JPhil XCVIII (July 2001): 331–350

On Putnam and his Models *

Timothy Bays

> It is not my claim that the 'Löwenheim-Skolem paradox' is an antinomy *in formal logic*; but I shall argue that it *is* an antinomy, or something close to it, in *philosophy of language.* Moreover, I shall argue that the resolution of the antinomy—the only resolution that I myself can see as making sense—has profound implications for the great metaphysical dispute about realism which has always been the central dispute in the philosophy of language.
>
> Hilary Putnam: Models and Reality

For the past twenty years or so, Hilary Putnam has advanced a family of arguments which go under the patronymic "the model-theoretic argument against realism." These arguments purport to show that basic theorems of model theory entail that realistic accounts of truth and reference are untenable and that, as a result, realistic metaphysics is incoherent. On the basis of these arguments, Putnam urges philosophers to abandon traditional realism—or "metaphysical realism," as he likes to call it—and to adopt Putnam's own, new-and-improved, "internal realism" instead.

In this paper, I examine one version of Putnam's argument—a version closely related to the traditional Löwenheim-Skolem Paradox. I defend three claims about this argument. First, I argue that a key step in Putnam's argument rests on a mathematical mistake, and I discuss some of the philosophical ramifications of this mistake. Second, I argue that, even if Putnam could get his mathematics to work, his argument would still fail on purely philosophical grounds. Third, I argue that Putnam's mathematical mistakes and his philosophical mistakes are surprisingly closely related. At the end of the day, I conclude that realists have little to fear from Putnam and his models.

1 Putnam's Argument

The version of Putnam's argument that I want to consider occurs in the first fifteen pages of a paper entitled "Models and Reality."[1] The paper begins with a short discussion of the Löwenheim-Skolem Paradox and then transforms this paradox in two ways. For expository convenience, I will follow this same progression in laying out Putnam's argument.

*This paper has benefited from discussions with members of the philosophy departments at UCLA, Yale, Notre Dame and Northwestern. I am particularly grateful to Tony Martin, Yiannis Moschovakis and an anonymous *Journal of Philosophy* reviewer for helpful comments on earlier drafts.

[1]Hilary Putnam, "Models and Reality," in *Realism and Reason* (Cambridge: Cambridge UP, 1983), 1–25. Hereafter, MR.

Begin with a standard, first-order axiomatization of set theory—say ZFC. On the assumption that this axiomatization has a model, the Löwenheim-Skolem theorems ensure that it has a countable model. Call this model M. Now, because ZFC $\vdash \exists x$ "*x is uncountable*," there must be some $\hat{m} \in M$ such that

$$M \models \text{"}\hat{m} \textit{ is uncountable.}\text{"}$$

However, since M itself is only countable, there are only countably many $m \in M$ such that $M \models m \in \hat{m}$. Thus, cardinality seems to be relative: from one perspective, $\hat{m}$ seems to be uncountable, while from another perspective, $\hat{m}$ is clearly countable.

Putnam begins his analysis of this paradox by noting that, whatever it might show about the countability or uncountability of $\hat{m}$, the paradox highlights the fact that ZFC has *different* models. Further, these models interpret many of the central definitions of classical set theory in structurally different ways. Some models satisfy "m is uncountable" only if m really is uncountable, while other models (like the M above) satisfy "m is uncountable" even when m is really countable. Similarly, some models satisfy sentences like "m is finite" or "m is the power set of n" if and only if m really is finite or really is the power set of n, while other models satisfy these sentences under (quite) different circumstances.

Skolem's paradox shows, therefore, that if the meanings of phrases like "is countable" or "is finite" are fixed only by the model theory of first-order set theory, then these phrases are semantically indeterminate. To use Putnam's terminology, the paradox shows that it is impossible to pin down the "intended interpretation" of set-theoretic language using only (first-order) axioms. Further, Putnam clearly thinks that axioms are all we have to go on here: "but if *axioms* cannot capture the 'intuitive notion of set' what possibly could?" (MR, 3) As a result, Putnam concludes that even traditional formulations of Skolem's Paradox uncover a genuine, and a fairly general, indeterminacy in our use of set-theoretic language.

So much, then, for the traditional Löwenheim-Skolem Paradox. Putnam's first transformation of this paradox involves parlaying the general indeterminacy of set-theoretic language into more specific indeterminacies in the truth-values of set-theoretic sentences. Putnam writes: "If I am right, then the 'relativity of set-theoretic notions' extends to a *relativity of the truth value of '$V = L$'* (and, by similar arguments, of the axiom of choice and the continuum hypothesis as well)" (MR, 7-8). Putnam's idea here is simple. It is a result of modern set theory that $V = L$ is "independent" of ZFC—that, on the assumption that ZFC itself is consistent, so are ZFC + V=L and ZFC + $V \neq L$. Hence, if the intended model of set theory is fixed *only* by the axioms of ZFC—and if there is, in fact, some such intended model—then there is an intended model in which $V = L$ is true and another in which $V = L$ is false.

Putnam's second transformation involves arguing that, to the extent that axioms fail to ensure the semantic determinacy of set-theoretic language, physical science won't pick up the slack. In particular, physical science won't fix a unique "intended interpretation" for the vocabulary of set theory, nor will it restrict such interpretations so as to eliminate the indeterminacy in truth-values of set-theoretic sentences. Now, at first glance, this second transformation may seem somewhat unmotivated: how, after all, *could* physical science affect the interpretation of basic set-theory? Later glances, however, reveal that Putnam has something to worry about here, and that his worries are closely related to the project of showing that $V = L$ has an indeterminate truth-value.

