Phil Quarterly 58 (April 2008): 193–213

Two Arguments against Realism

Timothy Bays

Over the last 20 years or so, Hilary Putnam has developed a series of arguments which use basic theorems of model theory to undermine semantic realism. Here, I discuss two generalizations of these arguments. The first employs some new forms of model theory to generate substantially stronger conclusions than Putnam's original arguments generated. The second illustrates a method for replacing the model theory in Putnam's arguments with results from other branches of science—in particular, with results from astronomy.

Now, I should say at the outset that neither of these new arguments is supposed to be persuasive: each of them fails, and fails rather badly, when regarded as a serious objection to realism. Nevertheless, the arguments serve three purposes. First, the parallels between my new arguments and Putnam's original arguments help to highlight what's really going on in the latter, and the obvious flaws in my arguments help to isolate the corresponding flaws in Putnam's arguments. Second, these new arguments expose the inadequacy of several recent defenses of Putnam. (Very roughly, I argue that if these defenses saved Putnam's arguments, then they would save my arguments as well. But, my arguments are unsalvageable. So, the defenses don't save Putnam.) Finally, the arguments present a new challenge to Putnam and his defenders: to provide a formulation of the model-theoretic argument which makes *that* argument seem compelling without doing the same for mine.

1 Putnam's Arguments

Let's begin with the sources. Over the years, Putnam has provided several versions of the model-theoretic argument. They differ both with respect to their targets and with respect to the model-theoretic tools they employ. On the "targets" side, some of Putnam's arguments focus exclusively on the semantics of mathematical English—of talk about numbers, functions, sets, etc.—while others focus on more ordinary talk about, e.g., cats and cherries, flies and spiders, tables and coffee cups. On the "tools" side, some of Putnam's arguments employ the Löwenheim-Skolem theorems, while others rely on results concerning the permutation of a model's domain. Despite all this variety, however, there is a common *structure* which all of Putnam's arguments share, and it's this common structure which concerns us here.[1]

[1]Versions of the model-theoretic argument can be found in [14], [15], and [16]. The argument in [16] focuses on mathematical language and makes use of the Löwenheim-Skolem theorem (along with some other sophisticated results). The arguments in [14] and [15] focus on ordinary talk about everyday objects and appeal mostly to permutation theorems. For more recent discussion of these arguments, see Putnam's comments in [18] and [19].

Suppose we have a first-order theory which is rich enough to describe some reasonable portion of the world. For our purposes, it doesn't really matter what kind of theory this is. It could be a mathematical theory like set theory, or a scientific theory like quantum mechanics, or even a theory concerning day-to-day things like automobile engines. Whatever the details, Putnam argues that basic theorems of model theory entail that this theory has many different models. As a result, the *mere formal stucture* of the theory doesn't pin down a unique interpretation for the theory's language—i.e., it doesn't fix a unique model for that language. To put this point in Putnam's terms, "theoretical constraints" cannot force a language to take on a unique "intended interpretation."[2]

Nor will "operational constraints" help here. No matter how many measurements we might take—indeed, even if we take a hypothetical *infinity* of measurements—these measurments can all be coded up as mere sentences of our theory: "the temperature of XYZ at time t_1 is n_1," "the temperature of XYZ at time t_2 is n_2," etc. So, taking empirical measurements into account won't solve the problems of the last paragraph: Putnam can still find multiple models which satisfy our (observationally) expanded collection of sentences.

[2] A few details are probably in order here. In all of his arguments, Putnam begins by assuming that the theory in question accurately describes the world. He then proceeds to treat the world as a *model* for the theory. That is, he fixes a set-theoretic interpretation function which "connects" our language to the world (and the power-set of the world, and the power-set of the two-fold cross-product of the world, etc.) so that this function mimics the ordinary relations of reference and predication. He then uses the first-order satisfaction relation to define the notion of truth. (I should note, here, that the idea that we can treat the world as an ordinary model is non-trivial, and I doubt that it's ultimately defensible. For the purposes of this paper, however, I will grant Putnam this move. For more on the issue, see section two of [2].)

At this point, Putnam's argument begins to bifurcate. In one version of the argument—the so-called "permutation argument"—Putnam appeals to two theorems of elementary model theory. First, if two models $\mathbb{M}$ and $\mathbb{M}'$ are *isomorphic*—i.e., if there exists a structure-preserving bijection $f : \mathbb{M} \to \mathbb{M}'$—then these two models are also *elementarily equivalent*—i.e., for every sentence ϕ, $\mathbb{M} \models \phi \iff \mathbb{M}' \models \phi$. Second, if $\mathbb{M}$ is a model and $f : \mathbb{M} \to A$ is a bijection, then f carries with it a canonical method for building a new model which has A as its domain and which is isomorphic to $\mathbb{M}$ (with f itself serving as the relevant isomorphism). Combining these two results, we note that any *permutation* of a model's domain brings with it a canonical method for creating a new model which 1.) has the same domain as the original model, 2.) satisfies exactly the same sentences as the original model, but 3.) has a radically different intepretation function from the original model.

Suppose, for instance, that we consider a permutation of the world which switches Spot the dog with Fluffy the cat. The model induced by this permutation has an interpretation function which uses "Spot" to name Fluffy and "Fluffy" to name Spot; it also thinks that Spot falls under the predicate "Cat" and Fluffy under the predicate "Dog". Hence, the mere fact that our two models satisfy the same *sentences* (and share the same domain) doesn't ensure that they reflect the same *semantics*—i.e., that their interpretation functions mimic the same reference and predication relations. This, then, is Putnam's first way of capturing the idea that "theoretical constraints" don't force a theory to take a unique "intended interpretation."

The second version of Putnam's argument eschews these permutation results in favor of the Löwenheim-Skolem theorems. Suppose that our original theory includes a fair bit of set theory. Then the Löwenheim-Skolem theorems ensure that the theory has models which "get it wrong" concerning basic issues of cardinality—e.g., models which satisfy formulas like "x is finite" or "x is uncountable" even when x is really infinite or x is really countable. As a result, the mere fact that these models satisfy the standard *axioms* for set theory doesn't ensure that they reflect the standard *semantics* for set-theoretic language—i.e., the semantics in which "$\exists$" ranges over the universe of sets and "$\in$" picks out the real membership relation. Once again, therefore, the "theoretical constraints" imposed by our set-theoretic axioms prove insufficient to pick out a unique "intended interpretation" for the language of formal set theory.

