Erkenntnis 67.1 (July 2007): 119–135

More on Putnam's Models

A Reply to Bellotti

Timothy Bays

In 2001, I published a paper which, among other things, criticized the mathematics involved in one version of Hilary Putnam's notorious model-theoretic argument against realism.[1] Recently, Luca Bellotti has taken issue with several of the arguments in that paper.[2] Here, I want to respond to Bellotti's objections and to take the opportunity to make a few, somewhat more general remarks concerning the mathematical side of Putnam's project.

1 Putnam and Bays

Let me begin by recapping the the main lines of Putnam's argument (or, at least, of the small portion of that argument which is under discussion here). In the first few pages of his 1980 paper, "Models and Reality," Putnam tries to show that there is an "intended model" for the language of set theory which satisfies the set-theoretic axiom V= L.[3] He begins by assuming that there are only two things which could play a role in fixing the intended model for set-theoretic language. First, there are what Putnam calls "theoretical constraints." These include the standard axioms of set theory (which the intended model must satisfy), as well as principles and theories from other branches of science (which the intended model must at least be compatible with). Second, there are "operational constraints." These are just the various empirical observations and measurements that we make in the course of scientific investigation.[4]

[1]Timothy Bays, "On Putnam and his Models," *The Journal of Philosophy* XCVIII (2001): 331–350. The version of Putnam's argument that I am responding to occurs on pages 4–8 of Hilary Putnam, "Models and Reality," in *Realism and Reason* (Cambridge: Cambridge UP, 1983): 1–25.

[2]Luca Bellotti, "Putnam and Constructibility," *Erkenntnis* 62 (2005): 395–409.

[3]Let me say a bit more about the context of this argument. Putnam's larger goal in this section of his paper is to show that sentences like V= L "have no determinate truth value. . . they are just true in some intended models and false in others." As a result, it doesn't "make sense" to think that "'V= L' is *really* false, even though it is consistent with set theory" (p. 5). Now, because Putnam takes himself to be arguing against Gödel—who thought that there was a *unique* "intended interpretation" of set theory and that V= L was false on that interpretation—Putnam doesn't feel the need to argue for an intended model satisfying V≠ L (he assumes that Gödel will grant him the existence of this model). As a result, Putnam thinks that if he can simply find an intended model which satisfies V= L, then he will have finished his argument; in his words, he will have shown that Skolem's famous "'relativity of set-theoretic notions' extends to a relativity of the truth value of 'V= L'" (p. 8).

[4]At first glance, it may be difficult to see how all this scientific material—i.e., operational constraints and those theoretical constraints which come from natural science—could even have a bearing on the intended model of *set theory.* Since this issue is tangential to the present paper, I won't discuss it in any detail. See pp. 5–7 of Putnam's paper, pp. 333–334 of my original paper, or p. 397 of Bellotti's paper for more detailed analysis of this issue.

Given this, Putnam argues that finding an intended model which satisfies V = L simply requires finding a model of ZF + V = L which satisfies the theoretical and operational constraints we inherit from the natural sciences (the constraints coming from set theory itself are covered by the stipulation that our model satisfies ZF). His strategy for finding this model rests on the following theorem:

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

Using this theorem, Putnam argues as follows. Let OP be a countable collection of real numbers which codes up all the measurements and observations human beings will ever make. By our theorem, there is a model of ZF + V = L which contains OP (or, at least, a formal analog of OP). Further, this model can be extended so as to take care of any theoretical constraints coming from natural science. Putnam 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).

In short: as long as the only constraints on the intended model of set theory 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, Putnam can find an intended model of set theory in which V = L comes out true.

This, then, is the version of Putnam's argument that will be at issue in this paper. To understand the debate over this argument, we need to start by looking more closely at Putnam's proof of the above theorem. The theorem says that if X is a countable collection of real numbers, then there exists an ω-model, $\mathbb{M}$, such that $\mathbb{M} \models$ ZF + V = L and $\mathbb{M}$ contains an "abstract copy" of the set X. Putnam's proof begins by noting that, in the special case in which we allow $\mathbb{M}$ to be countable, we can code both $\mathbb{M}$ 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 (. . . M, s, . . .). 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)

This completes Putnam's proof. Our discussion of this proof will focus on three substantive questions: 1.) does this proof work? 2.) to the extent that it doesn't work, can it be fixed? and 3.) how do the mathematical details of this proof—or of any potential replacement proofs—interact with the larger philosophical theses that Putnam is trying to defend in his paper?