Suppose we have a machine which takes a measurement—of something, it doesn't matter what—every three or four seconds. Suppose also that this machine gives a reading of 1 or 0 depending on the results of its measurements, and suppose finally that this machine manages to run for an *infinite* period of time and (so) produces an infinite sequence of measurements. In theory, the sequence of ones and zeros which result from these measurements could "code up" a non-constructible set—i.e., a set which lives in V but not in L. In this case, it might seem like *nature itself* manages to falsify the hypothesis that $V = L$. This possibility, together with Putnam's desire to make "$V = L$" come out indeterminate, explains why Putnam thinks he needs an explicit argument concerning the impact of physical science on the interpretation of set theory.

Putnam's explicit argument comes in two parts. The first centers on the following theorem which Putnam states and proves on pp. 6–7 of MR:

Theorem 1: *ZF plus V=L has an ω-model which contains any given countable set of real numbers.*

Given this theorem, Putnam argues as follows. Let OP be a countable collection of real numbers which codes up all of the measurements human beings will ever make (including measurements made by machines like those discussed above). By Theorem 1, there is a model of *ZF + V=L* which contains OP (or, at least, a formal analog of OP). Since this model satisfies ZFC it must be an "intended model," and since it both satisfies $V = L$ and contains OP, it takes care of the measurement problem discussed in the last paragraph.

Next, Putnam argues that this measurement problem is the *only* problem with physical science that *needs* to be taken care of. He writes:

> Now, suppose we formalize *the entire language of science* within the set theory ZF *plus* $V = L$. Any model for ZF which contains an abstract set isomorphic to OP can be extended to a model for this formalized language of science which is *standard with respect to OP*; hence... we can find a model *for the entire language of science* which satisfies '*everything is constructible*' and which assigns the correct value to all physical magnitudes (MR 7).

Thus, as long as the only constraints on the interpretation of set-theoretic vocabulary come from the formal structure of our scientific theories (including the explicit axioms of our set theory) and from the physical measurements we happen to make, there will be *an* interpretation of set theory on which $V = L$ comes out true. On the flip side, Putnam assumes that there is *some* interpretation of set theory—again, an interpretation compatible with the rest of our scientific theories and with all the physical measurements we might ever happen to make—on which $V = L$ comes out false.[2]

[2]Putnam makes this assumption for two reasons. From a dialectical perspective, Putnam takes himself to be arguing against Gödel. Gödel thought there *was* a unique "intended interpretation" of set theory and that $V = L$ comes out false on this interpretation. Hence, the existence of intended models satisfying $V \neq L$ isn't really at issue in this context.

From a somewhat deeper perspective, it's fairly easy to start with a model which *does* satisfy $V = L$ and expand it to one which *does not* (and to preserve nice properties like "being an ω-model" in the process). It's substantially more difficult to start with a model which *does not* satisfy $V = L$ and expand it to one which *does*. This is undoubtedly the main reason Putnam puts so much effort into obtaining a model which satisfies $V = L$ and so little into obtaining one which satisfies $V \neq L$.

This, then, gives us an initial premise in Putnam's model-theoretic argument. Let the phrase "theoretical constraints" refer to the set of sentences which constitutes our best theory of the world—i.e., our best physical theory together with our best axioms for set theory—and let "operational constraints" refer to all the measurements we might ever happen to make. Then we have:

1. Theoretical and operational constraints do not fix a unique "intended interpretation" for the language of set theory.

Further, Putnam clearly thinks that theoretical and operational constraints are the only factors which *could* fix an "intended interpretation" for set theory. His reasons for thinking this will be discussed in section 3. For now, let me simply lay out the remainder of his argument:

2. Nothing other than theoretical and operational constraints could fix a unique "intended interpretation" for the language of set theory.

So, 3. There is no unique "intended interpretation" for the language of set theory.

Finally, because different, equally "intended," interpretations of set theory disagree on the truth value of sentences like $V = L$, there simply are no determinate truth values for such sentence. They are, in Putnam's words, "just true in some intended models and false in others" (MR 5).

This, therefore, gives us the overall structure of Putnam's argument. The goal is to show that set-theoretic language is semantically indeterminate. To reach this goal, we first note that the axioms of set theory do not determine a unique interpretation for set-theoretic language. Next, we observe that throwing in scientific information—i.e., the physical theories and measurements which fill out Putnam's "theoretical and operational constraints"—will not improve the situation. Finally, we notice that the different "intended interpretations" which are compatible with our "theoretical and operational constraints" disagree on the truth value of sentences like $V = L$, and we conclude that these truth values are, themselves, indeterminate.

2 The Mathematics of Premise 1

In this section, I slow down and examine more carefully the details of Putnam's mathematics. My main theses are straightforward. I argue 1.) that Putnam's proof of Theorem 1 is mistaken, 2.) that this mistake cannot be "patched up" without weakening the overall force of Putnam's argument, and 3.) that even the weakened version of Putnam's argument leaves the realist with significant problems.

To begin, consider Putnam's proof of Theorem 1. The theorem says that if X is a countable collection of real numbers, then there exists an ω-model, M, such that $M \models ZF{+}V{=}L$, and M contains an "abstract copy" of the set X. Putnam's proof begins by noting that, in the special case in which we allow M to be countable, we can code both M and X by single reals. In this case, the theorem can be formulated as a Π_2 sentence of the form: (For every real s) (There is a real M) such that $(\ldots M, s, \ldots)$. From here, Putnam argues as follows:

> Consider this sentence *in the inner model* $V = L$. For every s *in the inner model*—i.e., for every s in L—there is a model—namely L itself—which satisfies '$V = L$' and contains s. By the downward Löwenheim-Skolem theorem, there is a countable submodel which is elementarily equivalent to L and contains s. (Strictly speaking, we need here not just the downward Löwenheim-Skolem theorem, but the 'Skolem Hull' construction which is used to prove that theorem.) By Gödel's work, this countable submodel itself lies in L, and, as is easily verified, so does the real that codes it. So, the above Π_2-sentence is true in the inner model $V = L$.
>
> But Shoenfield has proved that Π_2-sentences are *absolute*: if a Π_2-sentence is true in L, then it must be true in V. So the above sentence is true in V. (MR 6)