Hence, just as theoretical contraints are unable to fix the intended interpretation of our language, so also are theoretical *and* operational constraints unable to do so.[3]

This gives us an initial premise in Putnam's argument. Using "theoretical and operational constraints" in the sense just indicated, we have:

1. Theoretical and operational constraints do not fix a unique "intended interpretation" of our language.

Further, Putnam thinks that theoretical and operational constraints are the only things which *could* fix the intended interpretation of our language. His reasons for thinking this are complicated and lie at the heart of some controversial interpretive questions; I will return to them in sections 2 and 3. For now, I simply take this claim as given and continue with the central argument:

2. Nothing other than theoretical and operational constraints could fix the "intended interpretation" of our language.

So, 3. There isn't a unique "intended interpretation" of our language.

Here, 1–3 constitute the *core* of Putnam's model-theoretic argument. They show (purportedly) that our language has many different "intended interpretations" and that these interpretations are fixed by—and only by—the sum total of our "theoretical and operational constraints."[4]

Let me note two things about this core argument. First, once the core argument is in place, Putnam can (and does) go on to argue for more detailed conclusions. He argues that realism is false (because realists are committed to the claim that there *is* a unique intended interpretation for our language).[5] He argues that *ideal* theories—i.e., theories which are consistent, simple, elegant, etc.—must also be *true* theories (because a theory is *true* if it's true on its intended interpretation, the thing which makes an interpretation intended is the satisfaction of "theoretical and operational constraints," and every ideal theory has *some* interpretation which satisfies the relevant theoretical and operational constraints).[6] Finally, he argues that realists are committed to the rejection of naturalism (because the only way to evade premise 2 in the core argument involves appealing to mysterious, non-natural properties of the mind and/or world).[7]

Second, these more detailed arguments should be viewed as *extensions* of the core argument. The core argument—or something very close to it—underlies these more detailed developments of Putnam's thought, and the machinery used in the core argument is essential to the success of these other arguments. If the

[3]In [16], Putnam uses a slightly different technique to deal with operational constraints. For our purposes, however, these differences in technique are inessential and can safely be ignored.

[4]It's important to emphasize here that almost everyone—including Putnam's defenders—agrees that something like 1–3 lies at the heart of Putnam's argument. See, e.g., [1] and [5] for defenses of Putnam which accept essentially this formulation. Nor is there anything here to which the authors of [9] or [10] should object (though this isn't precisely the formulation they use).

[5]See [14], [15], [16], and [18].

[6]See [14] and [15].

[7]See [16] and [18].

core argument fails, therefore, and if Putnam's model-theoretic machinery can't do what he thinks it does, then Putnam's more detailed arguments will fail as well. In the next two sections, I will show that the core argument does, indeed, fail.

2 The Supermodel Argument

To motivate our first extension of Putnam's argument, it's useful to consider an old—and, by now, a relatively notorious—defense of premise 2 in the core argument. This defense, which is typically called the "just more theory" defense, rests on a single key idea: that the phrase "theoretical constraints" is broad enough to encompass *philosophical semantics* as well as mathematics and natural science. In particular, the defense insists that any philosophical account of how our language gets its intended interpretation should itself be viewed as a new theoretical constraint. So viewed, no such account will enable us to evade the argument for 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 cannot give our language a unique intended interpretation.

In effect, then, the just-more-theory defense claims that any mechanism which *seems* to fix the intended interpretation of our language turns out, upon reflection, to be a *special case* of the "theoretical and operational constraints" mentioned in premise 1. As a result, any purported counterexample to premise 2 has already been dealt with by the argument for premise 1. So, by simply adopting a particularly flexible—and a somewhat colonistic—reading of the phrase "theoretical constraints," Putnam ensures that no rival mechanism for fixing intended interpretations needs to be taken seriously.

Now, before examining what's wrong with this just-more-theory defense, I want to make two preliminary comments about it. First, it's worth emphasizing just how *often* Putnam has employed this defense.[8] We have, in effect, already seen it used to deal with "operational constraints" (which got interpreted, not as concrete observations of specific physical phenomena, but instead as mere *sentences* to be added to our overall theory). Similarly, Putnam has used the defense to defuse the suggestion that causal constraints help to pin

[8] This point goes to an interpretive dispute in the literature. Traditionally, most commentators have agreed that some version of the just-more-theory defense plays a central role in the model-theoretic argument (see [4] (chapter 11), [8], [11], and [22]). Recently, however, several commentators have challenged this interpretation (see [1]; see also [5] and [9]). They argue that Putnam's "just more theory" talk is merely supposed to highlight the theoretical inadequacy of several particular theories of reference (in particular, several versions of the causal theory of reference); it's not supposed to provide a *general* argument of the type sketched above.

As a rule, I find this revisionary line of interpretation unpersuasive. As indicated in the main text, Putnam has applied the just-more-theory defense quite widely, and he's applied it in ways which follow exactly the patern sketched above. So, while I certainly agree that Putnam has given *other* arguments against realistic theories of reference—and *many* other arguments in the case of the causal theory—and while I acknowledge that Putnam has recently started to back away from the just-more-theory defense (see, e.g., [18]), I think it's abundantly clear that this defense played a key role in Putnam's original formulation of the model-theoretic argument. I will say a bit more about some of the revisionary interpretations when we get to section 3.

down the intended interpretation of our language (arguing that the entire causal theory of reference should be regarded as "just more theory" and appended to our set of theoretical constraints).[9] Finally, Putnam has used the defense to deal with some technical objections involving modal and higher-order logics.[10]

Second, we should note that most realists have strongly resisted this just-more-theory defense. Their basic objection can be formulated in model-theoretic terms.[11] We begin by drawing a distinction between specifying the type of model theory to be used in interpreting our current theoretical constraints and simply adding new theoretical constraints to be interpreted using whatever type of model theory Putnam himself happens to prefer. Realists often want to do the former: e.g., when they specify that we should use a modal or second-order satisfaction relation, or when they suggest that we should limit ourselves to models whose interpretation functions respect certain kinds of causal constraints. Putnam's just-more-theory defense, however, systematically reinterprets them as doing the latter—as simply adding new sentences to be interpreted using *Putnam's* favorite model theory. On the realists' view, this reinterpretation constitutes an illicit—and a somewhat perverse—misconstrual of their overall position.

