true in a world, then ‘*x* does not exist’ must be satisfied by an object in that world which is *GWB*. But *GWB* would not be **actual**-in-that world, assuming ‘exist’ covers **actual** objects. So he must be actual+-yet-non**actual**-in-that-world.

Now *if* there are worlds where people (somehow) imagine *GWB*, there will be no new problem here. We can say that (14) is actually+-true-in-that-world, in virtue of *GWB* being a MIO in that world. What is new, however, is that (14) will uncontentiously be true at *GWB*-less worlds *where no one has even imagined GWB* (or hallucinated him, etc.). And so at those worlds, there will be no appropriate truth-maker for (14), not even a MIO. So absurdly, it seems (14) cannot be true at those worlds.⁹

In reply, we first need to distinguish worlds which contain (14) *qua* sentence of English, and worlds which do not. To take the former case: If a world contains (14) as an English sentence, then (14) is meaningful at that world (since sentences *qua* English sentences have meaning). And so by compositionality, the name ‘George W. Bush’ is meaningful in that world. But if ‘George W. Bush’ is meaningful in that world (i.e., if it is not like ‘Blurgafflug’), then the name must denote at least a MIO in that world. So if we assume that this world contains (14) as a sentence of English, then there must be at least an imaginary *GWB* in that world, *contra* assumption. Thus, the new problem is not a coherent one in worlds containing (14) *qua* sentence of English.

However: If a world does not contain the English sentence, it is ambiguous as to whether it will be true *in* that world. It would not be a true English sentence *found in* that world, since *ex hypothesi* the world now does not contain the sentence. Nevertheless, we can say (14) is true *at* a *GWB*-less world—after all, in *our* world, the subject-term denotes an **actual** object which is nowhere found in *that* world. And in saying this, the

⁹ My thanks to Meg Wallace for bringing this issue to my attention.

new problem does not arise. For (14) is true *at* that world, not because GWB makes it true in the GWB-less world, but because GWB in *our* world is not found in the GWB-less world.

IV. Some Limits on Interpretation.

There is a further difficulty with our treatment of negative existentials, when we defined ‘actual+’ and ‘**actual**.’ One can distill these definitions to the following:

(15) An object satisfies ‘actual+’ iff it is **actual** or merely intentional.

(16) An object satisfies ‘**actual**’ iff it is **actual**.

The issue is that the definition of ‘**actual**’ is clearly circular. This makes the definition uninformative; it tells you what is in the extension of ‘**actual**’ only if you already know. Of course, commonsensically we *do* know what is in its extension. Still, the circularity is frustrating if we want an informative account of what is *common* to **actuals**.

So instead, we might offer a different definition of ‘**actual**’ which avoids circularity. Consider:

(17) An object satisfies ‘**actual**’ iff it is not merely intentional.

Alternatively, one might propose:

(18) An object satisfies ‘**actual**’ iff it is mind-independent.

(19) An object satisfies ‘**actual**’ iff it is physical.

But in the first instance, these definitions select objects by their *metaphysical* rather *ontological* status. That is, the extension is defined according to the *nature* of objects—not according to whether objects really exist. Because of that, these will be inadequate to define ‘**actual**.’ For in some possible worlds, certain non-**actual** objects have the status

of being “not merely intentional” or “mind-independent” or “physical.” In a world described by Greek Myth, for instance, Pegasus is mind-independent, physical, and is not a merely intentional object. Yet still he is not **actual** (though he is **actual-in-that-world**).¹⁰ But this shows (17)-(19) are inadequate as definitions of ‘**actual**’; for when we go across possible worlds, these definitions count certain non-**actuals** as **actual**. Of course, one might clarify that, in (17)-(19), we have in mind **actual** mind-independent objects, **actual** physical objects, or **actual** objects that are not merely intentional. But this would be to use ‘**actual**’ in the definition of ‘**actual**.’¹¹

So if the metaphysical-idioms fail, we might go to an idiom that is clearly ontologically-loaded such as ‘exist,’ or better, ‘truly exist:’

(20) An object satisfies ‘**actual**’ iff it truly exists.

Yet as with the metaphysical idioms, ‘truly exist’ can apply to non**actual** objects in other worlds, e.g., (P) is true at some worlds, even though Pegasus is non**actual**. And so, it is non-equivalent to ‘**actual**’ across possible worlds. (Of course, in (20) we were probably thinking of objects that **actually** exist, but this reintroduces the circularity.)

Another proposal may be gleaned from David Lewis (1986). The idea would be to define ‘actual+’ ostensively, and then use ‘actual+’ to define ‘**actual**:’

(21) An object satisfies ‘actual+’ iff it is an object in *this* world.

(22) An object satisfies ‘**actual**’ iff it is actual+ and not merely intentional.

¹⁰ This kind of consideration also poses a problem for the definition of ‘actual+’ at (15), where Pegasus potentially gets *excluded* from the extension of ‘actual+.’

¹¹ These would not be completely uninformative; they would teach you, e.g., that **actual** objects are mind-independent. But ‘**actual**’ is doing the real defining work, since it alone is sufficient to define ‘**actual**,’ whereas ‘mind-independent’ is not.

Yet crucially, the demonstrative in (21) must be “rigidified”—it must be fixed on the demonstratum it has in the actual+ world. Otherwise, in a world with different MIOs than ours, it will count the MIOs in *that* world as actual+. But only MIOs in *our* world are actual+.^{12, 13}

However, the “rigidified” demonstrative is clearly defined as fixed on its *actual+* object, when we take its containing sentence to other possible worlds. Yet this definition of a “rigidified” demonstrative uses ‘actual+;’ hence, we cannot use such a demonstrative when defining ‘actual+,’ at least not without circularity.

There are a number of other idioms one might use in defining ‘**actual.**’ Idioms such as ‘real’ and ‘nonfictional’ come to mind. These form a family of idioms, along with ‘exist’ and ‘actual,’ which we might call “ontic-idioms”—idioms that attribute a “robust ontological status” to objects. But it seems ontic-idioms can always be rightly used in relation to **nonactual** objects. Thus, in the world of Greek Myth, it is true that:

(23) Pegasus is real.

(24) Pegasus is nonfictional.

Even if we amass ontic-idioms together, the result will be true in a world of Greek myth:

(25) Actually, Pegasus is nonfictional.

(26) Pegasus really and truly exists.

(27) Pegasus actually is a nonfictional entity that truly exists in the real world.

¹² The point also holds if we rely on pronominal reference, as in ‘actual+ objects are objects in *our* world,’ where ‘our’ must rigidly designate *us*.

¹³ It may be problematic to speak of “MIOs not of our world.” If that phrase is to avoid the meaninglessness of ‘Blurgafllurg,’ it must denote something actual+. But by definition, it would denote MIOs that are not actual+. So it seems the phrase denotes MIOs that are both actual+ and not actual+; contradiction.

Yet even if we cannot *speak* of nonactual+ MIOs, they may be metaphysically possible. If so, then the point would still hold that the demonstrative must be rigidified in (12), so that nonactual+ MIOs do not get counted as actual+. (In fact, I think further issues remain even here, but I shall not pursue them further.)

So these constructions also cannot define ‘**actual**,’ for in some possible worlds, **nonactuals** satisfy those constructions.

Surprisingly, there are even contexts in *our* world where ontic-idioms correctly apply to **nonactuals**. Suppose someone is adamant that the fairies in *A Midsummer Night’s Dream* are ruled only by a Queen, Titania. Then, you might stammer in reply:

(28) *There really is a King of the fairies, Oberon.*

Here you have said something true, though you have asserted the existence of something which is not **actual**. Or in discussing the characters in the “play-within-a-play” (Act V.i), if someone avows that there is no lion, you may respond:

(29) *I’m not making up this up; there really and truly is a lion, Snug.*

Again, in the right context, you would be saying something true, even though you have asserted the existence of something that is not **actual**, not even in Titania’s world.

Speaking generally, the problem is that different objects in different situations will qualify as having “ontological status.” Thus, the idioms which attribute such status will apply to different objects in different contexts. Since some of these objects will be *merely* possible objects, such idioms cannot define ‘**actual**.’ For regardless of what scenario we are dealing with, ‘**actual**’ is satisfied by **actual** objects alone.

Here is a final attempt to get rid of circularity. Let us say that boldface-type indicates the *commissive* use, so that a sentence stating that the object is **actual** entails that it *actually exists* (insert foot stomp). This use of ‘actual’ differs from the various noncommissive uses where this entailment does not result, as with ‘actual+.’ Now the above remarks show, in effect, that ‘physical,’ ‘real,’ ‘nonfictional’ etc. can be used noncommissively, in addition to their commissive uses. ‘Pegasus is real,’ we saw, can be

read as not entailing the *actual existence* of Pegasus, though of course it could be read that way too. If so, then we might define the set of **actual** objects as the set of objects that are real (nonfictional, worldly, etc.), where the relevant idiom is read commissively.

However, the commissive uses are defined by whether the idiom is satisfied only by objects that *actually exist*—where ‘*actually exist*’ is co-extensional with ‘**actual**.’¹⁴ But in saying this, we already need to have the extension of ‘**actual**’ defined before we can define the commissive reading of this idiom. And that means we cannot define without circularity the extension of ‘**actual**’ using an ontic idiom, read commissively.

I have been canvassing a few options for defining ‘**actual**’ in an extensionally adequate, non-circular manner. However, Chapter Two presented a general argument to the effect that *there is no way* to define ‘**actual**’ in this fashion. In particular, it was shown that if we are to define/interpret an ontic-idiom on its commissive use, we must employ a commissive ontic-idiom. If so, this means circularity will always be an issue. (Either the idiom will interpret itself, or it will be interpreted by some idiom which interprets itself, or several of the idioms will be interdefined.)

Earlier I said that circularity is a problem because it makes for a relatively uninformative definition. But circularity turns out to even more serious in the present case. If ontic-idioms have commissive and non-commissive readings, this means ontic-idioms are *ambiguous* taken as they are. Yet as the previous paragraph shows, if we want to interpret an ontic-idiom as commissive, then we must use a commissive ontic-idiom. But all ontic-idioms are ambiguous on the commissive/non-commissive dimension. (After all, it is not as if we are banned from using certain idioms in writing a novel.)

¹⁴Interestingly, an ontic-idiom could conceivably be “non-commissively commissive” iff a sentence using the idiom entails that the relevant object is actual+. The truth of the sentence would guarantee that the object is in *our* world, rather than some other possible world, though not that the object is **actual**.

Hence, it seems any attempt to disambiguate an ontic-idiom will use an ambiguous ontic-idiom, meaning that the disambiguation will fail. And if we try to disambiguate the second idiom, we shall use a third ambiguous ontic-idiom—and thus a regress. So apparently there can be no explicit disambiguation of ontic-idioms.

V. Wittgensteinian Understanding and Primitive Ontic-Idioms.

Nonetheless, we obviously can *understand* an ontic-idiom as unambiguously commissive. But if there is no disambiguating an ontic-idiom, then our understanding cannot be a matter of knowing an unambiguous interpretation. In this, our view of ontic-idioms is congenial to Wittgenstein’s (1953) claim about usage-rules for idioms: “there is a way of grasping a rule which is *not an interpretation*” (§201). So even though we may not be able to *explicitly* disambiguate an ontic-idiom, we can still *implicitly* understand such an idiom unambiguously, on an occasion of use.

Yet if the Wittgensteinian view is correct, then in one sense we *can* interpret ‘**actual**’ as unambiguously commissive. Namely, if we utilize an ontic-idiom which is *understood* (in the Wittgensteinian sense) to be commissive, then an interpretation of ‘**actual**’ can be understood unambiguously—though the interpretation *per se* will be ambiguous, strictly speaking.

In the spirit of this proposal, one might suggest:

(30) An object satisfies ‘**actual**’ iff it is *real*.