Ironically, the problem with this "proof" involves Putnam's application of the Löwenheim-Skolem theorem. The "short version" of the problem is simple: the downward Löwenheim-Skolem theorem applies only to structures which have sets for their domains, and L—the structure to which Putnam applies the downward Löwenheim-Skolem theorem—doesn't have a set for its domain. Hence, Putnam cannot (legitimately) use the downward Löwenheim-Skolem theorem to obtain "a countable submodel which is elementarily equivalent to L and contains [the real] s." Without this submodel, Putnam cannot ensure that L satisfies the Π_2 sentence which he needs it to satisfy; hence, he cannot apply Shoenfield absoluteness to "reflect" this sentence up to V. In the absence of Putnam's countable submodel, therefore, Putnam's overall proof simply collapses.[3]

[3] The "long version" of this problem simply expands the "short version" with a fair bit of terminological clarification. To begin, set theorists typically distinguish between *sets* and *proper classes*. Roughly, sets are classes which are "small enough" to be members of other classes, while proper classes are classes that are "too big" to count as sets. Examples of sets would include $\emptyset$, $\aleph_0$, and the power set of $\mathbb{N}$. Examples of proper classes would include the class of all sets, the class of all ordinals, and the class of all countable sets. Most pertinently, the class of all constructible sets—i.e., L—is a proper class.

It is important here, for reasons which will become clear shortly, that proper classes are usually required be *definable*. That is, for some formula in the language of set theory, $\phi(x, y_1, \ldots, y_n)$, and some sequence, $a_1, \ldots, a_n$, of sets, we can consider the class of all sets b such that $\phi(b, a_1, \ldots, a_n)$ is true. To insist that proper classes be definable is to insist that *all* classes be picked out by a formula of this kind.

Second, we need to distinguish between two interpretations of the terms "model" and "satisfaction." In ordinary model theory, the term "model" refers exclusively to structures which have *sets* for their domains. In turn, the "satisfaction" relation is defined as a relation *between sets*—i.e., between the sets which constitute the domain (and relations) of a model and those which code up the formulas of our language. Set theorists often use the terms "model" and "satisfaction" differently. They speak of "class models" when they want to refer to proper classes in which certain collections of sentences hold, and they use "satisfaction" to refer to the fact that certain sentences become true when their quantifiers are relativized to a proper class—i.e., explicitly relativized using the formula which defines the class in question (see above). It is in this latter sense, for instance, that Putnam refers to "inner models."

For our purposes, the key fact concerning all this terminology is the following: the Löwenheim-Skolem theorems apply to set models, but they do not apply to class models. Hence, we cannot use the downward Löwenheim-Skolem theorem to find an elementary submodel of some class model of ZF. In particular, Putnam cannot use the downward Löwenheim-Skolem theorem to find a (countable) model which is elementarily equivalent to L.

Given this, Putnam faces a dilemma. If he intends the term "model" to refer to set models, then his proof goes wrong when he says: "for every s in L—there is a model—namely L itself—which satisfies '$V = L$'." For, since L is not a model at all, it is not a model satisfying $ZF + V = L$. If, on the other hand, Putnam intends the term "model" to refer both to set models and to class models, then he can legitimately call L a "model." But even then, *he cannot apply the downward Löwenheim-Skolem theorem to this model*. On any interpretation, then, Putnam's proof is mistaken. However "model" gets defined, Putnam's proof requires that we apply the downward Löwenheim-Skolem theorem to L, and, as we have seen, this just isn't possible.

2.1 Five Technical Comments

Before I discuss the philosophical ramifications of the failure of Putnam's proof, I want to make five technical comments about the failure itself. First, Putnam's proof is not saved by his qualification: "Strictly speaking, we need here not just the downward Löwenheim-Skolem theorem, but the 'Skolem Hull' construction which is used to prove that theorem." To be sure, the Skolem Hull construction allows us to prove so-called "reflection theorems" in which some *finite* collection of sentences is "reflected" from a proper class to a set. That is, if we start with a proper class which "satisfies" some finite collection of sentences, then the Skolem Hull construction lets us find a countable set which satisfies the same collection of sentences.[4] However, this construction *only* works when when we try to reflect *finite* collections of sentences. In particular, then, it does not allow Putnam to reflect the *infinite* collection, $ZF + V = L$.

Second, there is nothing particularly surprising about the fact that Putnam's proof fails. Leaving aside the details of this proof, consider just the *form* of Theorem 1: for any countable set of reals X, there is an ω-model M, such that $M \models ZF{+}V{=}L$ and $X \in M$. Now, since any model of $ZF + V{=}L$ is also a model of ZFC, Theorem 1 entails that there is a model of ZFC. And since this, in turn, entails that ZFC is consistent, Theorem 1 also entails that ZFC is consistent.

However, by Gödel's second incompleteness theorem, the consistency of ZFC *cannot* be proved within ZFC itself. As a corollary, then, Theorem 1 cannot be proved in ZFC. Therefore, we should not be surprised to find that there is a mistake in Putnam's proof: Putnam's proof must be mistaken, because Theorem 1 *can't* be proved using the set theory with which Putnam is working. Hence, unless Putnam is willing to adopt some stronger set theory, his overall argument is bound to fail.[5]

Third, Putnam's proof is relatively easy to "patch up." If Putnam *is* willing to adopt a stronger set theory, then he can salvage his theorem (and, for the most part, his proof). So, for instance, if Putnam were willing to accept the existence of inaccessible cardinals, then his proof could be reconstituted with only minor modifications.[6] Similarly, if Putnam were willing to extend ZFC with a collection of axioms governing the behavior of proper classes, then he could probably prove "full reflection theorems" which would allow him to obtain (elementary) submodels of proper classes like V and L; this would, once again, allow him to

[4] Here, we have an example of a *legitimate* shift between the two uses of "satisfies" that I mentioned in the last footnote.