To evaluate the realists' objection here—and to show more generally why the kind of reinterpretation at issue in the just-more-theory defense is *way* too powerful to be plausible—I turn to my first new argument. Following Putnam, assume that we have expressed our "theoretical and operational constraints" as a set of first-order sentences. Let $\mathbb{G}$ be a model whose domain contains only one thing—my cat Gandalf—and which interprets all relations as maximal (i.e., which makes every n-ary relation, $R(x_1, \ldots, x_n)$, true at all n-tuples from $\mathbb{G}$). Finally, let $\models_g$ be a "satisfaction" relation which agrees with the ordinary first-order satisfaction relation, *except that* it interprets negation as a redundant operator. That is, let the recursion clause for "$\neg$" in the definition of $\models_g$ read as follows: for any model $\mathbb{N}$, any assignment of variables ν, and any first-order

[9] See [16], pp. 17–18 and [17] pp. vii–xii.

[10] In [16], Putnam uses the Löwenheim-Skolem theorems to generate "intended interpretations" for a collection of theoretical constraints which includes both set theory and *all of natural science.* Modal considerations come into play when we consider scientific theories which involve counterfactual talk—e.g., claims about what *would happen* if we made certain measurements. For such theories, it might seem that Putnam should use *modal models* to generate his "intended interpretations"; if so, then the Löwenheim-Skolem theorems might not apply. Higher-order logic comes into play when we consider second-order formulations of set theory (to which the Löwenheim-Skolem theorems certainly don't apply).

Putnam deals with these concerns on pp. 8–9 and 23 of [16]. In the modal case, he suggests that we "first-orderize" our counterfactual talk—i.e., that we reformulate such talk in terms of a new "subjunctively necessitates" predicate and then add (first-order) axioms governing this predicate to our overall theory. In the second-order case, Putnam uses Henkin models to the same effect. (Essentially, Henkin models provide a way of "first-orderizing" second-order theories: they treat second-order theories as if they were simply first-order theories with extra axioms governing the relevant second-order constructions.)

In both cases, therefore, Putnam employs a version of the just-more-theory strategy. The realist wants his theory to be interpreted using a certain *kind* of semantics—one which will fix his theory's interpretation more tightly than Putnam's model-theoretic semantics would. Putnam responds by transforming *descriptions of* the realist's semantics into collections of first-order sentences, and then insists that we interpret these sentences using his own preferred, first-order semantics.

[11] This isn't the way the objection is usually formulated in the literature, but I think it's the most perspicuous formulation for the purposes of this paper. For examples of this objection, see [11], 221–236, [22], 344–361, and chapter 11 of [4].

formula ϕ,

$$\mathbb{N}, \nu \models_g \neg\phi \iff_{\text{Def}} \mathbb{N}, \nu \models_g \phi.$$

Given these definitions, it's straightforward to show that $\mathbb{G} \models_g \phi$ for *any* ϕ in our language.[12] So, if we understand "satisfaction" after the manner of $\models_g$, then it turns out that $\mathbb{G}$ satisfies *absolutely everything.*

Now, since $\mathbb{G}$ satisfies absolutely everything, it clearly satisfies all of our theoretical and operational constraints. This lets us run the following argument (which I will dub the "supermodel argument"):

1′. Theoretical and operational constraints do not commit us to the existence of more than one object (my cat Gandalf).

2′. Nothing other than theoretical and operational constraints could commit us to the existence of more than one object.

So, 3′. Nothing commits us to the existence of more than one object.

Here, a simple modification of our understanding of negation allows us to generate surprisingly strong results. From a metaphysical perspective, the argument shows that we have no ontological commitments to objects other than my cat. From a semantic perspective, the argument shows that nothing about our use of language prohibits *all singular terms* from referring to Gandalf and *all predicates* from applying to Gandalf (e.g., the name "Spot" and the predicate "is a Dog").

Clearly, something has gone wrong here. I suggest, however, that if we accept Putnam's just-more-theory defense, then it's hard to see *what* has gone wrong. There are two things to note. First, the just-more-theory defense can be used to defend the supermodel argument against exactly the same criticisms as were raised against Putnam's argument. Consider any mechanism which purports to fix the intended interpretation of our language, and, in particular, which purports to *rule out* the interpretation given by $\mathbb{G}$ and $\models_g$. Following Putnam, the defender of the supermodel argument can simply insist that a description of this mechanism be added to our overall collection of theoretical constraints. Since $\mathbb{G}$ satisfies everything, $\mathbb{G}$ continues to satisfy these expanded constraints. Hence, the mechanism at issue doesn't *really* rule out $\mathbb{G}$ after all.

Second, the just-more-theory defense can be used to defend the supermodel argument against some more-specific criticisms that might be directed against it. Someone might, for instance, complain that the supermodel can't give the intended interpretation of our language because it contains only one thing (while it is clear that we intend to talk about more than one thing!). But, while it's certainly true that the supermodel

[12] The argument for this claim is a straightforward induction on the structure of ϕ. The fact that $\mathbb{G}$ contains only one element, along with the fact that all relations are maximal, ensures that $\mathbb{G}$ "satisfies" all atomic formulas. Given this, the passage through binary connectives is trivial (since standard binary connectives map (T, T) to T). Similarly, the fact that negation is redundant allows us to move from $\mathbb{G} \models_g \phi$ to $\mathbb{G} \models_g \neg\phi$. Finally, because $\mathbb{G}$ contains only one object, $\mathbb{G} \models_g \exists x\, \phi(x) \iff \mathbb{G} \models_g \forall x\, \phi(x) \iff \mathbb{G} \models_g \phi[\text{Gandalf}]$; so, passage through quantifiers is simple.