These questions lead us back to my original 2001 paper. There, I made four claims concerning the mathematical side of Putnam's argument. First, I claimed that Putnam's proof is mistaken: in ZFC, you can't apply the downward Löwenheim-Skolem theorem—or even the "Skolem Hull construction"—to obtain an elementary submodel of L. Second, I noted that Putnam's theorem can't be proved in ZFC at all: Putnam's theorem entails the consistency of ZFC, so the second incompleteness theorem rules out a proof of this theorem *in* ZFC (unless, of course, ZFC itself is inconsistent). Third, I observed that Putnam's theorem *can be* proved in relatively modest extensions of ZFC—e.g., ZFC plus "there exists an inaccessible cardinal" or ZFC plus "there exists an uncountable transitive model of ZFC."[5]

Clearly, these first three claims are pretty technical. The fourth claim is somewhat more philosophical. Suppose that Putnam *does* decide to use some modest extension of ZFC to prove his theorem. Then, although it's clear that a proof will now go through, it's no longer clear that Putnam's theorem will still serve the philosophical purposes that he originally wanted it for. After all, anyone who accepts the mathematics needed for Putnam's new proof has, *eo ipso*, theoretical constraints which go somewhat beyond ZFC. Hence, it's not clear why a model of ZF + V = L should satisfy *these* philosophers' theoretical constraints. As I put the matter in my earlier paper (p. 399):

> Recall, here, the philosophical *point* of Putnam's theorem. He 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... then it is hard to see how his 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 is not enough for Putnam to build a model which satisfies ZF + V = L; he 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 does not do what he wants it to do.

This, therefore, highlights an essentially technical problem for Putnam's argument. The argument depends on a key theorem which Putnam is not in a position to prove (his own proof doesn't actually work). Nor, for reasons relating to the incompleteness theorem, can Putnam put himself in a position to prove this theorem without jeopardizing the very philosophical point which the theorem is supposed to support. In short: the above-mentioned mistake in Putnam's proof (claim 1) cannot be fixed without undercutting Putnam's overall argument (claim 4).

---

[5]As Bellotti notes, these first three points were made independently by Daniel Velleman in a *Mathematical Reviews* note of Michael Levin's "Putnam on Reference and Constructible Sets," *British Journal for the Philosophy of Science* 48 (1997): 55–67 (see *Mathematical Reviews* 98c (1998): 1364). This note appeared after I had begun circulating my earlier paper, and I remained unaware of it until I saw Bellotti's paper. Similar points have subsequently been made by Haim Gaifman in "Non-Standard Models in a Broader Perspective," in *Non-Standard Models of Arithmetic and Set Theory*, ed. Ali Enayat and Roman Kossak (New York: American Mathematical Society, 2004).

For reasons of space, I won't spell out all the mathematical details of these first three claims here. The reader who would like a more extended discussion of these details is encouraged to look back through pp. 335–340 of my earlier paper (see especially the lengthy explanations in footnotes 3, 6 and 7).

2 Bays and Bellotti

Turn, now, to Bellotti's new paper. There, Bellotti presents three objections to the technical arguments I've just sketched.[6] The first concerns claim 1. In claiming that Putnam's proof was mistaken, I assumed that Putnam was working in ZFC when he presented this proof. As Bellotti notes, however, "Putnam is not clear about the theory in which he is working" (p. 404). Given that there are several modest extensions of ZFC in which Putnam's theorem *can* be proven (see claim 3 above), principles of charity require us to assume that Putnam was working in one of these alternate theories when he proved his theorem. Hence, it "seems unfair" to charge Putnam with a mathematical mistake (p. 404); at most, Putnam is guilty of some sloppy presentation insofar as he failed to make his background assumptions sufficiently explicit. In Bellotti's words: "the trouble does not come from a mathematical mistake, contrary to what Bays claims, unless the absence of a clear statement of the theory in which one is working can be considered an outright mistake" (p. 405).