[5] In conversation, several people have suggested that Putnam might not be working in ZFC in the first place—i.e., that he might *already* be working in some stronger form of set theory. But, while this suggestion helps to save Putnam's mathematics, it faces two major difficulties. The first is textual: Putnam's entire paper is *about* ZFC, and throughout the paper Putnam writes as though ZFC is the *obvious* set theory in which to work. Hence, it would be strange to find that Putnam's entire argument rests on an (utterly unmentioned) version of set theory which Putnam keeps hidden in the background. The second difficulty is philosophical: for reasons which will be discussed shortly, working with a stronger set theory would not, in the long run, save Putnam's argument (although it *might,* in the short run, allow Theorem 1 to be proved properly).

[6] To see this, let κ be the inaccessible cardinal in question. At the point in Putnam's proof where he claims that L is a model for $ZF + V{=}L$ and that L contains the real s, Putnam can simply argue that L_κ is a model for $ZF + V{=}L$ and that L_κ contains the real s. Since L_κ *really is* a set model for $ZF + V{=}L$, Putnam can proceed to apply the downward Löwenheim-Skolem theorem to L_κ to obtain his desired model M. From here, Putnam's proof goes as before.

reconstitute the essentials of his original proof. In either case, therefore, a slight strengthening of Putnam's background mathematics allows Putnam to save his proof from the problem discussed above.

Fourth, although these "patching strategies" allow Putnam to salvage his proof, they do very little towards salvaging his overall philosophical argument. Very roughly, patchings of the type just mentioned leave Putnam's critics with two fairly obvious lines of response. On the one hand, some philosophers might reject Putnam's argument simply because they reject his new mathematics. Whereas ZFC is a relatively widely-accepted axiomatization of set theory, the extensions discussed above are less-widely accepted. Hence, a philosopher (or mathematician) who has reservations about inaccessible cardinals and/or strong class axioms might well reject Putnam's new argument just because it employs this extra mathematics. Indeed, such a philosopher might even think that she has been given new *reasons* for rejecting Putnam's extra mathematics: "if Putnam's new math generates arguments for semantic indeterminacy, then so much the worse for Putnam's new math."

On the other hand, even philosophers who *do* accept Putnam's strengthened mathematics—say, those who accept the axiom of inaccessible cardinals—have ample grounds for rejecting Putnam's overall argument. In particular, they should reject Putnam's claim that his model M—the model guaranteed by Theorem 1—satisfies "all theoretical constraints." Since Theorem 1 does not guarantee that M satisfies the sentence "there exists an inaccessible cardinal," M may not even satisfy the "theoretical constraints" imposed by *set theory.* Contra Putnam, then, there is no reason to think that M provides an "intended interpretation" for set-theoretic language.

Recall, here, the philosophical *point* of Putnam's theorem. Putnam wants a model of *ZF + V=L* which "satisfies all theoretical constraints... [and] all operational constraints as well" (MR 7). His theorem is supposed to provide such a model: the theory of the model takes care of "theoretical constraints" and the fact that s is a member of the model takes care of "operational constraints." With this model in hand, Putnam tries to argue that $V = L$ is true in the (or at least in *an*) intended model of set theory.

My point is simply this: if our working set theory is *stronger* than ZFC—because we accept inaccessible cardinals, or class axioms, or whatever else is needed to patch up Putnam's proof—then it's hard to see how Putnam's theorem accomplishes its goal. Given that we accept more mathematics than ZFC, this new mathematics should count as part of our "theoretical constraints." Thus, it's not enough for Putnam to build a model which satisfies *ZF + V=L;* Putnam needs a model which satisfies *ZF + V=L plus whatever else we happen to have added to* ZFC. Since Putnam's theorem does not, so far as we know, provide a model satisfying this *extended* theory, it doesn't do what Putnam wants it to do.

Fifth, this problem is intrinsic to the *kind* of argument Putnam wants to make. Returning again to the incompleteness theorem, I note that there is no way for Putnam to both 1.) use a particular axiomatization of set theory (say, ZFC + XYZ) as his background set theory and 2.) prove the existence of a model satisfying ZFC + XYZ +V=L. Hence, *whatever* version of set theory Putnam winds up working in, he will be unable to tailor his overall argument so as to take care of the "theoretical constraints" this version imposes.

As a result, Putnam faces an *inescapable* dilemma. If he pitches his argument towards philosophers who accept less set theory than he himself does, then these philosophers will reject his argument simply because they reject the set theory used in proving Putnam's key theorem. If he pitches his argument towards philosophers who accept the same set theory that Putnam does, then his argument can't take care of *these philosopher's* "theoretical constraints." At the end of the day, therefore, *no one* has adequate grounds for accepting Putnam's model-theoretic argument.[7]

This, therefore, gives us a first, and essentially technical, response to Putnam's argument. Putnam's argument depends on a key theorem which Putnam is not in a position to prove. Nor, for reasons relating to Gödel's second incompleteness theorem, can he place himself in such a position without jeopardizing the very philosophical point which his theorem is supposed to support. In short: the mathematical mistake which we discussed at the beginning of this section is one which cannot be fixed without undercutting the model-theoretic argument as a whole.

2.2 Two Philosophical Comments

How strong is this argument which I have just sketched? Unfortunately, much as I like the argument, I'm afraid the answer is: "not very." On the positive side, the argument shows that Putnam cannot *conclusively prove* that set theory is semantically indeterminate. That is, if we want Putnam to stand toe-to-toe with the realist and *prove* semantic anti-realism, then my argument shows that he cannot do it.