It's worth noting that this construction doesn't really depend on the fact that we're working with a first-order language. Trivial modifications will let us prove the result for modal or higher-order languages. Nor must we apply Putnam's strategy of "first-orderizing" such languages in order to obtain this preservation result (although it does make things somewhat easier).

contains only one thing, it's also true that the supermodel "satisfies" the formal sentence which expresses the condition "contains more than one thing." That is,

$$\mathbb{G} \models_g \exists x\, \exists y\, (x \neq y).$$

Similarly, someone might complain that $\models_g$ violates the principles of bivalence and excluded middle. Again, though, the supermodel does satisfy the formal versions of these principles. For any ϕ,

$$\mathbb{G} \models_g \neg(\phi \wedge \neg\phi) \;\; \textit{and} \;\; \mathbb{G} \models_g \phi \vee \neg\phi.$$

Hence, as long as we follow Putnam and view conditions on interpretations—i.e., specifications or partial specifications of the model theory under which certain sentences are to be interpreted—as mere "theoretical constraints" to be interpreted using our own favorite model theory, we can save the supermodel argument from the most natural criticisms that could be leveled against it.

On the surface, this all suggests that something has gone badly wrong with Putnam's just-more-theory defense. If the defense provides adequate support for premise 2 in Putnam's model-theoretic argument, then, by parity of reasoning, it should also provide adequate support for premise $2'$ in the supermodel argument. But, the supermodel argument is *clearly* unsound, and the just-more-theory strategy *doesn't* provide adequate support for premise $2'$. So, it doesn't provide adequate support for premise 2 either.

Now, as it stands, this parity-of-reasoning argument may seem somewhat flippant. So, I'd like to slow down a bit and examine the serious philosophical point which the supermodel argument is supposed to bring out. I'll start with five general observations. First, at the most basic level, Putnam's model-theoretic argument turns on the fact that certain sentences—or sets of sentences—don't pick out unique models for themselves. Second, it only makes sense to talk about sentences "picking out" models against the backdrop of a fixed background semantics—i.e., a fixed conception of what counts as a model and a fixed "satisfaction" relation that ties sentences to models. Third, there are many different kinds of semantics which could, in principle, play this background role: propositional semantics, first-order semantics, second-order semantics, semantics with built-in causal constraints, etc. Fourth, Putnam's model-theoretic argument depends on making some fairly specific choices about these background semantics (in general, only standard, first-order model theory will do the trick for Putnam).[13] Finally, realists tend to prefer stronger background semantics than those favored by Putnam—i.e., semantics which connect languages to models in ways which preserve

[13]Putnam's emphasis on first-order semantics can be be seen most explicitly in his repeated insistence that theoretical constraints be formalized in first-order languages. It can be seen implicitly in his use of model-theoretic results which only apply to first-order theories—e.g., the Löwenheim-Skolem theorem—and in his repeated attempts to reduce strong semantics to first-order semantics—e.g., in the modal and higher-order cases discussed in footnote 10.

In fairness, I should note that some versions of Putnam's argument—the permutation arguments, in particular—work for a broader class of background semantics, including higher-order semantics. Still, they're pretty restrictive: they don't work for semantics with built-in causal constraints, and they probably don't work for modal semantics either. See [4] (chapter 11) for a discussion of the causal case and [12] (section 4) for an interesting analysis of the modal case.

more of the ordinary English significance of those languages than the first-order model-theoretic semantics that Putnam tends to favor.

Given these observations, we can view Putnam's just-more-theory defense as an attempt to close the gap between the kinds of strong background semantics preferred by realists and the substantially weaker background semantics needed for the model-theoretic argument. More specifically, the defense tries to eliminate this gap by *reducing* the realists' strong semantics to the first-order semantics needed for Putnam's model theory. The supermodel argument, in turn, shows that there is nothing *special* about our ability to perform this particular reduction. Just as Putnam can reduce the realist's strong semantics to his own first-order semantics, so also can we reduce first-order semantics—or, for that matter, even stronger semantics—to the supermodel's utterly trivial semantics.[14] So, unless Putnam thinks that *all* reductions of this kind provide legitimate tools for interpreting other people's theoretical constraints, he needs to explain the special status of his own favored reductions—i.e., he needs to explain why his own first-order reductions are theoretically significant while things like the supermodel reduction are not.

Let me be clear about the point that I'm making here. I'm not claiming that anyone who accepts the model-theoretic argument must also accept the supermodel argument. Clearly, the supermodel argument turns on treating as problematic a conception of truth in a structure which Putnam's core argument takes for granted (ordinary first-order satisfaction). So there's room in principle for accepting Putnam's argument while rejecting the supermodel argument. That being said, Putnam's own argument turns on treating as problematic conceptions of truth and reference which realists want to take for granted (second-order satisfaction, modal satisfaction, causal theories of reference, etc.). My point, then, is simply this: if Putnam's only justification for treating the realists' semantics as problematic comes from the mere fact that he *can* reduce these semantics to ordinary first-order semantics—in the manner of the just-more-theory defense—then *this* justification generalizes all the way to the supermodel case. After all, we *can* reduce first-order semantics to supermodel semantics.[15]

This, then, is the serious point which the supermodel argument is supposed to bring out. What can Putnam say in response? It seems to me that there are two lines he could take. On the one hand, Putnam could simply bite the bullet and accept the implications of the supermodel argument. That is, he could accept that all reductions—including various trivial reductions—provide philosophically legitimate interpretations of our theoretical constraints. Note that on this line, things like the supermodel argument may actually wind up supporting Putnam's overall position. Putnam's model theory was supposed to show that our theoretical and operational constraints have many different intended models; the supermodel argument shows that they have *even more* intended models. Hence, it shows that the situation for realists is *even worse* than Putnam's

[14] Of course, it's not just supermodel semantics which work here. We could make a similar point using propositional semantics, or we could invoke the purely stipulational semantics given by "$\mathbb{M}$-Γ-satisfaction" (where $\mathbb{N} \models_{\mathbb{M},\Gamma} \phi \iff_{\text{Def}} \mathbb{N} = \mathbb{M}$ and $\phi \in \Gamma$).

[15] As a textual matter, I should note that Putnam himself never gives a more-detailed justification for treating the realists' semantics as problematic. He simply shows that he can reduce their semantics to first-order semantics, and then gets on with his business. This, I think, is what makes the just-more-theory defense susceptible to challenges like the supermodel argument.

original arguments might have indicated.