Here, ‘*real*’ is undefined, though understood commissively in the present context. This seems legitimate for our purposes, since our aim was not to define commissive ontic-idioms *in general*. We just wanted a non-circular (and unambiguous) definition of

‘**actual**’ in our account of negative existentials. So we might let (30) stand, and just rely on the contextual cues (e.g., the use of italics) to prompt speakers into understanding ‘*real*’ commissively.¹⁵

But in taking ‘*real*’ as primitive, one may object that the Paradox of Analysis (or some variant) seems particularly compelling. Since ‘*real*’ is (roughly) just another word for ‘**actual**,’ then interpreting ‘**actual**’ by ‘*real*’ seems fairly uninformative. It would be like defining ‘moral person’ as follows:

(31) x satisfies ‘moral person’ iff x is an ethical person.

It is informative to be given (31) in that one at least can deduce that the term ‘moral person’ is interchangeable with the term ‘ethical person’ (in extensional contexts). Yet beyond this, the informative worth of (31) seems fairly trifling. Certainly no moral theorist would tout this as a worthwhile account of what moral persons are.

Still, in some cases this issue does not arise: ‘H₂O’ interprets ‘water’ informatively, even though ‘H₂O’ may be “just another word” for ‘water.’ Yet these terms clearly have a different “mode of presentation” (“MOP”).¹⁶ ‘Water’ “presents” a clear, odorless liquid (around here)—while ‘H₂O’ “presents” a certain kind of chemical compound. Of course, ‘H₂O’ cannot be used to give the meaning of ‘water,’ if the MOP constitutes part of a term’s meaning. But in the present discussion, I am not concerned to define the *meaning* of terms—it is enough if we can define *extensions* in an informative way. And ‘H₂O’ can informatively define the extension of ‘water.’ Indeed, it is precisely

¹⁵ Of course for Wittgenstein, our understanding of ‘**actual**’ would not be a matter of knowing (30) either. (30) is only meant to fill out the account of negative existentials, in as much as this is possible.

¹⁶ As argued by Fodor (1998), a MOP is apparently not a term’s intension, since the MOP does not fix extension. Thus ‘beech’ and ‘elm’ have disjoint extensions, though they may “present” identical-looking trees. Still, it remains mysterious what a MOP is, but I assume we can commonsensically judge sameness and difference of MOP in the present cases, even if we do not have a filled-out metaphysics of MOPs.

the difference in MOP which makes ‘ ‘Water’ denotes H₂O’ informative—this teaches us that the word for a clear, odorless liquid denotes a certain kind of chemical compound.

Yet the point is, unlike ‘water’ and ‘H₂O,’ the difference in MOP between ‘moral person’ and ‘ethical person’ is fairly thin. At worst, there is no difference in MOP—both just “present” people who follow right rules of conduct (or some such thing). Similarly, although (30) at least informs us that the sign-designs ‘**actual**’ and ‘*real*’ are interchangeable, the informative worth of (30) seems negligible.

Apparently, then, there is no substantial way to define the extension of ‘**actual**.’ Since ‘**actual**’ is a commissive ontic idiom, it must be interpreted by a commissive ontic idiom, as was demonstrated earlier. But if (30) is any indication, this interpretation will be fairly uninformative, even if the second idiom is understood commissively.

Certainly, one could define ‘**actual**’ extensionally by a list of objects, though given the number of **actual** objects, such a definition would (of practical necessity) be incomplete. So if our hope is for an extensionally adequate, informative interpretation of ‘**actual**,’ it seems we will be frustrated. And if an account of negative existentials requires the use of ‘**actual**,’ as I think it does, any account will have to suffer this limitation.

VI. Ontological Commitment.

As advertised, I shall now illustrate how the foregoing remarks influence a criterion of ontological commitment. I shall argue here that (i) ‘*y* is ontologically committed to *x*’ will be ambiguous in the same ways that the ontic-idioms are, (ii) as a result, there can be no criterion of ontological commitment in the standard sense of

‘criterion,’ yet (iii) there can be criterion of ontological commitment in a nonstandard Wittgensteinian sense.

First, I assume it is fairly uncontentious that:

(32) A person is “ontologically committed” to an object x iff the person accepts (if only tacitly) that x exists, in some sense of ‘exist.’

Now it may seem that we would want to specify a commissive reading of ‘exist’ in (32). Indeed, I think this correctly defines ontological commitments for most purposes. Yet it seems clear that ‘exist’ in (32) should sometimes be read noncommissively. For instance, it is plausible to say that:

(33) Macbeth is ontologically committed to his murder weapon.

Yet (33) is true not because Macbeth holds that his murder weapon **exists**, i.e. that it is an object of *our* (nonfictional) world. Rather, (33) is true because Macbeth accepts (at least tacitly) that the murder weapon **exists-in-his-world**, the world of the Shakespearean play. (Of course if asked, he might *say* that the weapon “exists” in the commissive sense. But interpreted charitably, this remark uses the noncommissive ‘commissive;’ see n. 12.) The upshot is that, if accepting x ’s existence-in-Macbeth’s-world sometimes suffices for ontological commitment, then ‘exist’ should not always be read commissively in (32).

This noncommissive variety of ontological commitment is exemplified in different kinds of cases as well, e.g.

(34) The author of this work is ontologically committed to MIOs.

There is a reading of (34) where it is true. But if ‘exist’ were read commissively in the definition at (32), the statement at (34) would be false, I assure you. For the author is not committed to the **existence** of Pegasus, though he is ontologically committed to Pegasus

as a MIO in our world, the actual+ world. But again, this indicates that in (32), ‘exist’ should not always be given the commissive reading.

Plausibly, what’s going on here is that ‘*y* is ontologically committed to *x*’ is context-sensitive. In the standard cases, only an acceptance of *x*’s **existence** will suffice for ontological commitment—thus, paradigmatically, only the theist is ontologically committed to God. But (33) and (34) occur in contexts where acceptance of *x*’s existence-in-Macbeth’s-world or existence+ suffices. Thus we might propose:

(35) A person is “ontologically committed” to an object *x* iff the person accepts (if only tacitly) that *x* exists, in a sense of ‘exist’ selected by context.

Ideally, however, we would now describe the principles and parameters by which context selects a reading of ‘exist.’ Without this, a contextualist view is not so much a *bona fide* account of ontological commitment, but the promise of such an account. However, entering into the details of the pragmatics is beyond what I can accomplish here. Suffice it to say that ‘*y* is ontologically committed to *x*’ is indeed context-sensitive, though the details of this context-sensitivity may be unknown at this stage.

But whatever the pragmatic details, other sections show what we *cannot* accomplish in an account of ontological commitment. Specifically, if a context assigns ‘exist’ a *commissive* reading on an occasion of use, our account will be unable to describe unambiguously this reading of ‘exist.’ And further, if context reads ‘exist’ as ‘exist+,’ this also seems to be something we cannot describe unambiguously.

Crucially, however, this also indicates that no filled-out version of (35) can stand as a *criterion* of ontological commitment, if such a criterion must be free from ambiguity. Again, (35) uses an ontic-idiom, and by the previous section, all ontic-idioms are inevitably ambiguous, strictly speaking.

Nonetheless, this problem isn't special to (35). I assume that any adequate criterion of ontological commitment will make use of an ontic-idiom (directly or indirectly) at some point. But then, such criteria will be equivocal, insofar as ontic-idioms are equivocal. And that means any criterion will fail as a criterion of ontological commitment, if such criteria need idioms to be explicitly disambiguated.

Still, I think (35) can function like a criterion, for practical purposes. For we can univocally *understand* ontic-idioms (in Wittgenstein's sense), even if we cannot univocally articulate this understanding. Thus in paradigm cases, we can understand unequivocally that ontological commitment a matter of accepting that something is **actual**. And so (35) would function a kind of criterion, though we should acknowledge it does not offer an *unambiguous formula* which such criteria are standardly thought to provide.

VII. Closing Remarks.

In this chapter, I have defended a solution to the problem of negative existentials, and pulled from this a new criterion of ontological commitment, where we understand 'exist' in (35) only in the Wittgensteinian sense of 'understand'. From the present point-of-view, then, a pessimistic view emerges of philosophical ontology, the project of specifying what exists. If there is no substantive, unambiguous interpretation of ontic-idioms, then it seems there is no unequivocal, substantive answer as to what **existing** objects have *in common*. Thus, if philosophical ontology hoped to win to this kind of answer, its hopes will be frustrated.

Nevertheless, the upshot of this chapter is not all negative. Since we are competent in using ontic-idioms commissively, we exhibit a kind of “understanding” of what **exists**. But this will be a kind of Wittgensteinian understanding. It will be an understanding which does not include knowledge of answers to what **existing** objects have in common—nor to which particular objects **exist** (when the model-theoretic considerations are in play). This kind of understanding will just reflect our linguistic competence with ontic-idioms, rather than some philosophical insight into what the world is really like.

FOURTH CHAPTER: SEMANTIC PESSIMISM AND TRUTH-CONDITIONAL SEMANTICS

In the First Chapter, I articulated the pessimistic view that no interpretation answers the question of what a term denotes, when certain model-theoretic alternatives are under consideration. At best, we can get a kind of “counterfeit” answer to this question, formulated in terms that are used unquestioningly. But although the pessimistic approach indicates a certain limit on our theories of language, I think this awareness also helps dispel certain problems for truth-conditional semantics (“TCS”).

The chapter proceeds by first abductively arguing that sentential meaning is to be understood *at least partly* in terms of truth-conditions. (This is done in light of the fact that, as far as I can tell, the view has no obvious deductive argument from uncontroversial premises.) I then raise the well-known objection that, even under this weak version of TCS, we are forced to characterize truth and meaning in a circular manner (the “Tarski-Davidson circle”). It is at this point I appeal to semantic pessimism to show that in lieu of a circular account, we can instead give a non-circular and informative account of *all* semantic notions from one language, if we use a different background language for the account.

I. The Determination Argument.

Perhaps there are many ways in which one could argue for TCS, but the best case for the view, I take it, stems from what Bar-On et al. (2000) call the “Determination

Argument.” The argument here is attributed to David Lewis (1972) based on the following passage:

In order to say what a meaning *is*, we may first ask what a meaning *does*, and then find something that does that.

A meaning for a sentence is something that determines the conditions under which the sentence is true or false. It determines the truth-value of the sentence in various possible states of affairs, at various times, at various places, for various speakers, and so on. (p. 22, italics original)

The argument which Bar-On et al. pull from this passage is (essentially) the following:

- (1) Given a possible world, a meaning for a sentence determines its truth-value in that world. [Assumption]
- (2) So, a sentence-meaning is a function from possible worlds to truth-values. [From (1)]
- (3) Such a function is a truth-condition (as conceived in intensional logic). [By Definition]
- (4) So, a sentence-meaning is at least a truth-condition. [From (2)-(3)]

Lewis is of course a believer in TCS, though I am a bit unclear on whether the particular passage above expresses this argument. But regardless, Bar-On et al. nonetheless do an admirable job of defending the argument from several objections. And indeed, the argument does seem to capture something about what makes TCS appealing. The first premise is just the virtual truism that a sentence’s meaning determines whether it is true in a possible world. This makes meaning a function from sentences to truth-values, and such a function just is a truth-condition (in intensional logic). Admittedly, however, (4) is weaker than what is usually claimed in Davidsonian truth-conditional semantics, viz., that a sentence’s meaning is *nothing more than* a truth-condition. But (4) would at least show that truth-conditions are nonetheless essential to sentential-meaning.