To assess this first objection, we need to start by drawing a distinction between, on the one hand, the background set theory needed to prove Putnam's theorem and, on the other hand, the background set theory needed to make Putnam's *own* proof of his theorem go through. These turn out to be rather different. As Bellotti notes, Putnam's theorem can be proved in rather weak extensions of ZFC. In general, however, these weak theories won't license the key move in Putnam's own proof—i.e., the use of the Skolem Hull construction to build an elementary submodel of L.[7] To make Putnam's own proof go through, we'll need some kind of strong class theory—say, something on the order of Morse-Kelley set theory.[8]

[6] I should emphasize that Bellotti's paper is a good deal broader than my discussion here may suggest. Bellotti has his own (quite persuasive) criticisms of Putnam's argument, and his paper includes a detailed analysis of a recent discussion of Putnam's argument by Michael Levin (see Levin, "Putnam on Reference and Constructible Sets"). For the purposes of this paper, however, I'm going to leave these other aspects of Bellotti's argument aside and focus rather tightly on the paper's treatment of the arguments sketched in the last section.

[7] On pp. 396–397 and 404–405 of his paper, Bellotti discusses several strategies for correcting Putnam's proof. These range from assuming the existence of an inaccessible cardinal to simply assuming that "for any constructible set of natural numbers *s*, there is a *standard* model M of ZF such that *s* is in M (and M is in L)." Let me make two comments concerning these suggestions. First, and as Bellotti is clearly aware, most of these suggestions don't save Putnam's original proof (they simply permit alternate proofs of Putnam's theorem). As noted in the main text, assuming the existence of large cardinals or of small transitive models of ZF won't license Putnam's use of the Löwenheim-Skolem theorem (or even the Skolem Hull construction).

Second, at least one of Bellotti's suggestions *does* seem to be aimed at saving Putnam's original proof (see, e.g., Bellotti's comments on p. 397; see also the above-cited passage from p. 405). On p. 396—and again on p. 404—Bellotti suggests that we could simply "assume the existence of a countable family of Skolem functions for L." The idea, presumably, is that we could add some kind of class variables to our language and then add axioms which ensure that some of these variables code Skolem functions for L. This strategy *would* allow something like Putnam's original proof to go through, but it would involve going well beyond ZFC to make that proof work. (Note, here, that it's not enough to simply add new function symbols to our language and then axiomatize the claim that these functions act as Skolem functions: the resulting theory is conservative over ZFC, so it doesn't let Putnam prove anything that he couldn't prove before.)

[8] At any rate, this will be true if we limit ourselves to fairly standard extensions of ZFC. There are, of course, more *recherché* and *ad hoc* extensions which would license the constructions Putnam needs. So, for instance, we could simply add a new truth

In thinking about claim 1, therefore, the question isn't whether Putnam slipped up and neglected to mention a relatively modest background assumption—e.g., something like Con(ZFC) or "ZFC has an ω-model"—it's whether he used a fairly powerful theory of classes to prove his theorem while writing as though ZF exhausted our theoretical constraints. It seems to me that there are two reasons to be cautious about making this latter assumption. The first is just the fact that Putnam *does* write as though ZF includes all of our theoretical constraints (or, at least, all of these constraints which come from set theory). On page 7 of his paper, for instance, Putnam explicitly claims that his model "satisfies all theoretical constraints," and he then discusses some problems with simply adding new axioms to ZF (claiming that adding such axioms "can hardly be acceptable from a realist standpoint"). Similar remarks occur elsewhere in Putnam's paper.[9] Given this, it's at least awkward to assume that Putnam was—either knowingly or inadvertently—keeping some kind of strong class theory hidden in the background of his argument.

Second, using such a class theory would lead to a rather serious version of the problem isolated at the end of the last section (i.e., in claim 4). Many set theorists have concerns about using strong class theories and would be reluctant to accept a proof which depended on such theories, while those who are comfortable with strong class theories clearly have theoretical constraints which go (well) beyond those dealt with in Putnam's theorem. Further, I think this problem would be pretty obvious if Putnam were to make his argument explicit in the way Bellotti seems to suggest: e.g., if Putnam were to simply come out and say "I'm going to use something like Morse-Kelley set theory to prove that your theoretical constraints—which, I insist, *must be* limited to ZFC—have an intended model which satisfies V= L." The objections to this kind of move would, I'm pretty sure, come in thick and fast, and I see nothing in Putnam's paper which would enable him to deal with those objections.