On the negative side, Putnam has clearly given an argument which raises the *possibility* of semantic indeterminacy. Suppose that premise 2 in Putnam's argument is correct. Then the only way for set-theoretic language to be semantically determinate is for there to be a *unique* model which satisfies all our "theoretical and operational constraints." To the extent that alternative models *happen to exist,* set theory winds up being semantically indeterminate.

In this context, the mere fact that Putnam cannot *prove* his central theorem should provide very little comfort to the realist. If the technical response to Putnam's argument is all we have to go on—if, that is, we are willing to accept premise 2 and to rest our challenge to premise 1 solely on the considerations discussed above—then realism depends on the *mere hope* that Putnam's non-standard models don't exist. This not much to stake a metaphysics on![8]

All that being said, this argument clearly depends on the assumption that premise 2 is true. If it is not, then the existence of non-standard models for our "theoretical and operational constraints" becomes considerably less troubling. It is high time, therefore, that we turn and examine premise 2.

[7]Note that this argument is not specific to the search for a model satisfying $V = L$; it works equally well against attempts at finding models for ZFC $\pm$ CH or ZF $\pm C$. Indeed, when framed as a simple incompleteness problem, it applies even to the "plain vanilla" version of Skolem's Paradox which simply tries to find structurally dissimilar models for ZF.

[8]Note, for instance, that we have to hope that there are no large cardinals stronger than the ones explicitly mentioned in our "theoretical constraints." Similarly, we have to hope that a great many small models (including those witnessing the consistency of our *own axioms*!) don't exist. While this is certainly a *consistent* hope, that's about all that can be said for it.

3 Just more theory I

Although premise 1 of Putnam's model-theoretic argument has attracted a great deal of attention in the literature—and although it's clearly the premise that gives Putnam's argument most of its philosophical "sex appeal"—it should be fairly clear that premise 2 is where the real philosophical action takes place. For one thing, premise 1 only eliminates one method for fixing the intended interpretation of set theory, while premise 2 eliminates all other methods in one fell swoop. For another thing, most of the methods that have actually been proposed for fixing the intended interpretation of our language fall under premise 2 rather than premise 1.[9] Hence, 2 is the premise which warrants the most sustained philosophical attention.

The heart of Putnam's defense of premise 2 is the observation that the phrase "theoretical constraints" is broad enough to encompass *philosophy* as well as mathematics and natural science. In particular, any philosophical account of the way set theory gets its "intended interpretation" can itself be viewed as just one more theoretical constraint. Hence, no such account allows us to evade the problems introduced by premise 1: since Putnam can always find an assortment of models which satisfy *both* our original theoretical constraints *and* our new philosophical semantics, the philosophical semantics do not force our language to take a unique "intended interpretation."[10]

In effect, then, Putnam argues that premise 2 is *vacuously* true. It's not that there are a variety of genuine alternatives to using theoretical and operational constraints to fix the intended interpretation of set theory (and that Putnam's argument for premise 2 takes care of these genuine alternatives). Rather, it's that every *seeming* alternative turns out, upon reflection, to be a *special case* of "theoretical and operational constraints." Hence, every alternative is really covered by the argument for premise 1. In short: because the phrase "theoretical and operational constraints" is flexible enough to *commandeer* rival mechanisms for fixing intended interpretations, all such mechanisms are ultimately taken care of by premise 1.

Now, this argument for premise 2—the so-called "just more theory" argument—has been extensively discussed in the literature. It is useful, therefore, to outline a standard objection to this argument and to sketch Putnam's response to that objection.[11] The objection comes in two parts. First, we distinguish

[9] E.g., causation for language referring to material objects and set-theoretic "perception" for language referring to sets.

[10] This way of formulating the argument may be slightly misleading. Putnam thinks that, to the extent that our philosophical semantics are any good, they should *already count* as part of our "original theoretical constraints." Hence, they should *already be covered* by his argument for premise 1.

[11] The objection at issue has been raised in various forms by various people (see, e.g., chapter 11 of Michael Devitt, *Realism & Truth* (Princeton: Princeton University Press, 1984), David Lewis, "Putnam's Paradox," *Australasian Journal of Philosophy* 62 (1984): 221–236, and James Van Cleve, "Semantic Supervenience and Referential Indeterminacy," *The Journal of Philosophy* 89 (1992): 344–361). Putnam's response to this objection comes in the introduction to his *Realism & Reason,* pp. x–xii.

It's worth noting that several recent papers claim that this objection rests on a misinterpretation of Putnam's original argument (see David Anderson, "What is the Model-Theoretic Argument," *The Journal of Philosophy* 93 (1993): 311–322 and Igor Douven, "Putnam's Model-Theoretic Argument Reconstructed," *The Journal of Philosophy* 96 (1999): 479–490). Although I disagree with this claim, it would take us too far afield to discuss the relevant interpretive questions here. I discuss these questions in "Revisionary Interpretations of the Model-Theoretic Argument," (in preparation).

between *describing* the features of a model which make it an "intended interpretation" and simply *adding new sentences* for model to satisfy in order to count as an intended interpretation. More perspicuously, we distinguish between *changing* the semantics under which a collection of axioms (or "theoretical constraints") gets interpreted—e.g., by restricting the class of structures which count as "models" for our language and/or strengthening the notion of "satisfaction" which ties sentences to models—and simply *adding new axioms* to be interpreted under the same old semantics.

Second, we note that many natural conditions on the intended interpretation of set theory fall on the description side of this distinction. When a philosopher claims that the intended models of set theory should be transitive, she is *describing* the structures which are to count as models for her axioms; she isn't just adding new sentences to be interpreted at Putnam's favorite models. Similarly, when she claims that intended models should satisfy second-order ZFC, she is explaining *which semantics* (and, more specifically, which satisfaction relation) her axioms should be interpreted under; she isn't just adding new axioms to be interpreted under a (first-order) semantics of Putnam's choosing.