As Bar-On et al. are aware, however, the present formulation of the determination argument occurs in the paradigm of Carnap's (1971) intensional logic. Yet the authors suggest that there is an epistemic version of the argument that does not presuppose such a logic (n. x). But before considering this, it is useful to consider whether the basic idea in the non-epistemic version can be captured without assuming Carnapian logic. Suppose that instead of (3), we just think of a truth-condition (rather trivially) as a condition under which a sentence is true. Thus the truth-condition for 'Snow is white' (in English) would just be the condition or state-of-affairs where snow is white (where this condition occurs in several possible worlds). Note that you do not need to accept that such a condition is a metaphysical robust condition in a really existing world—the idea is simply that, whatever states-of-affairs are, the state-of-affairs where snow is white is the truth-condition for 'Snow is white' (in English). Then, one may plausibly suggest:

(5) A meaning for a sentence determines the condition under which the sentence is true (i.e., its truth-condition). [Assumption]

(4R) So, a meaning for a sentence is at least a truth-condition. [From (5)]

The premise is plausible if we accept the natural thought that a sentence's truth-value is determined entirely by its meaning and the states-of-affairs in a world. For since non-linguistic states-of-affairs *themselves* do not determine whether they are the truth-maker for some sentence, it must be that the sentence's meaning does.

But the problem is that the inference from (5) to (4R) is patently invalid. It is valid only if we add the premise that:

(6) If x determines y , then x is at least y .

But this premise is susceptible to counter-example. For instance, the radius of a circle determines the diameter, but it does not follow that the radius is constituted at least by the

diameter. [Of course, the diameter also determines the radius and the diameter is at least the radius—but we only the earlier instance to show (6) is false.]

So at this point, it seems the determination argument is either controversial *qua* piece of Carnapian logic, or it ends up requiring a false premise. So it seems to me that TCS would need some other version of the argument to support it. It is here that the epistemic version of the argument in Bar-On et al. becomes crucial.¹ Consider:

- (1.1) If you know a sentence's meaning and you are omniscient as regards nonlinguistic fact in a world,² then you know the sentence's truth-value in that world.
- (2.1) So, knowing a sentence's meaning is *at least* knowing enough to assign the sentence a truth-value in a world, given omniscience about nonlinguistic facts in that world. [From (1.1)]
- (3.1) To know enough to assign the sentence a truth-value in a world, given omniscience about nonlinguistic facts in that world, is to know a truth-condition. [Assumption]
- (4.1) So, knowing a meaning is at least knowing a truth-condition. [From (2.1), (3.1)]

Bar-On et al. recognize that premise (3.1) is not definitional in the way of (3), and so they spend a bit of time defending it. But one concern they do not address is that (4.1) apparently does not suffice for weak TCS. One would need to add the further premise:

- (6.1) If knowing x is at least knowing y , then x is at least y .

Yet the same kind of counterexample exists for (6*). Knowing the radius may be enough to know the diameter of a circle. But again, it would not follow that radius *is* constituted at least by the diameter.

¹ Bar-On et al. cite Lycan (1984) and Field (1977) as proffering the epistemic argument as well, though to me it seems easier to read instead into these texts the abductive inference (discussed below).

² If omniscience of the non-linguistic world is crucial here, the argument would be concerned with only non-metalinguistic sentences. But of course, one could see easily enough how the argument might extend to metalinguistic sentences.

But even if the argument is unsound, I think the premises cite phenomena that make for a good *abductive* argument for (4). In the simplest case, (5) is given a fairly plausible explanation by (4) [i.e., (4R)]. Of course, the strength of an abductive argument comes in degrees, and ideally, we want the argument to be not only an inference to a plausible explanation, but also to the *best* explanation available. But there are no obvious alternate explanations that provide serious competition, and indeed, it simplifies our ontology to assume that truth-conditions are simply part of meaning, rather than that they are assigned to a sentence as some independent thing.

[Note: If a truth-condition is a state-of-affairs, then from (4), it follows that a sentence's meaning partially constituted by a non-linguistic state-of-affairs. This may be a counterintuitive result, since the tradition has been to think of meanings as “in the head,” though I find the anti-individualist arguments on the matter convincing (see the Fifth Chapter).]

Yet why settle for the weak version of TCS? Why not say that truth-conditions are *wholly* constitutive of meaning? Perhaps that would be an abductive inference with some weight, but it would not have as much weight, thanks to Fregean intuitions. Even if we are not Fregeans, we would like an explanation of (at least) the *appearance* of there being more to sentential meaning than truth-conditions, before we completely dismiss the issue. That is, we would want an *error theory* of why some of us are inclined to say that ‘Hesperus = Hesperus’ and ‘Hesperus = Phosphorus’ are different in meaning, despite having identical truth-conditions.³ (Of course, one can say that the first sentence is

³ Bill Lycan points out that for Russell (and arguably for Frege himself), the two Frege-sentences do *not* have identical truth-conditions—Russell's view is that there are worlds where ‘Hesperus = Hesperus’ is true and ‘Hesperus = Phosphorus’ is false. Thus for Russell, TCS could still be a full account of meaning,

uninformative while the second is not, but then we'd want an error theory for why it is intuitive to say that the difference in informativeness counts as a difference in meaning.) Without such an error theory, there seems to be little reason at this stage for thinking it is the Fregeans who are in error, rather than TCS.⁴ My own view is that the Fregeans are not in error, although I will not be able to defend this here.

A final worry is this: How can we rule out the alternative hypothesis that sentential meaning is partly constituted not by truth-conditions, but by the sentence's Fregean "sense"? The Fregean sense of a sentence, as I understand it, would be a function which takes a world as input and outputs a truth-value. Considering the worlds collectively, then, the function determines which conditions make the sentence true, i.e., the function determines its truth-conditions. Thus, if sentential-meaning consisted (partly?) in such a Fregean sense, we could explain why sentential-meaning *determines* truth-conditions, without assuming that sentential-meaning is partly *constituted* by truth-conditions. Thus, weak TCS would not be the only abductive inference available from (5); there is the competitor hypothesis that sentential meaning is (partly?) constituted by these Fregean senses.

I suspect, however, that the issue is merely verbal. Since the Fregean sense of a sentence is identical to a function's input-output pairs (where each pair consists of a world and a truth-value), such a function can be viewed as partly constituted by the pairs consisting of a world and the truth-value "True". Moreover, those worlds (I assume) are

and still assign different meanings to the Frege sentences. However, I assume the Russellian spin on TCS is now archaic; this is why I assume the Frege sentences threaten TCS directly.

⁴ Barwise and Perry (1981) exhort us to reclaim our "semantic innocence," claiming that Frege corrupted us into confusing epistemic differences with differences in meaning. But when asked whether the Frege sentences are same or different in meaning, my undergraduates overwhelmingly side with Frege. So it seems that if our intuitions are corrupted, it isn't due to Frege.

partly constituted by the condition under which the sentence is true (i.e. its truth-condition). So by transitivity of the “parthood” relation, the Fregean function is partly constituted by the sentence’s truth-condition. Thus, even if you view a sentential meaning as partly constituted by a Fregean sense, this does not preclude weak TCS—quite the contrary, it *entails* weak TCS.

II. Tarski-Davidson Circles.

Again, although (4) is weaker than Davidsonian TCS, the Davidsonian programme would still be (at least) a part of the theory of meaning (and an important part at that). For the machinery Davidson uses defines the truth-conditions for any sentence in a given language, which according to (4) is to define (at least part of) the sentence’s meaning. Nevertheless, there is an important difficulty with this machinery which I will be discussing in this section.

Davidson’s machinery is of course borrowed from Tarski’s (1935), (1944) theory of truth; this theory entails, for given a language L, all instances of the following schema, where S is an arbitrary sentence that means p:

(T) S is true iff p.

For Tarski, then, the theory tells us what it is for a sentence (with a specified meaning) to be true. Davidson’s observation, however, was that instances of the (T)-schema also might tell us what a sentence (with certain truth-conditions) means. So given a Tarskian theory of ‘true’ relative to L-sentences, Davidson proposed that such a theory might stand as a theory of *meaning* for L, by generating the truth-conditions for any sentence in that language. [Under (4), however, we would say that the Tarskian truth-theory for L can act

as at least *part* of a meaning-theory for L.] To dig into this further, then, we need to look at the Tarskian definition of ‘true.’

Tarski begins by classifying the various uses of ‘true’ into different orders of language, so that no language at any given order contains its own truth-predicate. (Thus, Liar Paradoxes are avoided.) Consequently, given an object-language and a well-ordering of metalanguages, we can say in a metalanguage ML what the predicate ‘true’ means, as it applies to sentences of a language L. Thus, for a particular ML, we define ‘true’ *but only with respect to L-sentences*.

For a particular ML, Tarski analyzes the truth-predicate into the notion of *satisfaction*. Specifically, he defines the extension of ‘true’ as the set of L-sentences where the predicate is satisfied by the object(s)⁵ substituted in for the variable(s). But, as we have indicated, a predicate is satisfied by the object(s) just in case replacing the variable(s) in the predicate with the object(s) results in a *true* sentence. The upshot would be that, given language L, an L-sentence is in the extension of ‘true’ iff it is true, which seems quite an uninformative analysis. Yet in lieu of this, Tarski proposes an alternative, recursive definition of ‘satisfaction.’

In brief, given an *n*-place predicate and *n* objects, the recursive definition of ‘satisfaction’ will output the conditions under which the predicate is satisfied by the objects. (Thus, as one application, if we input the predicate ‘*x* is white’ and the object snow, the recursive definition of ‘satisfaction’ will output that the predicate ‘is white’ is satisfied by snow iff snow is indeed white.) The recursive definition, specifically, first

⁵ I have some discomfort with Tarski’s talk of *objects* being substituted in for such variables rather than *terms*. It is not as if the result of a substitution is some kind of hybrid between linguistic and non-linguistic objects. However, such talk may be shorthand for substituting the variable with a *term* denoting that object. In any case, this is how I would like to be understood when I adopt Tarski’s way of speaking.

lists the satisfaction-conditions of (finitely) many predicates, and then gives rules which determine the satisfaction-conditions of more complex predicates. (Example: If ‘Fx’ is satisfied iff x is F, and ‘Gx’ is satisfied iff x is G—then ‘Fx and Gx’ is satisfied iff x is F and G.) As a special case, Tarski notes that for a predicate with no free variables (i.e. a sentence), *any* sequence of objects will satisfy this expression iff it is true, whereas *no* sequence of objects will satisfy this expression iff it is false. Tarski then defines a true sentence as a predicate with no free variables that is satisfied by all sequences; otherwise it is false.

In sum, then: To give a nonparadoxical analysis of ‘true,’ we must define ‘true’ relative to a particular language L (not containing its own truth-predicate); and in particular, define a set of things (in this case, a set of L-sentences) as its extension. Tarski’s proposal is that we define this set as the set of sentences which is satisfied by all sequences of objects. And since the recursive definition of ‘satisfaction’ defines the truth-conditions for such sentences, we can then substitute ‘is true’ in for ‘is satisfied’ in the recursive definition, and derive an instance of the T-schema for each sentence.

Again, Davidson’s key observation is that these T-schema instances suffice as *interpretations* for every sentence of the language, at least in the sense that they assign a truth-condition to each sentence. However, it seems that TCS would only be a (partial) theory of *sentential* meaning at best. Regarding subsentential content-words (i.e. terms), the Tarskian definition of ‘true’ simply presupposes a list of reference- and satisfaction-conditions for (finitely many) such words. We might very well ask then “what defines the reference- and satisfaction-conditions for terms?” It seems that TCS might talk of these conditions as conditions under which certain sentences are *true*. But then, the question

gets pushed back onto the notion of truth. Moreover, we could not appeal to Tarskian truth-theory to resolve the matter, on pain of circularity. Since ‘true’ in Tarskian theory is defined by the reference- and satisfaction- conditions of terms, our inquiry would land us back where we started.

But instead, one could use ‘denotation’ to define the satisfaction-conditions of a predicate, so that what satisfies the predicate just is the denotation of the predicate. Moreover, we can see reference-conditions as a special case of satisfaction-conditions—after all, the reference-conditions for a singular term X are those conditions where X satisfies the predicate ‘is identical to X.’ If so, then TCS seems able to use ‘denotation’ to define the reference- and satisfaction-conditions of terms. And this would still allow TCS to draw upon the Tarskian theory of truth, without the circularity from before.