These, then, are some reasons for being cautious about assuming that Putnam is—without bothering to

predicate to our language and then expand our comprehension and replacement schemes so as to include formulas involving this predicate (in effect, this is *how* Morse-Kelley would go about proving Putnam's theorem). Without classifying all of the different options on this front, let me simply highlight three general facts. First, any extension of ZFC that licenses Putnam's proof will involve expanding the *language* of set theory—e.g., by adding class variables or a new truth predicate. The property "x is an elementary submodel of V," can't be expressed in standard set-theoretic language; so, theories formulated in this language won't license Putnam's argument. In particular, therefore, normal large-cardinal axioms won't be enough to make Putnam's proof go through. Second, it's not sufficient to simply add a new constant to our language and to then axiomatize the claim that this constant designates an elementary submodel of V. Although the resulting theory is equiconsistent with ZFC, and although it *will* guarantee that the universe contains a set-sized elementary submodel, a simple compactness argument shows that it *won't* be enough to prove Putnam's theorem. Finally, I should note that the more specialized and non-standard Putnam's background set theory is, the less plausible it is to think that he simply *forgot* to make his background assumptions explicit. For philosophical purposes, therefore, it's important to focus our attention on relatively standard extensions of ZFC.

[9] So, for instance, the argument we're currently examining is an extension of another, somewhat simpler, Löwenheim-Skolem argument in which Putnam merely tries to build a *countable* "intended model" of set theory (without showing that this model satisfies V=L). Putnam claims that the model so constructed satisfies "a *formalization of all our beliefs*" (p. 3), and that it constitutes a "model for our *entire body of belief*" (p. 4). Similar comments can be found in the discussion on pp. 11–15. From a textual standpoint, therefore, it certainly looks as though Putnam thinks that his models satisfy *all* of our set-theoretic beliefs (and hence, that he regards these beliefs as a subtheory of ZF + V= L).

mention it—using a strong background set theory to prove his theorem. I'll admit that these reasons aren't conclusive. We're in the following situation. On the one hand, if Putnam is working in something like ZFC, then his proof doesn't go through (for the reasons I have already indicated). On the other hand, if Putnam is using a background set theory strong enough to make his proof go through, then he's keeping something pretty substantial hidden from view, something which would lead to immediate and significant philosophical objections if it were placed out into the open. Since both of these options leave Putnam in hot water, it's not clear which way principles of charity require us to go in interpreting him. Bellotti clearly thinks we should take the second option; I'm (still) inclined towards the first; I acknowledge, however, that it's a close call.

Let's turn to Bellotti's second objection. On pages 405–406, Bellotti argues that, even if Putnam does have to go somewhat beyond ZFC to prove his theorem, this doesn't have the deleterious consequences for his argument which I claimed it does (i.e., in claim 4). On Bellotti's analysis, anyone who accepts ZFC can, by a process of reflection, come to see that they should *also* accept the various set-theoretic assumptions needed to prove Putnam's theorem. Bellotti writes:

> First, despite the consistency of ZF plus the negation of the hypothesis needed by Putnam (if ZF is consistent), it seems that *in this context* nobody could be in a position to assume ZF as a "theoretical constraint", taking it seriously as a constraint, at the same time denying the needed hypothesis itself. For example, one who believes in ZF as part of our "best theory of the world" should be willing to concede the existence of (say) one inaccessible cardinal, since this is informally justified on the same (or a similar) theoretical basis as the other axioms. (p. 405)

Further, Bellotti seems to think that argument can be iterated, leading from one inaccessible to two, from two to three, from three to four, etc. In the end, he concludes that "Putnam can well assume the existence of as many inaccessibles as he needs, since this is surely part of 'our best axioms for set theory' " (p. 405).

Now, this is clearly an interesting and provocative suggestion, and it's one that I actually have a good deal of sympathy for.[10] That being said, I don't think it's a suggestion that helps Putnam very much. There

[10]In the long run, of course, the plausibility of this suggestion will depend largely on the exact hypotheses under discussion. As Bellotti notes (see p. 404 and 405), the argument is most plausible in the case where we're simply adding consistency statements for our current theoretical constraints (though, alas, this won't be enough to generate Putnam's own theorem). It's least plausible when we're moving directly to the kind of strong class theories needed to make Putnam's original proof go through. Bellotti's own example of small large-cardinals—inaccessibles, Mahlo's, etc.—is something of an intermediate case. Personally, I find the argument quite persuasive in this case; I don't, however, think it's *obviously* persuasive, and I think many philosophers would resist the argument here (and would be able to give good, if not overwhelming, reasons for doing so).