Now, if all this is correct, then Putnam's "just more theory" argument is in trouble, since it clearly assumes that all conditions on intended interpretations are really of the "adding axioms" variety. How can Putnam respond? Basically, Putnam claims that this objection *begs the question* against the model-theoretic argument. The argument, after all, involves the question of whether mathematical language has determinate significance, and the realist seems to *assume* that it has such significance when she uses phrases like "transitive," or "complete power-set" to describe her notion of "intended model" (or to explain her preferred semantics for interpreting axioms). To adapt a passage from *Realism & Reason*:[12]

> Here the philosopher is ignoring her own epistemological position. She is philosophizing as if naive realism were true of her (or, equivalently, as if she and she alone were in an *absolute* relation to the world). What *she* calls 'transitivity' really is transitivity, and *of course* there is a fixed, somehow singled-out, correspondence between the word and one definite property in *her* case. Or so she assumes. But how this can be so was just the question at issue. (*RR,* xi)

On Putnam's view, then, as long as the determinacy of mathematical language is still at issue, realists cannot—on pain of begging the question—use this language to describe the intended interpretation of their theories and/or to specify the semantics under which their axioms should be interpreted.

To summarize: Putnam's "just more theory" argument for premise 2 comes in two stages. First, Putnam considers realistic theories of interpretation and argues that these theories can be regarded as additional "theoretical constraints." As such, they do they do not add anything new to the "theoretical and operational constraints" of premise 1. Second, Putnam considers the claim that realistic theories *describe* the ways interpretations are fixed (or, equivalently, describe the kinds of semantics under which our axioms are to be interpreted). Since this claim assumes that we can use mathematical language to describe, in a determinate fashion, the intended interpretation of set theory, Putnam dismisses it as question-begging.

[12]This passage originally concerned the causal theory of reference; I have modified it to fit the mathematical case. In particular, I have replaced the term "causality" with "transitivity" in two places, and the phrase "one definite relation" with "one definite property." Finally, I've changed the gender of Putnam's philosopher to fit that in the main text.

4 Just more theory II

In this section, I give three responses to Putnam's "just more theory" argument. First, I show that this argument is incompatible with some of Putnam's other theoretical commitments; hence, whatever merits it may have in its own right, it is not an argument which is available to Putnam. Second, I show that analogous arguments can be made in situations where these arguments clearly don't work, and I argue that, by parity of reasoning, Putnam's argument doesn't work either. Finally, I give a straightforward explanation as to *why* Putnam's argument doesn't work: in brief, I show that Putnam's charge of question begging rests on a (fairly trivial) logical mistake. At the end of the section, I conclude that Putnam's "just more theory" argument provides no support for premise 2.

To begin, note that Putnam's charge of question begging in the last section rests upon an assumption: that it is illegitimate to use semantically *in*determinate language to describe the intended interpretation of set theory. After all, the problem with realists' attempts to use phrases like "is transitive" or "satisfies second-order ZFC" to describe the intended interpretation of set theory is that we cannot show that these phrases have determinate interpretations (and that simply *assuming* they have determinate interpretations would beg the question). For this objection to have force, however, we must first think that there is something *wrong* with using semantically indeterminate language to describe the intended interpretation of set theory. Only against the backdrop of this assumption would the realist even *need* to claim that, e.g., "is transitive" or "satisfies second-order ZFC" are semantically determinate.

Next, we need three technical facts about first-order logic. First, first-order logic lacks the resources to capture the notions of finitude and recursion. On the finitude side, there is no collection of formulas which characterizes exactly the finite models (or finite subsets of models). On the recursion side, there is no collection of formulas which captures the notion of a *recursive* definition. Second, the notion of finitude is needed to characterize the syntax of first-order logic: the sentences of first-order logic can be of arbitrary *finite* length, but they cannot be infinite.[13] Third, the definition of satisfaction for first-order logic is recursive: it starts with a definition for atomic formulas and then supplies recursion clauses—one for each connective and quantifier in our language—to extend this definition to arbitrary formulas.

Combining these technical facts with the assumption from the penultimate paragraph, we can see that Putnam has gotten himself into hot water. The following four claims are on the table:

1. The notions of finitude and recursion are needed to describe first-order model-theory.
2. First-order model theory cannot capture the notions of finitude and recursion.
3. It is illegitimate to use semantically indeterminate notions to describe "intended interpretations."
4. Only those notions which can captured by first-order model theory are semantically determinate.

[13]Note that this convention is crucial to the formulation of first-order ZFC. Two of the axioms of ZFC—comprehension and replacement—are really *axiom schemas.* That is, they are really infinite *lists* of axioms, one for every formula in the language of set theory. Hence, a firm grasp on the notion "arbitrary formula" is required to *formulate* the axioms of first-order ZFC.

Here, claims 1 and 2 are repetitions of the technical facts discussed in the last paragraph, and claim 3 is the assumption on which Putnam's charge of "begging the question" was seen to rest. Claim 4 is a version of Putnam's own premise 2: it ensures that the language of set theory is no more determinate than the argument for premise 1 would allow.[14]

Together, these four claims cause serious problems for Putnam. From claims 3 and 4, it follows that only notions which can be captured by first-order model theory can be used to describe the "intended interpretation" of set theory. Combining this with claim 2, we get that the notions of finitude and recursion cannot be used to describe the intended interpretation of set theory. But this fact, together with claim 1, entails that first-order model theory cannot be used to describe the intended interpretation of set theory. Since this is precisely what Putnam's own arguments require us to do (e.g., in the argument for premise 1), Putnam's position is internally inconsistent.