This is a view which so far takes ‘denotation’ as primitive. But then again, one can go further and define ‘denotation’ in terms of truth. Yet at that point it should be clear that our theory has re-engaged the circularity. To understand meaning, TCS utilized the notion of truth, which in turn, was analyzed into satisfaction-conditions and denotation—but now denotation would be analyzed back into truth. (See Figure I).⁶ [Note that under weak TCS, meaning *per se* is not given a circular analysis; rather, its analysis will use notions which are themselves given a circular analysis.]

Yet is the circularity vicious? David Wiggins (2002) suggests that the account of ‘true’ is only meant to map out the “conceptual connections” that truth bears to other

⁶ In Strawson (1969), the circularity between Tarskian truth and TCS is that we can generate instances of the (T)-schema only if we are given the meaning of sentences—but we would get the meaning of sentences only from instances of the (T)-schema. I avoid this way of putting the circularity, however, since one can derive instances of the (T)-schema without sentential-meanings, described as such. In Tarski’s own theory, one derives T-sentences from the denotation of terms plus the compositionality rules. Nonetheless, the basic problem is the same, viz., that semantic terms are (sooner or later) analyzed into a circle.

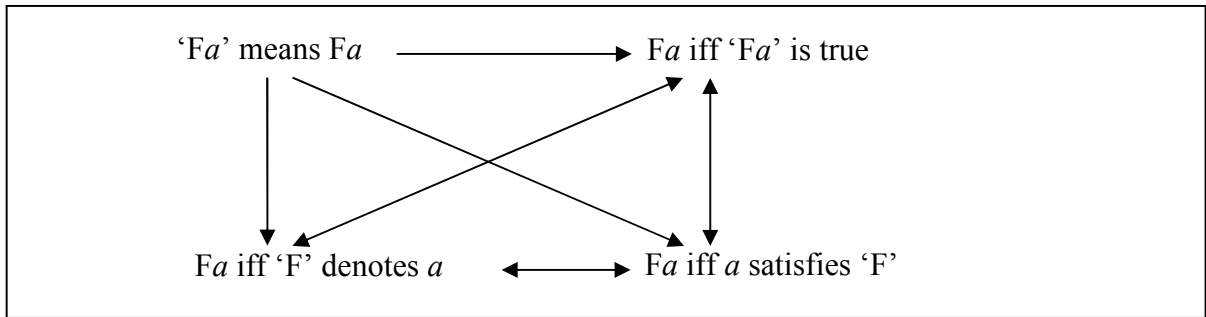


Figure 1. Tarski-Davidson Circles. Semantic terms such as ‘satisfies,’ ‘true,’ etc. seem interdefined in various ways, resulting in a circular definition for each term. [Arrows here are used as conditional and biconditional arrows.] Note that since we are assuming (4) instead of Davidsonian TCS, the arrows only go one direction away from ‘Fa’ means Fa’.

semantically interesting notions, e.g. the notion of meaning. These conceptual connections may form a circle, yet the claim is that mapping out these connections is theoretically illuminating nonetheless.

I am willing to grant that knowing the connection between truth and meaning is theoretically important in a certain respect. It is informative to know that the notion of truth has a strong conceptual tie to meaning (and vice-versa), as opposed to having a conceptual tie with, say, desire. But in other respects, this circular analysis seems impotent. For instance, it seems we will run into circular explanations of an uninformative sort, akin to the explanation “water is wet because water is wet.” E.g., we will explain the truth of a sentence by citing the denotations of its containing terms, but explain these denotations via the notion of truth. All told, this seems less than ideal.

III. A Proposal.

Perhaps a circular, Wiggins-style theory is the best we can do. But as inquirers we are pushed to search for something better before giving in to this pessimism. So how might we better deal with all this circularity? The most popular option is to break out of the circle by defining one of the semantic notions into non-semantic terms, in the style of

naturalized semantics. Thus, following Kripke (1972), one could define ‘*x* denotes *y*’ as “*x* bears the right kind of causal-historical relation to *y*”—and one might further unpack ‘the right causal-historical relation’ in the style of Fodor (1990), or Millikan (1984), or whatnot. There is no consensus, however, on how the relevant causal-historical relation is to be understood, and all such accounts face difficulties, such as the problem of empty names. (Even going by the Third Chapter, it is not obvious how a MIO would enter into a causal-historical relation with a term, since MIOs are not even **actual**).

But suppose that there is a way to analyze ‘denotes’ into non-semantic terms, in line with the Causal Theory. Of course, assuming circularity is not an option, there will eventually be some term or other which goes unanalyzed. This is just the semantic pessimist’s point that non-circular interpretation has to end somewhere. Yet part of the point here is that certain fundamental questions thus go unanswered by the theory. So for example, in the Causal Theory, if a “cause” is understood as an “actual sufficient condition” or perhaps an INUS-condition [in the sense of Mackie (1965)], it may turn out that the notion of a “sufficient condition” goes unanalyzed. And so, any question about the nature of such conditions goes unanswered. Of course, we could go on to identify a sufficient condition as what the antecedent of a material conditional expresses, but ‘expresses’ here would be just another word for “means”, and we would end up back in the circle.

Davidson’s own (1996) preference seems to be to take ‘true’ as primitive. Thus constructed, TCS would still nicely illustrate (at least in part) how sentential meanings “depend on” the meaning of the parts. Nonetheless, this choice requires the unhappy

stance that ‘true’ is *unanalyzable*. And besides, the longer we can stave off theoretical primitives, the better.

But is it any better to take a different term as primitive? In fact, I think there is a way to avoid making a tough choice for any given language. In the case of English (or more properly, metaEnglish), we can always define (at least one) of the semantic terms using a *different* language. That way, we would not be stuck with circular interpretations of metaEnglish semantic terms, though of course we would be taking as primitive a semantic notion in some other language.

To illustrate, weak TCS might first analyze just the metaEnglish predicates ‘means_{OL}’, ‘true_{OL}’, and ‘satisfies_{OL}’ in metaEnglish. Where ‘ x is F’ is an atomic sentence:

(10) a) ‘ x is F’ means_{OL} p only if [‘ x is F’ is true_{OL} iff p]

b) ‘F’ means_{OL} x only if ‘F’ is satisfied_{OL} by x

(11) ‘ x is F’ is true_{OL} iff ‘F’ is satisfied_{OL} by x .

(12) ‘F’ is satisfied_{OL} by x iff ‘F’ denotes_{OL} x

[The predicates would then get defined on composed sentences in the expected way.]

Now at (10)-(12) denotation is primitive since ‘denotes_{OL}’ goes unanalyzed. And so, even though the other metaEnglish semantic predicates are all defined in metaEnglish, thus far there is no circle if we take the stance that the clauses on the left are defined by the clauses on the right (and not vice-versa). When it comes to ‘denotes_{OL}’, then, the idea would be to define this term in a different language. But note that we *cannot* define ‘denotes_{OL}’ in metametaEnglish using metametaEnglish semantic terms, as in:

(13) ‘F’ denotes_{OL} x iff ‘F’ denotes_{ML} x .

In this case, (13) is false, since ‘denotes’ defined on object-language expression is non-equivalent to ‘denotes’ defined on metalanguage expressions. (That is how paradox is avoided in a classical system.) More concretely, (13) is false since ‘denotes_{ML}’ is defined on compound phrases that use ‘denotes_{OL}’, but ‘denotes_{OL}’ is not defined on such phrases.

Thus, if ‘denotes_{OL}’ is to be defined in metametaEnglish, it apparently *must* be defined by non-semantic terms, assuming such a thing is possible.⁷ But instead of waiting on an adequate naturalistic semantics, in the meantime we can simply use a different natural language such as French to do the work. Thus, (10)-(12) might be supplemented with the following thought:

(14) «‘F’ denotes_{OL} x » est vrai_{Anglais} si, et seulement si, «F» dénote_{Anglais} x.

Here I assume that ‘vrai_{Anglais}’ and ‘dénote_{Anglais}’ are defined on expressions of metaEnglish. But beyond that, we can leave these terms unanalyzed if we wish. For at this point, *all semantic terms in metaEnglish* have been given non-circular interpretations. Of course, we could subsequently go on to analyze ‘vrai_{Anglais}’ and ‘dénote_{Anglais}’ as well. Though at some point, if circularity is not an option, eventually some semantic element will go unanalyzed. This is of course what we’d expect under semantic pessimism. But under pessimism, we should also acknowledge that our interpretations of semantic terms merely function as “counterfeit” answers to what these terms mean. After all, our entire analysis of semantic terms would hang on the

⁷ Actually, though, if we called such a naturalistic *definiens* for ‘denotes_{OL}’ “D”, D would still have to be *equivalent* to ‘denotes_{OL}’ (since D is used to define that term). But apparently, anything which is equivalent to ‘denotes_{OL}’ must not occur in the object-language, on pain of paradox. This is a bit odd, since *ex hypothesi* D is supposed to be free of semantic terms. So it seems more than just semantic terms must be eliminated from OL, assuming there is such a D. In fact, I do not mean to be wedded to this argument here; I need to think more about it; but I mention it for the reader’s consideration.

interpretation of *one* semantic term taken as primitive—though *qua* primitive, the theory would say nothing about which relation the term expresses.

In addition, since some semantic element goes unanalyzed, our theory will not constitute a *general* analysis of semantic notions. This in fact is why Dummett (1974) objects to Davidsonian TCS: The complaint is that TCS is not a theory of Meaning-In-General, but only for a particular language L. But according to semantic pessimism, *one cannot give a General Theory of Meaning* since you are eventually forced to use a “background language” whose semantics go unspecified. It is understandable why some might be critical of a semantics which ultimately uses a set of unanalyzed terms, but this is the kind of limitation which semantic theory seems to face in general.

IV. Semantic Explanation.

Again, if we accept (10)-(13), we would essentially have a theory of all metaEnglish semantic terms without falling into circularity. To be clear, though, it is not as if one is *forced* to define ‘denotes’ and ‘satisfaction’ in the manner I have suggested. One can always settle for circular interpretations. But again, it seems to me that circular interpretation cannot offer the explanatory power which non-circular interpretation can.

To illustrate this, let us consider three characters, François, Francis, and Frank. François speaks only French, let us assume, while Francis speaks French and English, whereas Frank only speaks English. Suppose for his amusement, one day François goes to Germany and tries to acquire the language, *à la* the radical translator. (François has a sick sense of what a good time is.) Suppose that his success is almost nil, except he manages to observe that the sentence ‘Der Schnee is weiss’ is asserted if and only if snow

is white. (I leave the details of how François discovered this to the reader's imagination.)

Now François wants to boast his new discovery to his friend, Francis. And after telling Francis the details of his observations, François concludes, in the style of a radical translator:

(15) 'Der Schnee ist weiss' veut dire que la neige est blanche en L'Allemand, parce que 'Der Schnee ist weiss' est vrai en L'Allemand si et seulement si la neige est blanche.⁸

Now Francis knows that François is an intellectually careful fellow, and would not make such a claim unless the evidence was strong. So he comes to believe François as concerns (15). Nevertheless, François wants to prove his skills at radical translation—and so, he looks up the words in a German-French dictionary he happens to be holding. After a few moments, François shows Francis that the dictionary vindicates his claim.

Frank, who overhears them talking excitedly, asks Francis about what was the matter. Francis replies:

(16) François discerned that in German the sentence 'Der Schnee ist weiss' is true if and only if snow is white. And his observation was correct, since 'Der Schnee ist weiss' *just means* that snow is white in German.

Now from Francis' point of view, the meaning of 'Der Schnee ist weiss' was explained by appeal the truth-conditions of the sentence. Yet Francis then appeals to the *meaning* of 'Der Schnee ist weiss' to explain to Frank why François' assessment of the truth-conditions was correct. So in one sense, meaning is explained by truth-conditions, which is then explained by meaning. But despite the appearance of a circle, these explanations are each substantive, even from Francis' point of view.