Let me make two further comments concerning this line of argument. First, as Bellotti notes on pp. 406–407 of his paper, there's a danger for Putnam in following this line to its logical conclusion. If it's carried too far, after all, the line might lead to the acceptance of large large-cardinals—e.g., measurable or Woodin cardinals—and these would directly rule out the possibility that V = L. To make this line work, therefore, Putnam needs to strike a delicate balance. He needs to assume that we can reflect our way far enough up the large-cardinal hierarchy to make his overall argument work, without getting all the way up to the kinds of large large-cardinals which would trivialize his overall project.

Second, I should note that the term "reflection" itself does not occur in Bellotti's paper. I've introduced it as a convenient shorthand for arguments which use the informal justification of our current set-theoretic axioms—or perhaps simply the presumed consistency of those axioms—as grounds for accepting new set-theoretic axioms. For Bellotti's own discussion of this kind of process, see pp. 405–408 of his paper.

are two cases to consider. On the one hand, there's the case where Bellotti's process of reflection reaches some terminal point—some point where we accept a certain collection of axioms as theoretical constraints, but don't feel any impulse to reflect our way up to further axioms.[11] In this case, we'll eventually find ourselves back in the situation I described a few pages ago: we'll have a stable collection of set-theoretic axioms (our "best axioms for set theory"), we won't be inclined to accept any new axioms, but Putnam will have to *use* new axioms to prove his central theorem.[12] In this situation, *we* will have no reason to accept Putnam's theorem, and anyone who does accept his theorem will have grounds for denying that the theorem takes care of *their* theoretical constraints.

On the other hand, there's the case where Bellotti's process of reflection extends indefinitely: for any plausible set of axioms, T, there is always some further axiom, A, such that 1.) reflection on T naturally leads us to accept A and 2.) T+ A is strong enough to prove a version of Putnam's theorem *for* T. This is, I expect, the case that Bellotti is really interested in. It seems to me, however, that it's also a case where Putnam's argument goes wrong at a very early point—namely, at the point where Putnam insists that we start by formalizing our "theoretical constraints" as a fixed collection of sentences. On the current line, when we accept ZFC we implicitly commit ourselves to accepting an indefinitely large collection of further axioms; hence, no fixed collection of sentences will completely capture the "theoretical constraints" involved in our initial acceptance of ZFC. In this situation, therefore, Putnam's whole apparatus of (fixed) theories and models seems decidedly inappropriate.[13]

Let me make three comments on this dilemma. First, I should emphasize that it's Putnam's *own* development of the model-theoretic argument which turns on representing theoretical constraints as fixed sets of sentences; this isn't something I introduced into the argument in the course of my criticisms. As a textual matter, Putnam always starts his model-theoretic arguments with a fixed theory—usually, one reached at some hypothetical "ideal limit of inquiry."[14] More formally, Putnam *needs* fixed theories in order to apply his various model-theoretic constructions (since we don't have a worked-out model theory for "indefinitely

[11]For instance, perhaps we're willing to accept a proper class of inaccessible cardinals, but balk at accepting a Mahlo cardinal. Personally, I'm fine with Mahlo cardinals, but this doesn't strike me as an obviously silly position for someone to take.

[12]Note, here, that the axioms Putnam needs in this "case 1" situation may well be substantially stronger than those needed to prove Putnam's *original* theorem. Once our theoretical constraints have gone beyond ZFC, Putnam's new axioms will have to be strong enough to deal with these extended constraints—i.e., to build a model which satisfies these extended constraints as well as ZF. See pp. 399-340 of my original paper for a more detailed discussion of this kind of issue.

[13]What Putnam really seems to need here is an unstable amalgamation of the two cases just considered. He needs us to start with a fixed collection of sentences which constitutes the totality of our theoretical constraints. We then reflect *just enough* to generate some new principles which will allow us to prove a (relevant) version of Putnam's theorem. Finally, we *forget about* this whole process of reflection and return to viewing our original collection of sentences as constituting the totality of our theoretical constraints (without, in the process, abandoning our new-found commitment to Putnam's theorem). I see no reason why any realist should find this particular procedure at all compelling.