Unfortunately, this line has two major problems. First, it makes Putnam's overall position quite implausible. Whatever worries Putnam's model-theoretic argument may have engendered, it's hard to see the supermodel argument—or similar arguments based on propositional semantics or $\models_{\mathbb{M},\Gamma}$-semantics—engendering the same kinds of worries. If Putnam's just-more-theory defense requires us to take these trivial arguments seriously—if, that is, it really requires us to accept that *any* reinterpretation of our semantics gives rise to an intended model of our theoretical constraints—then this undercuts the overall plausibility of Putnam's argument. Indeed, I myself would view this conclusion as a straightforward *reductio* of the model-theoretic argument (or, at least, of the just-more-theory defense).

Second, this line essentially abandons the *model-theoretic* character of Putnam's core argument. Much of the initial attraction of that argument—much, if you will, of its "philosophical sex-appeal"—stems from its claim that basic theorems of model theory show that semantic realism is untenable. But, when the argument is filled out in the way we have been discussing, then all of the serious model-theory—the permutation theorems, the Löwenheim-Skolem theorem, the Shoenfield absoluteness theorem, etc.—becomes superfluous. The argument now rests on the more-or-less trivial observation that *if you get to reinterpret anything you want, any way you want, then you can make any sentences you want true under any circumstances you want.* We didn't need fancy model theory to tell us that.

In the end, then, I don't think that Putnam can plausibly respond to the supermodel argument by simply accepting that argument's implications (treating it, in effect, as a friendly amendment to his own model-theoretic argument). Instead, I think that Putnam needs to give an explicit argument for the special status of first-order model theory. That is, he needs to show that there's something special about first-order model theory which makes it theoretically significant that various theories can be recast in first-order terms, but there's no equivalent "something" which would make supermodel semantics theoretically significant. This would explain why it's philosophically legitimate for Putnam to reduce the realist's strong semantics to his own first-order semantics, but illegitimate for me to continue this reduction all the way to supermodel semantics (or propositional semantics, or $\models_{\mathbb{M},\Gamma}$-semantics, or ...).

In practice, this second line requires Putnam to show that first-order model theory is, in some fairly strong sense, semantically normative. Among all the possible background semantics which we could use for interpreting our language, first-order model theory is—and is uniquely—the *right* semantics for us to use. Let me make three points about this position. First, it's important to emphasize just how strong this position really is. To avoid the supermodel argument, Putnam needs to assume that first-order logical constants like $\neg$, $\rightarrow$ and (perhaps) $\exists$ have fixed interpretations which are given by the first-order satisfaction relation. Hence, anything which can be defined in terms of these constants will also have a fixed (though derivative) interpretation. But that's *all* that gets a fixed interpretation. Things which can't be defined in purely logical terms—i.e., all of the non-logical predicates and relations in our language—are simply indeterminate; they just have one interpretation in some intended models, and another interpretation in others.

In effect, then, this line commits Putnam to a fairly strong form of logicism—not just logicism about mathematics, but logicism about zoology, logicism about astronomy, logicism about auto mechanics, etc. Clearly, this isn't a very popular position in the literature; nor, for reasons that we'll discuss below, is it a position that I myself see much hope of seriously defending. Still, I think it's the position that Putnam *has to* defend if he wants his larger model-theoretic argument to have any real philosophical bite. For convenience, I'll dub this position "global logicism."[16]

This brings me to a second point. As far as I know, Putnam has never actually tried to defend the kind of global logicism that his just-more-theory defense now seems to depend on. As I noted earlier (p. 8, n. 15), Putnam's formulations of the just-more-theory defense don't go much beyond the simple observation that he *can* reduce various kinds of strong semantics to first-order, model-theoretic semantics. So, there's nothing in Putnam's own writing which would explain how he intends to defend the rather radical position on which the just-more-theory defense now seems to rest (nor, indeed, is there anything which indicates that he even recognizes the *need* for such a defense).

Finally, I think it's highly unlikely that this particular position *could be* given an adequate defense. On the surface, global logicism looks like a pretty loopy philosophical position. Further, and as I've argued elsewhere, it's a position that's subject to some deep internal tensions (since the set-theoretic machinery that's needed to define notions like *model* and *satisfaction* can't itself be specified in purely logical terms).[17] Hence, I'm inclined to think that there just *isn't* any good reason for accepting the kind of global logicism that's needed to make Putnam's just-more-theory defense cogent.

[16]Three comments are in order here. First, it's sometimes suggested that Putnam's model-theoretic argument tells against a view called "global descriptivism," the view that the intended interpretation of our language is simply that interpretation which best "fits" with our overall patterns of linguistic usage. I think that this is somewhat misleading. Global descriptivism is really a family of views which differ in the ways they flesh out the notion of an interpretation "fitting" our usage. And these differences matter here. If "fit" can be specified any way we want, then things like the supermodel will count as intended interpretations of our language, and global descriptivism will look pretty silly. If "fit" is specified more tightly—say, by building Lewis-style eligibility constraints and/or some version of the casual theory of reference into our specification—then the model-theoretic argument doesn't get any traction. The only way the model-theoretic argument can play a non-trivial role in countering global descriptivism, therefore, is if our notion of "fit" is itself specified in model-theoretic terms. In short: the only version of global descriptivism that Putnam's model theory really tells against is the version I'm calling "global logicism."

Second, global logicism goes well beyond the simple Quinean demand that we use first-order quantifiers to make our ontological commitments explicit. After all, the supermodel works just fine with first-order languages, and it even interprets first-order quantifiers in a standard fashion. More importantly, Quine's strictures are supposed to be compatible with the notion of a "fully interpreted first-order language"—i.e., a language which has a first-order quantificational structure but in which the other (non-logical) terms, predicates and relations also get fixed interpretations. In contrast, global logicism only fixes the interpretations of those terms, predicates, and relations which can be defined using purely logical machinery.

Finally, I should note that even those versions of Putnam's argument which apply to non-first-order languages—e.g., the permutation arguments—still retain this logicist character (though they allow a richer variety of background logical machinery). In general, it's the *model-theoretic* nature of Putnam's argument which forces him to accept some form of logicism.