One might think the need for such explanations is relatively rare. But perhaps it is rare to *invoke* such explanations, but I think the need for some explanation here is quite

⁸ Again, François' inference doesn't deductively follow here, but presumably it is a good abductive one.

real. Or more modestly, these explanations at least offer us something theoretically significant. And crucially, the kinds of semantic explanations are possible only on the views which give noncircular definitions of semantic terms. Such explanations thus constitute the primary advantage of our approach semantic terms, over theories that define semantic terms in circles.

**CHAPTER FIVE:
WHAT DOES THE SLOW-SWITCH ARGUMENT SHOW?**

The previous chapters concern the semantics of *linguistic* representations, and our limitations in answering certain questions in that domain. In the present chapter, however, I consider a pessimistic argument regarding the semantics of *mental* representations. Specifically, I consider the slow-switching argument from Burge (1988), which purports to show that we cannot know *a priori* what content our mental representations have. And although I conclude the argument is non-demonstrative, I think the argument can be extended to create an interesting skeptical problem, which is somewhat analogous to the epistemic problem discussed in Chapter One.

To start things of, consider (what we might call) the “inverted beetle-in-the-box.” Suppose that a group of speakers have lived their entire lives in a windowless box, and suppose they use ‘beetle’ to represent whatever is on the ground nearest to the Northwest corner, outside the box. Then, even if they are able to use ‘beetle’ to represent whatever that is, in one sense they have no idea what they are talking about. ‘Beetle’ could represent any number of things, for all they know—a hat, a motorcycle, a dead body, etc.

According to some philosophers, if semantic externalism is true, this would be the situation we are in, as regards knowledge of what we mean by ‘beetle’. More exactly, the externalist view may make it impossible to gain *a priori* knowledge of the content of my own use of ‘beetle’ and the correlative concept. (Hereafter, I shall call this knowledge of content *a priori* “self-knowledge”). According to externalism, the content of a natural

kind term like ‘beetle’ is determined in part by the world around us. But in *a priori* contemplation, we shut ourselves in our mental “box” in a way that occludes the external world. And from this position, it may seem that if externalism is true, we could not know *a priori* what content is expressed by ‘beetle’, since this would depend on features of the external world, which are inaccessible from inside the “*a priori* box.” From that position, it seems that the term ‘beetle’ could be about (say) flies, for all we know.

The issue here is important, since it seems to be nothing less than a clash between two landmark ideas in philosophy. On the one hand, Descartes convinced many of us that we can know mental contents introspectively, in a way that is independent of our knowledge of external objects. On the other hand, Frege convinced many of us that the content of ‘beetle’ (and of the correlative concept) determines what it denotes in the external world (or to use Frege’s dictum, that its *intension* determines its *extension*). Yet this seems to conflict with the Cartesian idea, since (as we’ll see) Frege’s claim implies the externalist view that content of ‘beetle’ is partly *individuated* by the external environment. In which case, knowledge of the content of ‘beetle’ would apparently require knowledge of its object—meaning you could not get knowledge of content just from inside the “*a priori* box.” You’d have to go out into the external world and have a look. The upshot is, apparently, one of philosophy’s long-standing intuitions has to go: Either Cartesians were wrong to think we could know contents by mere introspection, or Fregeans were misguided to think that content is what determines denotation.

However, I believe there is a way out of this dilemma. I think we can keep both the Cartesian and Fregean ideas (roughly) by giving up the idea *Cartesian* knowledge of content requires knowledge of denotation. In this paper, I shall first consider Burge’s

(1988) “slow-switching” argument, an argument which presents the Descartes-Frege conflict in one of its guises. Many writers (including Burge) have come to reject the argument—but in fact, I think there is more to it than is commonly assumed. I eventually agree, however, that the argument’s force is limited, due to an independently-motivated semantics for ‘knowing wh-’ locutions from Boër and Lycan (1986). More exactly, with this semantics in hand, we will see that although one variety of self-knowledge is threatened by Frege-*cum*-externalist, the Cartesian variety is not.

I. Externalism and Anti-Individualism.

Externalism, to be precise, is the view that the content of a subject’s natural and artificial kind terms/concepts is determined at least partly her social and physical environment.¹ The arguments for the view are based in the Twin-Earth thought-experiments in Putnam, (1975), Burge (1979), (1982), which I presume familiar. For our purposes, however, it will best to focus not on the positive thesis of externalism, but rather on its negative, anti-individualist consequence that:

(AI) Psychological states (described individualistically) do not wholly determine the content of a kind term/concept.

The psychological states I have in mind here are those individuated by physical states of the neurological system, which are (roughly) those states describable independently of the subject’s social and physical environment.²

¹ The view may also extend to singular terms/concepts, and others as well. But of course, a term such as ‘individualistic psychological states’ would be an exception to the view.

² N.B., Burge (1979) simply assumes that (AI) is an anti-Cartesian view, though since Descartes was a dualist, he could grant (AI) if narrow states are individuated physicalistically. However, Descartes himself plausibly individuates these states phenomenologically, so in that sense, (AI) would be anti-Cartesian. Nonetheless, to make the discussion more relevant to contemporary philosophers, I shall assume the physicalistic individuation of narrow states is appropriate.

In the case of kind *terms*, the thought-experiments purport to show (AI) on the Fregean assumption that a term denotes what it denotes in virtue of its content, so that:

(1) Content determines denotation.

I do not doubt that there are different senses of the term ‘content’, so that “content” is not always what determines extension. But it is the extension-determining sense of ‘content’ I intend here. So really, (1) is really more stipulative than anything. (This just expresses my commitment to the Fregean dictum.)

Now the thought-experiments suggest that for two uses of a kind term:

(2) Sameness of psychological state (described individualistically) does not determine sameness of denotation.³

Thus it follows from (1) and (2):

(3) Sameness of psychological state (described individualistically) does not determine sameness of content.

So from (3), (AI) would follow in the case of terms. And since the content of a term *just is* the content of the concept it expresses, the point easily carries over to concepts.

II. The Slow-Switch Argument.

In Burge (1988), the apparent difficulty for (AI) plus *a priori* self-knowledge arises from a different sort of thought-experiment. There, we are to imagine that Oscar (a normal, adult English speaker) is unwittingly switched to Twin-Earth. Now immediately after the switch, Oscar’s thoughts about water will presumably continue to be *de* water

³ Do we really need the thought-experiments to show (2)? The claim strikes me as plausible in its own right. Arguably, even Descartes in the first two *Meditations* grants that individualistic states do not determine whether a term/concept denotes an external object vs. an object in a dream; cf. Burge (2006b).

(i.e., H₂O), even though he is now in an environment that houses XYZ instead.⁴ But many anti-individualists have the intuition that, over time, those thoughts will eventually become *de* XYZ—even if he continues to be unaware of the switch. After all, Oscar will acquire the habit of using ‘water’ in the presence of XYZ rather than water, and he will thereby succeed in various language-games he plays with Twin-Earthians. But the intuition is also that Oscar will not know, at least not by introspection alone, that his thoughts are now *de* XYZ rather than water.

There may be several confusing features of the “slow-switch” example; but the point is supposed to be that in a slow-switch scenario, Oscar would not have introspective knowledge of his own mental contents. For he could not know introspectively whether thoughts using the concept expressed by ‘water’ (hereafter, “W-thoughts”) are *de* XYZ vs. water. Officially, we can formulate the slow-switch argument as follows:

- (4) If slow-switched, Oscar knows introspectively what content his W-thoughts have only if he can know introspectively that these thoughts are *de* XYZ vs. water. [Assumption]
- (5) Oscar cannot know introspectively whether his W-thoughts are *de* XYZ vs. water. [Assumption]
- (6) If slow-switched, Oscar does not know introspectively what content his W-thoughts have. [From (4), (5)]

It is then presumed that (6) indicates a conflict between anti-individualism and introspective self-knowledge.

⁴ I talk of thoughts as being “*de*” water, but this does not mean we can substitute co-referring terms in *ascriptions* of these thoughts. E.g., ‘S thinks that water is wet’ and ‘S thinks that H₂O is wet’ each ascribe a thought that is *de* water (on Earth), but I assume these ascriptions are non-equivalent. After all, the second ascription (but not the first) suggests that S deploys the concept H₂O rather than WATER. Even so, both thoughts are *de* water, since their truth depends on the properties of *water* as opposed to some other stuff. And to be clear, even though the thought-ascriptions themselves are opaque, we *can* substitute co-referring terms to say instead that S’s thoughts are *de* H₂O instead of *de* water. For this substitution does not alter the truth-condition of our claim about which stuff determines the truth-value of S’s thoughts.

The current literature on this argument primarily offers two objections. Some object that (6) is irrelevant to *our* self-knowledge. In the next section, however, I argue that this criticism, the “Relevance Objection” as I shall call it, is not entirely successful. Others (e.g., Burge himself) object to the argument by pointing to *self-verifying* second-order judgments, to show by example that slow-switching does not preclude *a priori* self-knowledge. In the end, I conclude in favor of the “Objection from Self-Verifying Judgments,” though with some important qualifications.

III. The Relevance Objection.

The Relevance Objection to the slow-switch argument is levied by a number of writers, e.g. Warfield (1992) and Brown (2004). The complaint is that it is unclear how (6) bears on us—after all, unlike Oscar, *we* are not victims of covert slow-switches. At best, (6) is relevant only if we add something like the following:

(7) If (6) and you have not “ruled out” the possibility of slow-switching, then you do not know introspectively the content of your W-thoughts. [Assumption]

This would mean that if you have not ruled out the possibility of slow-switching, you would lack self-knowledge, regardless of whether you have actually been switched. But of course, (7) is highly contentious. It is at least *prima facie* plausible that knowing *p* does not require you to rule out all skeptical scenarios where $\sim p$. To borrow the example from Dretske (1994), it seems you can know that a certain animal in the zoo is a zebra even if you cannot rule out that it is a cleverly painted mule. For in the usual cases, cleverly painted mules are not “relevant alternatives” to zebras. Similarly, in the case of self-knowledge, I can apparently know that a thought is about water, even if I have not

ruled out the skeptical possibility of a slow-switch to Twin Earth. For XYZ is not a relevant alternative to water; XYZ is just a fictional stuff made up by Putnam in the 70s.⁵

The relevant alternatives model is of course not accepted by everyone, but the point is that it indicates a gap in the slow-switch argument. If slow-switching is to show an incompatibility⁶ between (AI) and introspective self-knowledge, it appears we must also show that either that twins *are* relevant alternatives, or that the relevant alternatives model is incorrect.⁷ But thus far, we have not seen an argument for either claim.

Ludlow (1995), however, contends that sometimes twins *are* relevant alternatives. He considers the term ‘chicory’ which denotes chicory in American English and denotes endive in British English—though the two salad greens are indistinguishable to the untrained eye.⁸ Thus one can easily imagine an American expatriate in England who is unaware of the different usage, and is effectively the victim of a slow-switch. However, Warfield (1997) argues (and I concur) that such cases are quite rare.⁹ Slow-switching not only requires a term that actually switches denotation from place-to-place—it also requires that the denotations swapped are indistinguishable to the lay person. But the co-

⁵ A Moorean may also suggest on commonsense grounds that you do not need to rule out the cleverly painted mule. But since the relevant-alternatives response is more prominent in the literature, I focus on that instead. Besides, I eventually grant what the Moorean wants, viz., that commonsense knowledge of contents is ubiquitous.

⁶ The relevant sense of ‘incompatible’ is “materially incompatible,” where *p* and *q* are materially incompatible iff $\sim(p \ \& \ q)$. I thus agree with Ludlow (1997) that Warfield (1992), (1997) frames the issue incorrectly when asking if is a *possible world* where externalism is true and we have *a priori* self-knowledge.

⁷ Or show that the Moorean arguments are faulty (see n. 5).

⁸ Ludlow himself talks of arugula instead of endive in his example, though I am told that it is actually endive which ‘chicory’ denotes in England. (P.S. It may be helpful to know I speak American English.)