[14]Note that on the line we're currently considering, there *can't be* an "ideal limit of inquiry." Given any purported limit, we can always go one step further to generate a new axiom which will let us prove a version of Putnam's theorem for that "limit." Hence, the whole set-up of Putnam's argument would have to be seriously revised to make it fit the current line of argument.

extensible" terms and sentences). Further, one of the more controversial moves in Putnam's argument—his famous "just more theory" move—depends on recasting various (purported) constraints on interpretation as mere sets of sentences and then applying his model-theoretic constructions *to* those sentences. Hence, the insistence that we represent theoretical constraints as fixed sets of sentences is something which runs quite deep in Putnam's *own* presentation of his argument.

Second, although I've presented this dilemma as a response to Bellotti's objections, nothing in my argument really depends on Bellotti's own understanding of the ways we could/should extend our set-theoretic axioms. To formulate the argument as a direct challenge to Putnam, we simply ask whether our theoretical constraints should be represented by fixed sets of sentences. If Putnam thinks they *should be* so represented (as his own writings would certainly seem to indicate), then his argument will run aground on the incompleteness considerations discussed in the last section. If he thinks they *shouldn't be* so represented, then his argument won't even get out of port, since he will have no initial theory to which to apply his model-theoretic constructions. In either case, therefore, Putnam's overall argument seems to be in trouble.

Finally, I want to reiterate that these comments should not be misconstrued as criticisms of Bellotti's positive views on the ways we can/should be led to extend our set-theoretic axioms. As I mentioned earlier, I find the idea that reflection on the informal justification of ZFC should lead us to accept large-cardinal axioms quite persuasive. My only concern is with the way this analysis meshes with Putnam's larger model-theoretic argument. If the analysis is understood locally—i.e., as simply explaining how we could generate the principles needed to prove the *original* version of Putnam's theorem—then I think it only pushes Putnam's problems down the road a bit. If it's understood more globally, then it suggests a picture of mathematical understanding that's fundamentally incompatible with the model-theoretic machinery used in Putnam's argument. Whatever its independent merits may be, therefore, I don't think Bellotti's analysis helps to defend Putnam's argument against the criticisms in my earlier paper.

This brings me to Bellotti's third objection to my criticisms. So far, all of my arguments have focused on the fact that Putnam is unable to prove a version of his theorem that will cover *all* of our theoretical constraints (including any constraints introduced in the course of the proof itself). On pp. 407–408, however, Bellotti suggests that Putnam may not need to prove quite this much. He writes:

> Finally, Bays' request that Putnam prove that his final model satisfies the conditions assumed in order to prove its existence seems excessive from the start, since it is unfulfillable in any interesting case, because of Gödel's theorem on consistency proofs. . . . But, although Putnam is unclear, and he could at some points seem to be aiming at a *proof* that his model satisfies all theoretical constraints, in fact he is not obliged to *prove* that (he is well aware that, of course, he could not *prove* anything for operational constraints): it seems sufficient for his argument that the later *might be the case.* (p. 407)

On this reading, then, Putnam's argument isn't supposed to provide a knock-down refutation of realism; it's simply supposed to raise a kind of *hypothetical* problem—to show that there "might be" a case where serious difficulties for realism would arise. As Bellotti puts this point later: "there *might be* a Putnamian model satisfying all theoretical constraints . . . although we cannot prove its existence—and this is sufficient for Putnam's argument" (p. 407).

To see what Bellotti is getting at here, it's useful to consider what an *explicitly* hypothetical version of the model-theoretic argument might look like. Here's one possible sketch:

> Let T be a theory which comprises all of our theoretical and operational constraints, and consider the following large cardinal axiom, A. Note that *if* A is true, then we can find a non-standard model for T. So, *if* A is true, then the language of set theory is semantically indeterminate.
>
> Of course, A doesn't actually follow from T (so there's nothing in our theoretical constraints which *forces* us to accept A as true). Still, here are some reasons for thinking that A is somewhat plausible—that it's the kind of axiom which *might well be true.* Further, as far as we know, there's nothing in T which conflicts with A. Given this, you [the realist] should at least admit that A is *possible.* Hence, semantic indeterminacy is something that you should worry about.