This, then, is what I like to call the "stability" objection to Putnam's argument. The argument rests on the assumption that we cannot use semantically indeterminate language to describe "intended interpretations." But, by Putnam's *own standards,* the notions needed to formulate first-order model theory turn out to be semantically indeterminate. So, since Putnam's own techniques for obtaining intended interpretations involve first-order model theory, his overall position is logically unstable.

Leaving this stability objection aside, I note that Putnam's argument is also rendered suspect by the company it keeps. Consider the old, childhood puzzle about the three men and the thirty dollars:

> Three men walk into a bar. They order drinks. At the end of the evening, the bartender presents them with a bill for $30. Each man pays $10, and the three men leave the bar. Once they have left, the bartender realizes that the bill should only have been $25, so he gives the busboy $5 to take to the three men. The busboy, being a dishonest chap, gives each of the men $1 and pockets the remaining $2 for himself. At the end of the day, then, each of the three men has paid $10 and gotten $1 back. Now, $10 - 1 = 9$, and $3 \times 9 = 27$. If we add in the $2 in the busboy's pocket, we are left with $29. So, what happened to the missing $1?

Now, suppose that someone presents this puzzle as a serious challenge to arithmetic—say, as a challenge to the determinacy of ordinary talk about the natural numbers. Suppose also that we respond to this suggestion by giving the obvious solution to the puzzle—i.e., by explaining where the puzzle's arithmetic goes awry.[15] Why, in this case, can't the proponent of the puzzle simply follow Putnam and claim that this solution "begs the question"? After all, the puzzle is supposed to challenge the determinacy of ordinary talk about the natural numbers, and we have to *use* such talk to explain why the puzzle is mistaken.[16]

[14] That is, it ensures that notions which aren't fixed by (the first-order models of) our "theoretical and operational constraints" aren't fixed by other means either. These notions remain indeterminate.

[15] For those who haven't seen this solution, it's really quite simple. The three men *did* pay $27. This $27 can be split into two parts: $25 which is inside with the bartender and $2 which is in the busboy's pocket. There is also $3 which the men got back from the bartender (via the busboy). As expected, $27 + 3 = 30$. The puzzle turns, therefore, on a trick. Instead of adding the $3 which the men received to the $27 which they payed, the puzzle tries to convince you to add the $2 in the busboy's pocket to the $27 and to simply ignore the $3. Since the $2 is *already part* of the $27, this is clearly illegitimate.

[16] I.e., we have to refer to numbers like 2, 3, 25, 27, 29 and 30. Indeed, we even employ higher-order operations on these numbers: addition, subtraction, multiplication, etc.

This, then, gives us an argument which is structurally analogous to (one part of) Putnam's "just more theory" argument. Clearly, however, this new argument does not work. The "paradox of the three men" is not a real paradox, the solution given in footnote 15 is a good solution, the fact that this solution uses number theory is irrelevant, and we don't beg any questions in saying all of this. Given the structural similarities between this failed argument and Putnam's own "just more theory" argument, therefore, I think that considerations of parity provide us with ample grounds for rejecting the latter.

Of course, it would be nice if we didn't have to rely on parity—or on theoretical conflicts with *Putnam's* other views—to get a response to the "just more theory" argument. Fortunately, it's relatively easy to explain *why* Putnam's charge of question begging misses the mark. In evaluating Putnam's model-theoretic argument we are interested in the question: does Putnam's model theory entail that mathematical language is semantically indeterminate? To answer this question, we need to evaluate the following conditional:

$$\text{Putnam's Model Theory} \implies \text{Semantic Indeterminacy.}$$

Now, as a general rule, conditionals of the form $\mathbf{P} \Rightarrow \neg\mathbf{Q}$ are evaluated by asking whether we can accept both **P** and **Q**. If we can, then the conditional has to be rejected.

In the case at hand, therefore, we ask whether we can accept both Putnam's model theory and the semantic determinacy of mathematical language. That is, we (tentatively and hypothetically) accept both of these things, and we then check to see whether the resulting combination leads to a contradiction. In this context, our hypothetical acceptance of semantic determinacy does not constitute an instance of "begging the question." It's just part of the standard method for evaluating conditionals.

Two comments are in order here. First, this "evaluation of conditionals" point does not depend on controversial assumptions about the status of the larger dialectic between Putnam and the realist. It does not presume that we have an adequate account of mathematical reference; nor does it rely on fancy considerations regarding the burdens of proof in realist/anti-realist debates. Indeed, it is compatible with the assumption that mathematical language *really is* indeterminate—i.e., that Putnam is *right* on the deep philosophical issue. The point simply shows that one aspect of Putnam's model-theoretic argument—namely, his charge of "question begging" in his defense of premise 2—rests on a misapplication of the logic of ordinary conditionals.

Second, this "evaluation-of-conditionals" point is *completely uncontroversial* when questions other than semantic determinacy are at issue. Suppose that two philosophers—John and Alvin—are arguing about the existence of God. John presents a version of the problem of evil, and Alvin responds by sketching a hypothetical situation in which God co-exists with all of John's evils. In this case, Alvin has not "begged the question" against John. Even though the overall debate is *about* the existence of God, it is perfectly legitimate for Alvin to evaluate John's argument by constructing this kind of hypothetical. Although the hypothetical can't show that God exists, it can show that John's *argument* fails.

So too, then, in the case of Putnam's model-theoretic argument. Even though the overall debate between Putnam and the realist concerns the determinacy of mathematical language, it is perfectly legitimate for the realist to assume such determinacy in the course of evaluating Putnam's argument. Just as Alvin's argument begs no questions against John, so the realist's argument begs no questions against Putnam. To think otherwise, is to misunderstand the logic of ordinary conditionals.