⁹ Ludlow also cites examples of philosophical terms, such as ‘realism’, ‘pragmatism’, etc., whose extensions vary between philosophical circles. But I would typically think of these as causing “fast switches” rather than slow-switches, though I suppose a slow-switch could conceivably occur with these terms. However, that would also seem quite rare in the actual world.

incidence of these two factors is fairly uncommon; if so, then in the majority cases, there is no need to consider twins as relevant alternatives.

My own response to the Relevance Objection is to reject the relevant alternatives model, at least for the *Cartesian* variety of self-knowledge. In this, Cartesian knowledge is understood as *skeptic-proof* knowledge—you Cartesian-know that p only if you have ruled out every skeptical scenario where $\sim p$. So in particular, in order to Cartesian-know that your W-thoughts are *de* water, you must rule out any skeptical possibility where those thoughts are about something other than water, as in a slow-switch scenario. As concerns Cartesian knowledge, then, it seems (7) is correct to say that ruling out a slow-switch is necessary in order to know.

Yet we now seem to avoid the Relevance Objection by restricting the slow-switch argument to a very unusual type of self-knowledge. Yet I think this is apt, since in Burge (1988), the slow-switch argument was only meant to target Cartesian self-knowledge anyway (p. 653).¹⁰ And lest this sound too parochial, recall that externalism generally gained notoriety for challenging the Cartesian conception of mind specifically. So in this way, we can see the slow-switch argument as one manifestation of this challenge.

Thus, it may be true enough that, commonsensically, we do know the content of our concepts—given that twins are almost always irrelevant alternatives. Things are different, however, when it comes to Cartesian knowledge. For in the context of the *Meditations*, you Cartesian-know that p only if you have ruled out every skeptical

¹⁰ Burge's term on p. 653 is 'basic self-knowledge', though it is clear this is meant to be Cartesian self-knowledge (see p. 649).

scenario where $\sim p$.¹¹ But on the relevant-alternatives model, you can know p while presupposing (quite reasonably!) that some deviant scenario in the external world does not obtain, e.g., that zebras in the zoo have been replaced with cleverly painted mules. Yet this just means that the relevant-alternatives model is not a model of Cartesian knowledge; hence, contra the Relevance Objection, that model does not bear on the self-knowledge at issue in the slow-switch argument.

IV. The Objection from Self-Verifying Judgments.

Thus, since it is best to see Cartesian self-knowledge as the target of the slow-switch argument, I conclude that the Relevance Objection is unsuccessful. The objection indeed shows that non-Cartesian self-knowledge is possible under externalism. But again, the slow-switch argument is concerned with Cartesian self-knowledge.

Even so, it may be unclear why anyone should care about *Cartesian* self-knowledge. After all, super-strong Cartesian knowledge is generally seen as quite unattainable—and so, it may be immaterial if (AI) precludes that kind of knowledge. I shall say more in the last section about why the issue is of interest. But for now, I take it to be intuitive that we *do* have Cartesian knowledge of content, for our own kind terms/concepts, even if such knowledge is otherwise unattainable. More exactly, the intuition is that you can know what you are thinking, *even when entertaining hyperbolic doubts about the external world.*

The challenge for the anti-individualist, then, would be to find an example of introspective self-knowledge which is unaffected by slow-switching doubts. If we can

¹¹ Strictly speaking, this is true only in the context of Meditations One and Two, and so my talk of the “Meditations” really only concerns these.

find such an example, it would show that slow-switch argument is unsound (though it would take further reflection to say which premise is false). In any case, a putative example of such knowledge is the judgment expressed by (8):

(8) I am thinking that water is wet.

The idea is that when I judge such a thing, then even in the context of the *Meditations*, it seems *I just can't be wrong* about that. Sure, I might accidentally use the wrong words if I try to articulate my judgment. But as Burge (1988) suggests, there is something *self-verifying* about my judgment—it seems that if I occurrently judge I am thinking water is wet, then *I really am* thinking that water is wet.¹² For the first-order thought that water is wet is, as it were, “contained” in the second-order judgment. That is, if you token this second-order judgment, you must “run through” the first-order thought in the process. And so, if you token the former, you *ipso facto* end up thinking the latter. But since my thinking the latter is what the second-order judgment claims, that judgment is self-verifying; tokening the judgment creates the condition under which it is true. (Hereafter, I shall refer to such judgments as “SVJs.”) The existence of SVJs is defended by several writers besides Burge, such as Heil (1988), Bar-On (2004), and Parent (2007).¹³

Yet what of Freudian repression? Repression phenomena illustrate that introspection can mislead us, suggesting (at the very least) that skeptical doubts can arise

¹² In this, the verb ‘think’ is non-equivalent to ‘believe’. If you “think” that *p*, this only means that you are in a state with the content *p*; it does not entail that you have any particular attitude toward *p*. See my (2007) for an elaboration of this.

¹³ If the slow-switch argument (in Burge) concerns *skeptic-proof* self-knowledge, then I assume self-verifying judgments in Burge are meant to constitute this kind of knowledge. This reading is supported by his claim (on p. 647) that our knowledge of content is sometimes has the “certainty” of the *cogito*. However, SVJs are not claimed to constitute skeptic-proof knowledge in Heil (1988) and Bar-On (2004); these writers are more concerned with a broader sort of epistemic “privilege” we have to our own thoughts.

when I judge (8). And if so, then it appears that introspective knowledge of (8) would not count as *Cartesian* self-knowledge, thus undercutting the Objection from SVJs.

However, I think we need to be careful in identifying the kind of introspective knowledge for which Freudian repression occurs. Let us distinguish between:

- (a) introspecting that you are currently thinking that water is wet,
- (b) introspecting that you are currently thinking *nothing but* that water is wet,
- and-
- (c) introspecting that you have a thought that water is wet which has been a *standing* thought for some period of time.

I would submit that failures of introspection, due to Freudian repression and the like, occur only when you attempt to know *exhaustively* your current thoughts, as in (b)-type cases, and when you attempt to know what you *have been thinking* for a period time, as in (c)-cases. But if you introspectively judge only that you have the occurrent thought that water is wet, then *that* judgment seems to be sufficient for its own truth.

Be that as it may, it seems that only an extreme reliabilist would say there is *knowledge* in this case, even if the judgment is self-verifying. I shall say more about the reliabilist view in the next section. But as it stands, note that others besides reliabilists could count some SVJs as knowledge. For instance, if you token an SVJ *while aware* that such a judgment is self-verifying, then a more internally-minded epistemologist might count such SVJs as knowledge. And that may be enough to show that the slow-switching argument is non-demonstrative against all Cartesian introspective self-knowledge.¹⁴

¹⁴ Bill Lycan reminds me that further conditions may need to be in place, before we have *bona fide* Cartesian self-knowledge, e.g., perhaps your awareness that the judgment is self-verifying must be clear and distinct. Regardless, the point still stands that at least some SVJs may constitute Cartesian self-knowledge, for all that has been said so far.

V. Brown's Reply.

So far, then, it seems that the Objection from SVJs may have a point. At least, it is not obvious at this stage that the slow-switch argument undercuts all Cartesian self-knowledge. Yet Jessica Brown (2004) offers further considerations on behalf of the slow-switch argument: Brown first observes that (8) would express an SVJ in *both* English and Twin-English. That is so, even though the judgment expressed would concern *different first-order thoughts*. In one case, it would concern a water-thought—and in the other case, an XYZ-thought. In addition, even if SVJs are *perfectly reliable* about their respective first-order thoughts, Brown claims that the switching victim is unable to *discriminate* between water-thoughts and XYZ-thoughts, without empirical investigation. And inspired by Goldman's (1976) barn façades example, Brown further maintains that a discriminatory ability is necessary for knowledge of one's thoughts.

Essentially, we can see Brown as suggesting a variant of the slow-switch argument, which makes *discriminatory abilities* key to self-knowledge. Let us call an "(8)-judgment" a judgment which the subject would express using (8), where this expresses different SVJs in Twin-English vs. English. Then, the argument runs:

- (4') If a subject is slow-switched and tokens an (8)-judgment—then she introspectively knows what content her (first-order) thought has, only if she can introspectively discriminate water-thoughts from XYZ-thoughts. [Assumption]
- (5') A subject cannot introspectively discriminate water-thoughts from XYZ-thoughts. [Assumption]
- (6') If a subject is slow-switched and tokens an (8)-judgment, she does not introspectively know what content her (first-order) thought has. [From (4'), (5')]

If we want to apply this to the Meditator's case, we may add:

(7') If (6') and the Meditator has not “ruled out” the possibility of slow-switching—then even if she tokens an (8)-judgment (and is aware that it is self-verifying), the Meditator still does not introspectively know what content her (first-order) thought has. [Assumption]

(I speak of the “Meditator” here just to keep clear that *Cartesian* self-knowledge is the concern. Of course, in this I do not mean to do Descartes exegesis, but rather to discuss whether externalism precludes the kind of self-knowledge claimed in the *Meditations*.)

As with (7), premise (7') would be contentious if we were dealing with someone other than a Cartesian Meditator. For the switching possibility seems to be something we normally do not need to rule out. But with the Meditator, this seems necessary. Otherwise, since the Meditator takes seriously any skeptical scenario, it would be a live question for her as to which of two stuffs her thoughts are about. And in that case, it appears that the Meditator would not know what content she harbors.

So given (7'), it looks as though the Meditator would not introspectively know what content she harbors, even in tokening an (8)-judgment. That is so, given she cannot rule out slow-switching from her assumptions (or rather lack thereof). The upshot, then, would be that the Meditator lacks introspective knowledge of content. Thus, if Brown's refined slow-switch argument is successful, the Objection from SVJs would be wrong to hypothesize that an (8)-judgment can constitute a piece of Cartesian self-knowledge.

Now the refined argument relies heavily on the notion of a “discriminatory ability,” but what is that exactly? Brown first considers Goldman's (1986) view that this ability is an ability to form reliably true beliefs. (Definition: The belief that p is reliably true iff, in relevant alternative situations where $\sim p$, the believer does not believe p .) So on Goldman's view, a discriminatory ability is basically an ability to “track the truth,” i.e., track state-of-affairs where a belief is true, as distinct from those where it is not.

Brown grants, however, that the subject is indeed *perfectly* reliable in her SVJs about which first-order thoughts she has. So why say that a switching victim cannot discriminate water-thoughts from XYZ-thoughts? Given the subject's impeccable ability to "track" her own water-thoughts as distinct from XYZ-thoughts, why think the victim cannot discriminate between them, as (5') suggests?

Well, contra Goldman, Brown's contention is that reliability is insufficient for an ability to discriminate between thoughts. She has more than one argument for this, but I believe her most compelling point is that a slow-switch victim would *deny* that any shift in content has occurred, if we asked her. Even so, there is a slight infelicity here since our concern is not the victim's ability to notice *changes* in thoughts *per se*, but rather in her ability to notice whether her thought is *de* water in particular, as opposed to something that merely looks like water. But no matter. One could instead ask the victim which of the two liquids her current W-thoughts are about. The claim, then, would be that the switching-victim (once she is informed of the switches) would be unable to answer such a question. If so, then it may seem that the victim cannot discriminate water-thoughts from XYZ-thoughts, at least not without empirical investigation.

Yet now a different problem arises, noted by Brueckner (1999), (2003). Suppose the switching victim is asked 'Are your W-thoughts currently *de* water or something that merely looks like water?' Then, even if she is ignorant of which planet she is on, the victim could know *a priori* that the answer here must be 'water'. That is so, *regardless of which stuff her thoughts are about*. After all, "W-thoughts" are by definition thoughts that are *de* whatever she currently calls 'water'. But that means 'water' must be the correct answer to the question to what her W-thoughts are about. (This assumes that the

victim and the inquisitor mean the same thing by ‘water,’ but that is required for he victim to understand the question in the first place.)

Thus, apparently the victim *is* able to answer the question, and so nothing here shows she cannot discriminate water- and XYZ-thoughts. However, Brueckner (1999) identifies an alternative way to present the question so that it is not so easily answered. Suppose we say to the switching victim:

Your W-thoughts were originally *de* one kind of liquid; call it “home-water.” But at least once you were unwittingly switched to a different planet, where your W-thoughts eventually became *de* a liquid that only looks like water from home. The question is then: Are your present W-thoughts *de* home-water or not?