If I understand him right, Bellotti's point is simply that this kind of hypothetical argument would put real pressure on (at least some) realists. If realists manage to evade the argument, it will be because they reject Putnam's initial assumption that every model for T provides an "intended interpretation" of set-theoretic language; it won't be because Putnam's *proofs* turn out to be inadequate.

Note, here, that in taking this line Bellotti gains the resources to counter several of the arguments that I gave earlier in this paper. So, for instance, the dilemma I sketched back on pp. 6–8 rested on the assumption that, whenever we reflect our way to a new set-theoretic principle, we should then add that principle to our overall collection of theoretical constraints. On Bellotti's new analysis, however, reflection doesn't need to go this far. Instead of convincing us that a new principle is *true*—and hence that it should be added to our theoretical constraints—reflection only needs to convince us that a new principle is *possible*—that it's the kind of principle which might well be true, and that we should, therefore, take its potential consequences pretty seriously.[15]

Similarly, when I discussed the soundness of Putnam's original proof, I argued that attributing a strong class theory to Putnam would violate principles of charitable interpretation, since it would open him up to immediate (and powerful) objections from other philosophers: either he went beyond those philosopher's theoretical constraints in *starting* his argument or he hadn't taken care of their theoretical constraints by the *end* of his argument. However, these objections would be at least partially mitigated if Putnam was giving a purely hypothetical version of his argument. After all, the hypothetical version only uses class theory to form a *conditional* judgment, and it fairly explicitly *intends* the antecedent of this conditional to lie outside the realist's theoretical constraints. If this is right, then my charity concerns are also somewhat mitigated.[16]

[15] Of course, this analysis walks a very fine line. If we don't find a particular set-theoretic principle very plausible, then we are unlikely to take its consequences all that seriously. If we find a particular principle *too* plausible, then we'll want to add it to our theoretical constraints. If we find *too many* principles too plausible, then we may well conclude that our set-theoretic commitments are indefinitely extensible. Any of these three cases will be problematic for Putnam. For Putnam's purposes, we need to find a (single) set-theoretic principle which is *just plausible enough* to take seriously, but still *sufficiently implausible* that we can avoid the other cases just mentioned. Clearly, this is a pretty delicate position to establish.

[16] That being said, I don't think this argument touches any of the purely-textual evidence for thinking that Putnam's proof was originally supposed to take place in ZFC. I said a bit about this on pp. 4–5, and I'll say a bit more later in the paper. In the end, I think that purely-textual considerations make it very difficult to attribute a strong background set theory to Putnam (even in the hypothetical mode). Hence, I think it's hard to avoid the conclusion that his original argument turned on a serious,

Now, let me say from the outset that I think Bellotti is onto something important here and that, on one level, I actually agree with his overall analysis. As I argued in my earlier paper, the deepest problem in Putnam's argument lies, not in his inability to prove the existence of a particular kind of model, but in the very idea that proving such existence would show something significant about realism. If Putnam were right that any structure which happened to satisfy a first-order formalization of our "theoretical and operational constraints" would count as an "intended model" for our language, then realism would be in a precarious position (even if Putnam himself couldn't prove the existence of such structures). As I wrote previously:[17]

> 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 is not much to stake a metaphysics on! (pp. 340–341)

In short, then, Bellotti is certainly right to think that Putnam's inability to *prove* a certain theorem doesn't get realism completely off the hook. Given his other assumptions—e.g., the aforementioned "premise 2"—Putnam's argument raises the strong possibility of semantic indeterminacy. So, if Putnam's overall picture of the relationship between models and interpretations is correct, then realism is still on shaky ground (whether or not Putnam can actually *prove* a strong enough version of his theorem).

All that being said, I think Putnam's inability to prove a strong version of his theorem is a little more significant than Bellotti's analysis makes it out to be. There are two issues here. First, however interesting the hypothetical version of the model-theoretic argument may be, I don't think it's the version that Putnam himself actually defends. Earlier, I presented some textual evidence for thinking that Putnam takes himself to have really proven the existence of a model which satisfies *all* of our theoretical constraints (see pp. 4–5), and I think that similar claims can be found in Putnam's other presentations of the model-theoretic argument.[18] More importantly, I don't see any *positive* reasons for thinking that Putnam ever intended

---

though subtle, mathematical mistake.