[17]At least, not if we understand "purely logical" in terms of first-order model theory. See section 4 of [2] for a more-detailed discussion of this kind of problem.

To sum up, then, Putnam's just-more-theory defense requires him to walk a very fine line between the kinds of strong semantics which realists would like to use to rebut his argument and the kinds of weak semantics which threaten to trivialize his argument. Putnam himself has never explained why this particular line is philosophically significant; nor do I see any good prospects for developing such an explanation. I'll end this section, therefore, by simply issuing a challenge to Putnam and his defenders:

> Provide a version of the model-theoretic argument which makes *that* argument seem compelling without, at the same time, making the supermodel argument equally compelling.

3 The Astronomical Argument.

In this section, I turn to a second new argument. To motivate this argument, it's once again useful to look at a relatively simple defense of premise 2 in the core argument. This defense—which I will call the "no-explanation" defense—rests on two key observations. First, realists need a plausible account of how reference is supposed to work. If we maintain, for instance, that the truth of a sentence like "the cat is on the mat" *depends on* the fact that a particular cat is on a particular mat, then we seem to need an account of how "cat" relates to our cat and "mat" to our mat.

Second, realists have yet to provide a plausible account of reference. Although such an account is necessary for realism, and although this necessity has been obvious for some time, no attractive candidates—at least by Putnam's lights—have yet been put forward. Until such candidates are forthcoming, therefore, we should give provisional support to premise 2. In short: realists' obvious need for an explanation of reference, combined with their continuing failure to provide such an explanation, should lead us to the conclusion that no such explanation is ultimately possible.[18]

Now, before I examine what's wrong with this defense, it's worth saying something about its history. As far as I can tell, this defense of premise 2 wasn't used in Putnam's original presentations of the model-theoretic argument (i.e., those in [14], [15], and [16]). It is, however, the defense which he's adopted more recently (see [19], [20], and especially [18]). Further, it's a defense which Putnam's defenders have often used: Haukioja does so in [9] (responding to Chambers in [3]), and Anderson essentially does so in [1].[19] So,

[18] For Putnam, plausible explanations of reference must be (at least nominally) compatible with naturalism. Appeals to divine agency, noetic rays or Aristotelian essences are unacceptable. In [16] and [17], Putnam criticizes Chisholm and Lewis for relying on non-natural reference-fixers; in [19] and [20], he criticizes Boyd and Devitt on similar grounds. See [1] for further discussion of this aspect of Putnam's argument.

[19] Anderson's argument is subtle, and it probably deserves a bit more explanation. Anderson starts with the assumption that causal theories of reference are the only live candidates for modern-day realists. He then presents a series of arguments—his so-called "causality trilemma"—which purports to show that causation cannot, in fact, serve as a reference-fixer in the sense relevant to Putnam's argument. Hence, Anderson concludes, realists lack an adequate theory of reference.

Let me make three comments on this argument. First, as far as I can tell, Anderson's assumption that causation provides the only possible reference-fixer for modern-day realists is simply based on the state of the current literature—i.e., on the fact that no other proposals are seriously on the table. That's why I view Anderson's argument as a species of the no-explanation defense. Second, I should emphasize that Anderson himself views his argument as a mere *interpretation* of some of the "just-even if this defense didn't play a role in Putnam's *original* formulation of the model-theoretic argument, it has, over time, *become* an integral part of that argument.[20]

To see what's wrong with the no-explanation defense, let's turn to our second new argument. This argument—which I call "the astronomical argument"—starts with a basic result of modern astronomy: *there is no intelligent life on Mars.* Given this, we can easily generate the following:

1″. Little green Martians do not fix a unique "intended interpretation" of our language.

2″. Nothing other than Martians could fix the "intended interpretation" of our language.

So, 3″. There isn't a unique "intended interpretation" of our language.

There are three things to notice about this argument. First, and most obviously, the astronomical argument looks strikingly similar to Putnam's own core argument. The two arguments have the same conclusion, they have (essentially) the same logical structure, and their initial premises are both true.[21] The only significant difference is that the astronomical argument eliminates Putnam's model theory and replaces it with some more accessible results from astronomy.

Second, this replacement doesn't have any effect on the no-explanation part of Putnam's argument. After all, the no-explanation defense says nothing specific about the "theoretical and operational" constraints mentioned in Putnam's premise 2. It simply observes that presently-available accounts of reference-fixing don't stand up to philosophical scrutiny, and then infers that future accounts will also prove inadequate. Given this, replacing talk of "theoretical and operational constraints" with talk of "little green Martians" should have *no effect whatsoever* on the cogency of the no-explanation defense. Indeed, the no-explanation defense is *so* general that it should work with any version of premise 1—with that from the astronomical

more-theory" passages in Putnam. (Anderson himself is a realist, and he ultimately rejects Putnam's conclusions.) Finally, I should note that, for the reasons sketched back on page 4 (see especially fn. 8), I find Anderson's *interpretation* of Putnam unconvincing. Hence, I'm going to treat Anderson's argument as though it were an independent defense of Putnam's argument, rather than an explication of Putnam's own texts.

[20]Although I don't want to over-emphasize this point, I should note that Putnam's recent adoption of the no-explanation defense creates an awkward mismatch between the *rhetoric* of his original formulations of the model-theoretic argument and the conclusions his (modified) argument can actually support. Originally, the argument was supposed to support claims like the following:

> The idea that it is something *other* than operational and theoretical constraints that singles out the right reference relation . . . is an *incoherent* idea. ([18], 215)
>
> The supposition that even an 'ideal' theory (from a pragmatic point of view) might *really* be false appears to collapse into *unintelligibility.* ([14], 126)
>
> The 'Löwenheim-Skolem paradox' *is* an antinomy, or something close to it, in *philosophy of language.* ([16], 1)

Clearly, however, the no-explanation defense can't justify this talk of "incoherence," "unintelligibility" and "antinomy." At best, it justifies talk of "puzzles yet to be solved" and "phenomena yet to be explained." Hence, whatever independent merits the no-explanation defense may have, the defense introduces tensions into Putnam's overall rhetoric.