It seems the switching-victim would not be able to answer this question, at least not without some empirical knowledge.¹⁵ Thus in this sense she is unable to discriminate introspectively between thoughts of the two kinds. Of course, Goldman may again insist that she *is* able to discriminate them, thanks to her infallible SVJs. But it seems there is at least one sense of ‘discriminate’ where the victim cannot “discriminate” between these thoughts, given her inability to answer the challenge-question *a priori*. (Or if preferred, we may just give the word ‘discriminate’ to Goldman, and use some other term to express what the subject cannot do epistemically speaking, viz., answer our question.)

VI. Cartesian Self-Knowledge and “Knowing What”

So in one sense, it seems (5') is correct to say that the Meditator could not discriminate introspectively between water-thoughts and XYZ-thoughts. But does that mean Cartesian self-knowledge is impossible? Falvey & Owens (1994), as well as Goldberg (1999), suggest that an (8)-judgment would suffice for knowing what she

¹⁵ N.B., Brueckner also argues that the ability to answer this question is unnecessary for *a priori* self-knowledge. But as I argue below, it depends on one’s purposes as to whether this ability is needed.

thinks nonetheless, though they concede it would be a non-discriminating sort of knowing. But Brown, of course, would say “non-discriminating” knowledge is a *contradicto in adjecto*.

So who is right? Note that the issue concerns whether the Meditator would “know what” she thinks in virtue of an (8)-judgment. And here, it is important that the semantics for “know wh-” attributions are highly *purpose-relative*, as argued by Boër and Lycan (1986). In other words, the truth of a “know wh-” attribution depends on the purposes set in a context. To illustrate, consider that we would ordinarily say a five-year-old “knows what” water is. After all, the child has the kind of knowledge which enables him to fetch a glass of water (as opposed to orange juice, soda, molasses, etc.). Yet if asked what water is on a chemistry exam, the same child may not count as “knowing what” water is, if he cannot select the correct chemical formula among the answer-choices. Thus we might say that the child knows what water is for *ordinary* purposes, though not for the purposes of the chemistry exam. And so we get a purpose-relativity in “knowing what.”¹⁶

If this is correct, then whether the Meditator “knows what” she thinks will also depend on her purposes. If the aim is specifically to answer our challenge-question, then the Meditator apparently cannot “know what” she is thinking, at least not introspectively. For she cannot answer, without empirical investigation, if her W-thoughts are *de* home-water vs. something that merely looks like water from home.

¹⁶ Note that the purpose-relativity in “knowing what” is different from the context-sensitivity which some have argued in the case of “knowing.” Epistemic contextualism [as defended by Lewis (1996), DeRose (1995)] is the view that the truth-conditions of knowledge-ascriptions are conditioned by how high the evidential “standards” are in a context. However, the purpose-relativity in “knowing what” is different from the contextualist’s context-sensitivity. No matter what the evidential standards, a person “knows what” water is, for the purposes of a chemistry exam, if she knows that water is H₂O. It’s just that, in a skeptical context, the antecedent might not be satisfied, due to high evidential standards—whereas the antecedent may be satisfied when lower standards are in place. But whether or not the antecedent is satisfied, the fact remains that *for the purposes of the exam* she knows what water is *if* she knows water is H₂O.

Still, the Meditator might have other purposes which do not require her to answer our question. For example, an (8)-judgment may qualify her as knowing what she thinks, at least for the purposes of being *infallible* on the matter. Infallible knowledge is indeed one aim of the *Meditations*—and relative to that particular purpose, absolute reliability may be enough for “knowing what,” contra what Brown would suggest.

However, this kind of point is probably easier to make in relation to an SVJ that is more “descriptive.” Suppose that instead of (8), the Meditator asserts the following:

(8') I am thinking about a colorless, odorless liquid.

Even though (8') does not specify a complete first-order thought, (8') still looks self-verifying: If the Meditator judges that she is thinking about a colorless, odorless liquid, then it seems she really is thinking about such a liquid. Again, in order to judge she is thinking about this liquid in the first place, she needs to “run through” her descriptor ‘a colorless, odorless...’ in her mind. But that is sufficient for thinking about such a liquid.

Again, with this “descriptive” SVJ, it is easier to say the Meditator “knows what” she thinks, at least for some purposes, even if she cannot answer the challenge-question. For her use of the descriptors suggest she is able to answer other, substantive questions. Thus when she tokens (8'), she seems well-positioned to answer questions like “Are you thinking of something red? Are you thinking of a solid body? Are you thinking of something sweet-smelling?” And so, for the purpose of answering these kinds of questions, it seems she indeed “knows what” she is thinking.¹⁷

¹⁷ The point holds, even if the Meditator cannot answer “Are you thinking of something that exists?” In the context of *Meditation Two*, she probably cannot answer that question, though she can still answer whether she is thinking of something red, solid, etc.

Importantly, the Meditator is well-positioned to answer these other questions *even if she has actually been slow-switched*. For it is unnecessary to distinguish a content that is *de* water from one that is *de* XYZ, for the purposes of answering *those* questions. Thus, if the task is just to identify whether the object of thought is red, solid, sweet-smelling, etc., then the Meditator may indeed “know what” she is thinking when she judges (8’).¹⁸

Apparently, then, this shows that (4’) is susceptible to counter-example. If the Meditator judges (8’) by introspection, it may be that this qualifies her as knowing what she thinks (for at least some purposes) *even if she actually has been slow-switched*. That is so, even if she cannot know introspectively whether her thought is *de* water vs. XYZ. Of course, if the Meditator explicitly has the aim of answering the challenge question, then it seems she would be required to know whether her thought is *de* home-water or not. And for that purpose, it may indeed be impossible to “know what” she is thinking just by introspection. But ruling out slow-switches is just one of many goals which the Meditator might have—and oftentimes, the Meditator’s purposes might indeed qualify her as “knowing what” she thinks in virtue of her (8’)-judgment. At the very least, nothing has been said to defeat this kind of claim.

Nevertheless, it may sound odd to suggest that the Meditator could know what she thinks, even though (for all she can tell) her thoughts may have one of two contents. Yet this may just re-enforce that, for some of her purposes, the Meditator isn’t required to distinguish a content from a twin. Oddly, then, “knowing what” in the *Meditations* can be similar to knowing what water is for ordinary purposes. (But of course, there would be

¹⁸ The claim is congenial with contrastive accounts of knowledge [see Schaffer (2005)]: What the Meditator knows would contrast with the proposition that she is thinking about, e.g., a red ball. (Still, with contrastive accounts, I would argue that purpose-relativity can *explain* the relativity to certain contrast-propositions.)

stark differences as well.) In particular, both kinds of “knowing what” would not require the individual to know something that distinguishes water from a twin. Nonetheless, a child positively does know what water is, for ordinary purposes. Similarly, the Meditator may indeed know what content she thinks, for some of her purposes.

Thus, I conclude the Objection from SVJs is successful against the slow-switch argument. At the very least, the Objection shows that the argument is non-demonstrative: Even if a victim is slow-switched, some variety of Cartesian self-knowledge might remain, relative to some purposes. Nonetheless, I have also concurred that:

- (9) The argument indeed shows that the Meditator fails to know what mental content she harbors, for the purpose of answering the challenge-question.

But I have also shown that:

- (10) The argument does NOT show that the Meditator’s (8’)-judgment is insufficient for knowing what content she thinks for *every* purpose—even if she is slow-switched.

And so, I claim the slow-switch argument fails to show that anti-individualism precludes Cartesian self-knowledge. So, for all that has been said, a Fregean-*cum*-externalist can embrace the kind of self-knowledge claimed in the *Meditations*.

I should also like to add that:

- (11) (9) is no cause for philosophical anxiety.

I take it to be unsurprising if the relevant kind of introspective self-knowledge is impossible, viz., when the purpose is to answer our challenge-question. Who would have thought you could identify what the *actual truth-maker* is for your W-thoughts, from inside the “*a priori* box”? Certainly not Descartes. For all he claimed to know *a priori*, his W-thoughts might have been about a non-existent stuff. Perhaps W-thoughts are *de facto* determined partly by a certain kind in the world, but this is not something Descartes

assumed *a priori*. I doubt, then, that Cartesian self-knowledge itself was ever meant to answer our challenge question.

VII. The Charitable Skeptic.

I promised earlier to say more about why skeptic-proof knowledge of contents is of interest. To repeat, the issue is of interest partly because we *do* have an intuition that we can know what we think, even when entertaining skeptical doubts. Furthermore, if the the slow-switch argument is sound, that would demonstrate one anti-Cartesian consequence of (AI), the kind of putative consequence which made anti-individualism famous. Even so, the problem is that, today, many philosophers have no interest in skepticism and whether certain bits of our knowledge are skeptic-proof. Skepticism is often seen as an embarrassment to contemporary philosophy, an exaggerated and arcane anxiety from the 17th century. Now it is true that there are similarities between the usual skepticism and the skeptical problem about content we have discussed. [See Bar-On (forthcoming) for many of these similarities.] Both utilize arguments of the following form [featured in DeRose (1995)], where ‘H’ is a skeptical hypothesis, and ‘O’ is something you ordinarily take yourself to know:

- (i) If I know that O, then I know that \sim H.
- (ii) I do not know that \sim H.
- (iii) So, I do not know O.

H for the external-world skeptic will be something like “I am a brain-in-a-vat” and O will be some external-world proposition like “I have hands.” Whereas, the content-skeptic’s H would be the slow-switching hypothesis, and O would be something like “My current W-

thoughts are *de* the liquid from home.” In both cases, the idea would be that you don’t know what you think you know, on the assumption that you do not know $\sim H$.

However, there is an important difference which, I think, makes content-skepticism the more interesting one. External-world skepticism feeds off the paranoia that you may be *in total error* about the external world. Yet content-skepticism, even when generalized to all thoughts, need not suggest you may be in massive error. Indeed, it seems to me the content-skeptic can grant the truth of virtually *all* of your assertions—and do so *even when their justification is* a posteriori. In the *a priori* case, of course, we saw that the challenge-question can be pressed even if (8) is true. Basically, that’s because (8) turns out to be true both English and Twin-English. But notably, the following *a posteriori* claims would also be true in both languages:

(12) My W-thoughts are *de* the clear, odorless liquid found in lakes around here.

(13) My W-thoughts are *de* that [where the local watery stuff is ostended].

Since all these are true on both planets, that means that the skeptic can *grant* their truth—yet given the slow-switch hypothesis, he can still press the question of whether your W-thoughts are *de* home-water or not. Thus even granting these truths, the content-skeptic suggests you *still* do not “know what” your W-thoughts are about, for the purpose of answering his challenge-question.

Notably, if the slow-switch hypothesis is deviant enough, the skeptic could even grant you the truth of the following assertion as well:

(14) My W-thoughts are *de* hydrogen oxide.

That’s because the skeptic might raise the possibility that (14) is also true in both English and Twin-English. Indeed, model theory illustrates that one can reinterpret English to

preserve the actual truth-value of each sentence, while changing the denotation of its terms. Thus we could imagine that Twin-English has all English names for water as names for XYZ, and also that any English predicate satisfied by water on Earth is a predicate XYZ satisfies on Twin-Earth (and vice-versa). In this way, the model for Twin-English would be isomorphic to the model for English, meaning that not only would (14) as well as (8), (12), and (13) be true in both languages—but also the following:

(15) My W-thoughts are *de* the watery stuff on Earth (and not on Twin-Earth).

We could suppose (15) is true in Twin-English if ‘Earth’ refers to Twin-Earth and ‘Twin-Earth’ refers to Earth. Consequently, the content-skeptic could grant the truth of (15)—yet still, this leaves open which of two stuffs your W-thoughts are about. And more generally, given the isomorphism, it seems that virtually *anything* you say could be granted by the skeptic, without harming his point.