To sum up: Putnam's "just more theory" argument fails. There is a basic difference between *describing* the features which make a model "intended" and simply *adding* new sentences for a model to satisfy. Put otherwise, there is a difference between *changing* the semantics under which a set of axioms is interpreted and adding new sentences to be interpreted under the same *old* semantics. Further, Putnam's charge of "question begging" doesn't allow him to undercut—or to slur over—this distinction. At best, the charge rest on a misapplication of the logic of ordinary conditionals. At worst, it introduces assumptions which get Putnam into even more trouble than he was in before—e.g., the stability objection.

5 Some Connections

In this last section, I highlight a connection between the mathematical problems discussed in section 2 and the philosophical problems discussed in section 4. In particular, I show that both kinds of problems involve a certain dialectical assumption that Putnam tends to make, and I argue that realists have no reason to grant Putnam this assumption. In light of this, I conclude, once again, that there is little reason for realists to worry about Putnam's model-theoretic argument.

Let's begin with a passage originally discussed in section 3. Putnam is considering the claim that realists use mathematical language to *describe* the intended interpretation of set theory. He writes: "Here the [realist] is ignoring his own epistemological position. He is philosophizing as if naive realism were true of him... as if he and he alone were in an *absolute* relation to the world." Now, leave aside the question of whether there is anything *wrong* with such naive philosophizing—a question answered in the negative in the last section—and ask instead whether Putnam engages in any naive philosophizing of his own.

To see the worry here, recall the overall structure of Putnam's argument. Putnam begins by considering a theory which includes "all theoretical and operational constraints," a theory which includes *everything* we might ever need to say and (even) "all true sentences" (*MR* 18). Once this theory is in hand, Putnam steps back to examine this theory's semantics: he shows that the theory has many different (first-order) models, and he argues that these models all constitute "intended interpretations" of our language.

It is here, in this turn to semantics, that Putnam's own "naive realism" comes into play. Whatever problems our theory is supposed to have (e.g., semantic indeterminacy) these problems vanish when Putnam engages in semantics. Putnam allows himself to refer, naively and absolutely, both to a particular theory and to a whole menagerie of models for this theory. Further, for each of these models, Putnam spells out the relationship between model and theory in a thoroughly realistic manner. In short: Putnam philosophizes as though he and he alone were in an *absolute* relation to the world of models—what *he* calls "models" really are models, what *he* calls "satisfaction" really is satisfaction, and *of course* there is a fixed, somehow singled-out, correspondence between the language of model theory and the world of models in *his* case.

As we saw in section 3, however, this is precisely the standpoint which Putnam denies to the realist. When the realist tries to "stand back" from his set theory to talk about that theory's interpretation—to specify, for instance, that this interpretation must be transitive, or well-founded, or satisfy second-order ZFC—Putnam accuses him of "begging the question." Though Putnam's own model-theoretic talk should be viewed as talk *about* set theory, the realist's talk must be viewed as talk *within* set-theory.[17]

This asymmetry between Putnam's own position and the position he allows his opponents resembles closely the asymmetry discussed at the end of section 2.1. There, Putnam wanted to limit his opponents to a specific collection of set-theoretic axioms, while he himself used stronger axioms to prove his key theorem. Here, Putnam wants to limit his opponents to working *within* a particular theory, while he himself steps outside this theory to talk about its semantics. In both cases, then, Putnam's arguments depend on allowing himself just a little more than he allows those against whom he is arguing.

Clearly, however, there is no reason for realists to go along with all this. As long as realists insist on using the same resources that Putnam himself uses, neither of Putnam's key arguments can get off the ground. Putnam's argument for premise 1 will fail for the reasons discussed in section 2, and his argument for premise 2 will fail for the reasons discussed in section 4 (and in this section). With both of these arguments so disabled, Putnam's larger model-theoretic argument becomes completely unthreatening.

6 Conclusion

This paper has isolated two flaws in (one version of) Putnam's model-theoretic argument against realism. I have shown that one half of Putnam's argument—i.e., his argument for premise 1—rests on a mathematical mistake, and I have argued that Putnam cannot fix this mistake without jeopardizing the very philosophical point which his mathematics is supposed to support. On a more philosophical note, I have argued that the non-mathematical portion of Putnam's argument is just as problematic as the mathematical portion. Putnam's defense of premise 2 introduces inconsistencies into Putnam's overall position, renders his argument strikingly similar to some obviously faulty arguments, and, in the final analysis, depends on a straightforward misapplication of the logic of ordinary conditionals. What's more, it seems likely that *any* defense of Putnam's premise 2 will face similar problems—e.g., will run aground on the "stability objection" of section 4. In the end, therefore, I think that realists have very little to fear from Putnam and his models.

[17] Clearly, this point amounts to a less-technical reworking of some of the ideas behind the "stability objection" in section 4. Putnam needs to use notions like "finite" to pick out his own preferred "intended interpretations" of set theory; however, these are precisely the notions he wants to deny the realist. Hence, it is only by assuming a stark asymmetry between his own position and that of his opponents that Putnam can make his overall argument hold together. (A note: if Putnam *were* to allow realists to use the notion "finite," then they could define the notion "well-founded." This would be enough to rule out the models generated by Theorem 1.)

References

Anderson, David. "What is the Model-Theoretic Argument." *The Journal of Philosophy* 93 (1993): 311–22.

Devitt, Michael. *Realism & Truth*. Princeton: Princeton University Press, 1984.

Douven, Igor. "Putnam's Model-Theoretic Argument Reconstructed." *The Journal of Philosophy* 96 (1999): 479–90.

Lewis, David. "Putnam's Paradox." *Australasian Journal of Philosophy* 62 (1984): 221–236.

Putnam, Hilary. "Models and Reality." In *Realism and Reason*. Cambridge: Cambridge UP, 1983.

Van Cleve, James. "Semantic Supervenience and Referential Indeterminacy." *The Journal of Philosophy* 89 (1992): 344–361.