[21]Actually, there may be grounds for doubting the truth of the first premise in *Putnam's* argument. For convenience, however, I'll grant Putnam this premise. In any case, premise 1″ is clearly true.

argument, that from the zoological argument, that from the quantum-mechanical (or even auto-mechanical) argument, etc.

Finally, given points one and two, the astronomical argument creates another parity-of-reasoning problem for Putnam. If the no-explanation defense provides adequate support for premise 2 in Putnam's argument, then it should also provide adequate support for premise $2''$ in the astronomical argument. But the astronomical argument *isn't* a genuine challenge to realism, and the no-explanation defense doesn't make it into one. So, it shouldn't make Putnam's argument into a genuine challenge either.

Now, just as before, the somewhat loopy nature of the astronomical argument makes it important to slow down and examine the serious philosophical point which this argument is supposed to bring out. It seems to me that there are two different ways to understand the astronomical argument. First, we could view the argument as a straightforward *reductio* of Putnam's position. The astronomical argument isn't a serious challenge to realism; so, the no-explanation defense doesn't work; so, premise 2 in Putnam's argument isn't well-supported; etc.; etc. This is the line we'd be inclined to take if we didn't find the no-explanation defense very persuasive in the first place.

Second, and more importantly, we can view the astronomical argument as a tool for highlighting the role that model theory plays—or, more accurately, *doesn't* play—in Putnam's overall argument. The reason the no-explanation defense works so nicely with the astronomical argument—and its zoological, quantum-mechanical and auto-mechanical cousins—is that the defense says nothing *specific* about the kinds of reference-fixing mechanisms it's trying to rule out.[22] Instead, it provides an entirely *general* argument against realist theories of reference. In effect, therefore, the no-explanation defense really amounts to a *direct* argument for the *conclusion* of the astronomical argument. To the extent that it supports the astronomical argument at all, therefore, it does so only by rendering the astronomy in that argument irrelevant (i.e., by providing us with an independent and non-astronomical argument for $3''$ which allows us to bypass all the astronomy used in arguing for $1''$).[23]

Clearly, this point carries over to the model-theoretic case as well. Once we see that the no-explanation defense provides an independent argument against realism—i.e., an independent argument for 3 and $3''$—then Putnam's model theory becomes just as superfluous as the astronomical argument's astronomy. Even if the no-explanation defense makes Putnam's overall argument sound, it does so only by reducing Putnam's model-theory to mere technical window dressing. Further, and as I noted in the last section (see p. 9), this is a genuine problem for Putnam. The philosophical appeal of Putnam's argument stems almost entirely from

[22]That is, there's nothing in the basic structure of the no-explanation defense which limits its applicability to non-model-theoretic accounts of reference-fixing (or non-martian-based accounts, or non-quantum-mechanical accounts, or . . .).

[23]Let me emphasize, here, that I really do think that anyone who accepts the no-explanation defense should also regard the astronomical argument as *sound.* There shouldn't be anything problematic about this position. Given *any* sound argument for P, we can always add an extra premise from astronomy/zoology/quantum-mechanics to generate an "astronomical/ zoological/quantum-mechanical argument for P." As long as these new premises are true, the resulting argument will be *sound* (though it's philosophical significance won't, of course, have very much to do with the new premises we're using to promote it).

it's model-theoretic character; if the model theory turns out to be irrelevant, then Putnam is guilty of some serious false advertising.[24]

This, then, is the serious point which the astronomical argument serves to highlight. Whatever merits the no-explanation defense may have in its own right, the defense constitutes an independent and self-contained argument against realism. If the defense fails, then it doesn't support premise 2 in Putnam's model-theoretic argument. If it succeeds, then it provides a direct argument for Putnam's conclusion and thereby renders the rest of his argument—including *all* of the model theory in that argument—irrelevant. In neither case, therefore, does the no-explanation defense really help Putnam's overall position.

Let me close with two final points on this matter. First, I want to be clear about just what I am and am not claiming here. I'm not claiming that anyone who accepts the model-theoretic argument must also accept the astronomical argument: clearly, anyone who eschews the no-explanation defense of premise 2 is completely off the hook on this one. Nor am I claiming that the astronomical argument shows that the no-explanation defense is unsound. Although I don't think that the no-explanation defense can justify some of Putnam's more heated rhetoric (cf. fn. 20), I do think that Putnam's charge that realists tend to leave reference a matter of "we know not what" fixing interpretation "we know not how" has some real bite. Instead, I'm simply highlighting two *costs* of using the no-explanation defense of premise 2. First, the defense commits us, for better or for worse, to the soundness of the astronomical argument. Second, the defense is general enough that it renders Putnam's model theory superfluous and (thereby) undercuts the philosophical appeal of his larger model-theoretic argument.

This brings me to a second point. On the surface, it might seem like Putnam has an obvious response to the astronomical argument, insofar as this argument is pretty clearly a *silly* argument, while the original model-theoretic argument is, presumably, not so silly. More formally, it might seem like the model theory in Putnam's argument serves to eliminate a *genuine candidate* for reference-fixing, while the astronomy in the astronomical argument serves only to eliminate a straw man. If this is right, then it might explain why the model-theoretic argument is philosophically significant while the astronomical argument is not.

In effect, this response concedes my claim that the no-explanation defense is doing most of the real work for Putnam, but it suggests that Putnam's model theory still plays a small role in establishing that defense's initial claim—i.e., that "presently-available accounts of reference-fixing are inadequate." In particular, the model theory helps to explain why one particular "presently-available account of reference-fixing" is, in fact, inadequate. The astronomical argument, in contrast, doesn't play even this small role. That's why it's not analogous to the model-theoretic argument.

Now, although this response may be initially attractive, I think it suffers from two, fairly-straightforward problems. First, even if the account of reference ruled out by premise 1 were superficially plausible, it still wouldn't be the account which most contemporary realists have actually tried to defend (as evidenced, for

[24]In [6], Garcia-Carpintero expresses a similar worry about the role that model theory ultimately plays in Putnam's overall argument (see pp. 312–13).

instance, by the responses they've given to Putnam in [4], [11], [12], and [22]). So, it's still the case that the non-model-theoretic parts of the no-explanation defense do the *vast majority* of the work in Putnam's overall argument, and it's still the case that Putnam's own emphasis on the model-theoretic side of his argument amounts to some pretty shady promotion.