The skeptical argument, then, bears a striking similarity to the pessimistic argument from Chapter One. Both arguments are raising the possibility of non-standard denotations for your (linguistic or mental) representations (In the present case, the non-standard isomorphic model being considered one which simply swaps H₂O with XYZ, and makes minor adjustments to the extensions of certain predicates.) Subsequently, both arguments raise a question about what is denoted by some representation and its co-referring representations—and both indicate a limit on providing an answer which truly settles the question.

However, as things currently stand, there is a crucial difference between the linguistic and the mental cases. Chapter One demonstrated there *cannot* be an answer which settles the question of what a term denotes in the standard model. Socrates, if you

recall, could grant the truth of *every* assertion about the denotation of ‘Pollux’, yet still have a question concerning its standard denotation. But in the mental case, although the skeptic can grant *virtually* all of your assertions regarding the object of your W-thoughts, there is one exception. When asked by the skeptic whether your W-thoughts are *de* home-water or not, one response that could settle the issue is:

(16) My W-thoughts are *de* water from home.

The reason why this might settle the issue is that the skeptic cannot raise two possible interpretations of ‘home.’ After all, that term was key to raising the challenge-question in the first place. (The question assumed that ‘home’ was used to denote *home* as opposed to some other place.) So if a skeptic started to question this answer, he would be inconsistent to assume and not assume a unique interpretation of the term.¹⁹

Now there is a question of whether the Moorean is entitled to his anti-skepticism, in both the case of content and the case of the external world. But that is not something I mean to discuss. My point is rather that, since content-skepticism does not depend on the possibility of systematic *error*, the content-skeptic is importantly different from the external-world skeptic. The content-skeptic is vastly more charitable. And that makes him a bit more interesting, I think, than his external-world cousin.

¹⁹ However, I think that the content-skeptic could ask a similar question which has no adequate answer. In particular, we can make the parallel with Chapter One strict, where the content-skeptic instead asks “What are your W-thoughts about (*de re*)?” This formulation of the question avoids treating the term ‘home-water’ as having special status. Moreover, we can obviate introducing the term ‘home-water’ by just defining the standard model by using the term ‘water.’ Granted, however, you could correctly respond the skeptic in this case by saying “My W-thoughts are *de* water.” Still, it may not be an appropriate answer since it is rather uninformative to say “The thoughts I express using the term ‘water’ denote water.” However, this line requires further development, and it is undesirable to attempt this at the present time. But for the record, I think one could make a sound argument in the case of mental representations that is more strictly parallel to the Chapter One argument vis-à-vis linguistic representations.

BIBLIOGRAPHY

- Azzouni, J. (2004). *Deflating Existential Consequence: A Case for Nominalism*, New York: Oxford UP.
- _____. (2007). Ontological Commitment in the Vernacular,” *Noûs*, 41(2), 204-226.
- _____. (in progress). Ontological Dependence, unpublished manuscript.
- Bar-On, D. (forthcoming). Externalism and Skepticism: Recognition, Expression, and Self-Knowledge. [Title of Anthology Unknown], Oxford: Oxford UP.
- Bar-On, D., Horisk, C., & Lycan, W. (2000). Deflationism, Meaning, and Truth-Conditions, *Philosophical Studies*, 101, 1-28.
- Barwise, J. & Perry, J. (1981). Semantic Innocence and Uncompromising Situations, *Midwest Studies in Philosophy*, 6, 387-403.
- Boër, S. & Lycan, W. (1986). *Knowing Who*, Cambridge, MA: Bradford Books.
- Brentano, F. (1874). *Psychology from an Empirical Standpoint*, translation by Rancurello, Terrell and McAlister, New York: Humanities Press, 1973.
- Brown, J. (1998). The Incompatibility of Anti-Individualism and Privileged Access, *Analysis*, 55, 149-156.
- _____. (2004). *Anti-Individualism and Knowledge*, Cambridge, MA: MIT Press.
- Brueckner, A. (1999). Difficulties in Generating Skepticism about Knowledge of Content, *Analysis*, 59(1), 59-62.
- Brueckner, A. (2003). The Coherence of Scepticism about Self-Knowledge. *Analysis* 63(1), 41-48.
- Burge, T. (1979). Individualism and the Mental, *Midwest Studies in Philosophy*, 4, 73-121.
- _____. (1982). Other Bodies, in *Thought and Object*, Woodfield, A. (ed.), New York: Oxford UP, 97-120.
- _____. (1988). Individualism and Self-Knowledge, *The Journal of Philosophy*, 85(1), 649–663.
- _____. (2006a). Introduction, in *Foundations of Mind*, New York: Oxford UP, 1-31.
- _____. (2006b). Descartes on Anti-Individualism (with supplement), in *Foundations of Mind*, New York: Oxford UP, 420-439.

- Carnap, R. (1947). *Meaning and Necessity*, Chicago: U of Chicago Press.
- Davidson, D. (1967). Truth and Meaning. In *Inquiries into Truth and Interpretation*, 2nd edition (2001), Oxford: Clarendon Press, 17-42.
- _____. (1984). Belief and the Basis of Meaning, in *Inquiries into Truth and Interpretation*, 2nd edition (2001), Oxford: Clarendon Press, 141-154.
- _____. (1996). The Folly of Trying to Define Truth, in *Truth, Language, and History* (2005), Oxford: Clarendon Press, 19-38
- DeRose, K. (1995). Solving the Skeptical Problem, *The Philosophical Review*, 104(1), 1-52.
- Dretske, F. (1970). Epistemic Operators, *Journal of Philosophy*, 67, 1007–1023.
- Dummett, M. (1974). What is a Theory of Meaning? II, in *The Seas of Language*, New York: Oxford UP.
- Evans, G. (1982). *The Varieties of Reference*, New York: Oxford UP.
- Field, H. (1972). Tarski's Theory of Truth. *Journal of Philosophy*, 69, 347-375.
- Fodor, J. (1975). *The Language of Thought*, Cambridge, MA: MIT Press.
- _____. (1990). *A Theory of Content and Other Essays*, Cambridge, MA: MIT Press.
- _____. (1998). *Concepts: Where Cognitive Science Went Wrong*, New York: Oxford UP.
- _____. (2006). *Hume Variations*, New York: Oxford UP.
- Goldberg, S. (1999). The Relevance of Discriminatory Knowledge of Content. *Pacific Philosophical Quarterly*, 80, 136-156.
- Goldman, A. (1976). Discrimination and Perceptual Knowledge. *Journal of Philosophy*, 64, 355-372.
- _____. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard UP.
- James, W. (1907). *Pragmatism: A New Name for Some Old Ways of Thinking*, Hackett Publishing, 1981.
- Kaplan, D. (1978). Dthat, *Syntax and Semantics*, 9, P. Cole (ed.), New York Academic Press.

- Kripke, S. (1972). "Naming and Necessity," in Davidson, D. & Harman, G., (eds.) *Semantics of Natural Language*, Dordrecht: Reidel, 253-355 & 763-769.
- Lewis, D. (1984). Putnam's Paradox, *Australasian Journal of Philosophy*, 62, 221–236.
- _____. (1986). *On The Plurality of Worlds*, Cambridge, MA: Blackwell.
- _____. (1996). Elusive Knowledge, *Australasian Journal of Philosophy* 74.4, 549-567.
- Ludlow, P. (1995). Externalism, Self-Knowledge, and the Prevalence of Slow-Switching, *Analysis*, 55, 45-49. Reprinted in Ludlow and Martin (1998), 225–230.
- _____. (1997). On the Relevance of Slow-Switching, *Analysis*, 57, 285–286. Reprinted in Ludlow and Martin (1998), 235–238.
- Ludlow, P. & Martin, N. (eds.). (1998). *Externalism and Self-Knowledge*, Stanford: CSLI Press.
- Lycan, W. (1979). "The Trouble with Possible Worlds," in Loux, M. (ed.), *The Possible and the Actual*, Ithaca, NY: Cornell UP, 274-316.
- _____. (1984). *Logical Form in Natural Language*. Cambridge, MA: MIT Press.
- _____. (2008) *Philosophy of Language: A Contemporary Introduction*, 2nd edition, London: Routledge.
- Mackie, J. (1965). Causes and conditions, *American Philosophical Quarterly* 2.4, 245-255 & 261–264.
- Millikan, R. (1984). *Language, Thought, and Other Biological Categories*, Cambridge, MA: MIT Press.
- Parent, T. (2007). Infallibilism about Self-Knowledge, *Philosophical Studies*, 133, 411-424.
- Port, R. & van Gelder, T. (eds.). (1998). *Mind as Motion: Explorations in the Dynamics of Cognition*, Cambridge, MA: MIT Press.
- Putnam, H. (1975). The Meaning of 'Meaning', in *Mind, Language, and Reality*, Cambridge: Cambridge UP, 215–271.
- _____. (1976). Models and Reality, in *Realism and Reason: Philosophical Papers*, vol. 3. Cambridge: Cambridge UP (1983), 1-25.

- _____. (1977). Realism and Reason, in *Meaning and the Moral Sciences*. London: Routledge and Kegan Paul (1978), 123-138.
- _____. (1981). A Problem about Reference, in *Reason, Truth, and History*, Cambridge, MA.: Cambridge University Press.
- Rosenberg, J. (1974). *Linguistic Representation*, Dordrecht, Holland: Kluwer.
- _____. (1986). *The Thinking Self*, Philadelphia: Temple UP.
- Russell, B. (1905). "On Denoting," *Mind*, New Series, 14(56), 479-493.
- _____. (1913). "Philosophy for Laymen," in *Unpopular Essays*, New York: Simon and Schuster, 21-33.
- Quine, W. V. O. (1948). On What There Is, *Review of Metaphysics*, 2, 21-38.
- _____. (1960). *Word and Object*, Cambridge, MA: Harvard UP.
- _____. (1969). Ontological Relativity. *Journal of Philosophy*, 65(7), 185-212.
- Rumelhart, D.E. (1989). The Architecture of Mind: A Connectionist Approach, in M. Posner (ed.) *Foundations of Cognitive Science*, Cambridge, MA: MIT Press, 133-159.
- Shaffer, J. (2005). Contrastive Knowledge, in *Oxford Studies in Epistemology, Vol. 1*, Gendler, T. (ed). Oxford: Clarendon Press, 235-271.
- Skolem, T. (1920). Logico-combinatorial investigations on the satisfiability or provability of mathematical propositions: A simplified proof of a theorem by Loewenheim, Reprinted in van Heijenoort, J. (1967). *From Frege to Godel: A Sourcebook in Mathematical Logic, 1879-1931*, Cambridge, MA: Harvard UP, 252-263.
- Soames, S. (2003). *Philosophical Analysis in the Twentieth Century*, 2 vols., Princeton, NJ: Princeton UP.
- Strawson, P. (1969). Meaning and Truth, *Proceedings of the British Academy*, Oxford: Oxford UP.
- Tarski, A. (1935). The Concept of Truth in Formalized Languages. In *Logic, Semantics, and Metamathematics*, 2nd edition, J. Corcoran (ed.), Indianapolis: Hackett, 1981.
- _____. (1944). The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research*, 4, 341-376.

- Warfield, T. (1992). Privileged Self-Knowledge and Externalism are Compatible, *Analysis*, 52, 232-237. Reprinted in Ludlow and Martin (1998), 215–222.
- _____. (1997). Externalism, Privileged Self-Knowledge, and the Irrelevance of Slow-Switching. *Analysis* 57, 282-284. Reprinted in Ludlow and Martin (1998), 231–23
- Wiggins, D. (2002). An Indefinabilist *cum* Normative Theory of Truth and the Marks of Truth, in *What is Truth?* R. Schantz (ed.), Berlin: Walter de Gruyter, 316-332.
- Wittgenstein, L. (1953). *Philosophical Investigations*, trans. G.E.M. Anscombe, Cambridge, MA: Blackwell.