Second, and more importantly, it's not clear that the account of reference ruled out by premise 1 *is* even superficially plausible. As we saw in the last section (p. 9), premise 1 only serves to rule out an extremely strong form of logicism. As I noted earlier, there's no reason to take this kind of "global logicism" very seriously; nor can I find evidence that other philosophers have ever championed it. Given this, I'm inclined to think that the kind of logicism ruled out by Putnam's premise 1 is just as implausible as the hypothesis about martians that's ruled out by the astronomical argument's premise 1$''$. Hence, I don't think that Putnam can safely rely on plausibility considerations to distinguish between the model-theoretic argument and the astronomical argument.

At the end of the day, then, I don't think that the no-explanation defense provides Putnam with much of an improvement on the just-more-theory defense. Either the defense is too weak to adequately support Putnam's premise 2, or it's *so* strong that it renders the rest of Putnam's argument—including all of Putnam's model theory—completely irrelevant. Neither horn of this dilemma is helpful to Putnam. I'll end this section, therefore, by simply issuing a second challenge to Putnam and his defenders:

> Provide a version of the model-theoretic argument which makes *that* argument seem compelling without, at the same time, making the astronomical argument equally compelling.

4 Concluding Remarks

In the last two sections, I've examined in some detail the philosophical implications of two different generalizations of Putnam's model-theoretic argument. Here, I want to step back and highlight a few of the broader points which this paper has tried to establish. First, most of the explicit model theory in Putnam's argument comes in his defense of premise 1:

> 1. Theoretical and operational constraints do not fix a unique "intended interpretation" of our language.

Despite the scientific-sounding reference to "theoretical and operational constraints," this premise doesn't actually engage with theories of interpretation drawn from empirical linguistics. Instead, it serves to rule out a view which I've called "global logicism"—the view that all terms, predicates and relations in our language can be given *purely logical* definitions. This isn't a view which many philosophers have championed; nor, as far as I can see, is it a view which has much to recommend it. Nevertheless, I think it's the only view on which premise 1 actually gets some philosophical traction.

Second, given the implausibility of the view ruled out by Putnam's premise 1, most of the real work in Putnam's argument has to occur in his defense of premise 2:

2. Nothing other than theoretical and operational constraints could fix the "intended interpretation" of our language.

This, after all, is where Putnam takes on all of the theories of reference that other philosophers have actually proposed. It's also, I would argue, where he takes on all of the *plausible* theories of reference.

Given all this, it's important for Putnam's defense of premise 2 to retain at least some of the model-theoretic character of his defense of premise 1. If it doesn't, then it will almost certainly amount to an independent argument against realism and (so) render Putnam's model theory irrelevant. We've already seen how this can happen in the case of the no-explanation defense, and I think that the point generalizes pretty widely. Hence, I think that Putnam's only real option is to provide a genuinely *model-theoretic* defense of premise 2—preferably, a defense which uses the same kinds of model theory as Putnam used in his defense of premise 1.

Finally, with the exception the just-more-theory defense, I don't know of any defenses of premise 2 which meet this final condition. Unfortunately, as we saw in section 2, the just-more-theory defense is deeply flawed. To keep the just-more-theory defense from slipping into sheer triviality—*a la* the supermodel argument—Putnam needs to provide a positive argument for global logicism. There's absolutely nothing in his writings, however, which would suggest that such an argument is even possible. I conclude, therefore, that unless Putnam can find some genuinely new—and genuinely model-theoretic—defense of premise 2, his larger argument will continue to strike most philosophers as singularly unconvincing.

References

[1] David Anderson. What is the model-theoretic argument. *The Journal of Philosophy*, 93:311–22, 1993.

[2] Timothy Bays. On Putnam and his models. *The Journal of Philosophy*, XCVIII:331–50, 2001.

[3] Timothy Chambers. A quick reply to Putnam's Paradox. *Mind*, 109:195–197, 2000.

[4] Michael Devitt. *Realism & Truth*. Princeton University Press, Princeton, 1984.

[5] Igor Douven. Putnam's model-theoretic argument reconstructed. *The Journal of Philosophy*, 96:479–90, 1999.

[6] Manuel Garcia-Carpintero. The model-theoretic argument: Another turn of the screw. *Erkenntnis*, 44:305–316, 1996.

[7] Alexander George, editor. *Reflections of Chomsky*. Blackwell, Cambridge, 1989.

[8] Bob Hale and Crispin Wright. Putnam's model theoretic argument against metaphysical realism. In Bob Hale and Crispin Wright, editors, *A Companion to the Philosophy of Language*, pages 427–457. Blackwell, Oxford, 1997.

[9] Jussi Haukioja. Not so quick: A reply to Chambers. *Mind*, 110:699–702, 2001.

[10] Frederick Kroon. Chambers on Putnam's Paradox. *Mind*, 110:703–708, 2001.

[11] David Lewis. Putnam's paradox. *Australasian Journal of Philosophy*, 62:221–236, 1984.

[12] Van McGee. Inscrutability and its discontents. *Noûs*, 39:397–425, 2005.

[13] Hilary Putnam. *Meaning and the Moral Sciences*. Routledge, New York, 1978.

[14] Hilary Putnam. Realism and reason. In *Meaning and the Moral Sciences* [13], pages 123–138.

[15] Hilary Putnam. *Reason, Truth and History*. Cambridge University Press, New York, 1981.

[16] Hilary Putnam. Models and reality. In *Realism and Reason* [17], pages 1–25.

[17] Hilary Putnam. *Realism and Reason*. Cambridge UP, Cambridge, 1983.

[18] Hilary Putnam. Model theory and the 'factuality' of semantics. In George [7], pages 213–231.

[19] Hilary Putnam. A defense of internal realism. In *Realism with a Human Face* [21], pages 30–42.

[20] Hilary Putnam. Is the causal structure of the physical itself something physical. In *Realism with a Human Face* [21], pages 80–95.

[21] Hilary Putnam and James Conant. *Realism with a Human Face*. Harvard UP, Cambridge, 1990.

[22] James Van Cleve. Semantic supervenience and referential indeterminacy. *The Journal of Philosophy*, 89:344–361, 1992.