[17]Here, "premise 1" is just the claim that there are many models which satisfy our theoretical and operational constraints. "Premise 2" is (essentially) the claim that any such model should count as an "intended interpretation" for our language.

[18]So, for instance, when Putnam's recaps his "Models and Reality" argument in the introduction to *Realism and Reason,* he writes "what I show is that no matter what operational and theoretical constraints our practice may impose on our use of language, there are always *infinitely many different reference relations*. . . which satisfy all of the constraints" (p. ix). Similarly, in his more recent paper, "Model theory and the 'factuality' of semantics," Putnam claims that the existence of multiple models for our theoretical and operational constraints is "an undisputed result of modern logic" (p. 214).

More broadly, I think the overall *rhetoric* of Putnam's presentation of the model-theoretic argument is incompatible with giving that argument a merely hypothetical reading. Putnam regularly claims that his argument shows realism to be "incoherent" or "unintelligible." In "Models and Reality," he claims that his Löwenheim-Skolem argument uncovers an outright "antinomy" in the philosophy of language. To support such strong claims, I think Putnam would need to *really prove* a full to give a hypothetical version of his argument: there's no place in his writing where he acknowledges that his proofs may require premises which go beyond his interlocutor's theoretical constraints, there's no place where he admits that his conclusions are "merely conditional," and there's no place where he comes out and says that he's only talking about what "might be the case." From a textual standpoint, therefore, I think it's pretty clear that Putnam himself *was not* giving a hypothetical version of his argument. Hence, I think the criticisms I sketched in section 1 (and in my earlier paper) still work against the argument that Putnam was actually making.[19]

Second, I think the purely mathematical problems discussed in section 1 highlight some more general problems in Putnam's argument. As we saw earlier, the mathematical problems all turn on the fact that Putnam's argument requires him to use slightly more mathematics than that which is accepted by those against whom he is arguing. This need to go beyond what his opponents accept—or, perhaps, what he is willing to *allow* them to accept—is characteristic of Putnam's overall argument. In my original paper, I showed that Putnam's (in)famous "just more theory" strategy turns on exactly this kind of maneuver.[20] Similarly, Bellotti's own criticisms of Putnam focus on the fact that Putnam's argument uses the notion of *finitude* in ways that Putnam will not—and, indeed, cannot—allow the realist to use it.[21] Hence, even if Putnam *could* retreat to a hypothetical version of his argument, the mathematical problems in the original version are still worth exploring, insofar as they bring out more general features of Putnam's argument.

These, then, are two reasons for shying away from Bellotti's suggestion that we read Putnam as giving a hypothetical version of the model-theoretic argument. Although I agree that the hypothetical argument is

---

version his theorems; I don't think the claims are adequately supported by a mere "well, there *might be* a problem here" kind of hypothetical.

[19] In my original paper, I formulated this point somewhat differently. While acknowledging that Putnam's argument raises the serious possibility of semantic indeterminacy, I wrote:

> On the positive side, [my] 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 antirealism, then my argument shows that he cannot do it. (p. 340)

What I may, perhaps, have insufficiently emphasized is that Putnam himself seems to want to stand toe-to-toe with the realist and to *prove* semantic antirealism. So, even if Putnam *could* fall back on a hypothetical version of his argument, it wouldn't be the argument he originally gave, and my criticisms would still show that his *original* argument doesn't work.

[20] Very roughly, the problem is this. Putnam's model-theoretic argument requires him to stand back from our best overall theory of the world in order to describe that theory's semantics—e.g., to show that the theory has many different models and that some of these models are quite pathological. In contrast, his "just-more-theory" argument refuses to allow the realist to stand back to describe her preferred semantics—e.g., to insist that models of set theory must be transitive or well-founded, or that models of ordinary language must respect certain causal constraints. At best, Putnam allows the realist to add new *sentences* to her theory—sentences which get interpreted using *Putnam's* favorite semantics. In effect, then, Putnam allows himself just a little more than he allows those against whom he is arguing: *they* must work within a particular theory, while *he* can step outside that theory to specify its intended semantics. For a more detailed discussion of this issue—and of its relation to the mathematical issues discussed in section 1—see sections IV and V of my original paper.

[21] Putnam uses the notion of finitude to describe the ω-models mentioned in his main theorem. If the realist could also use this notion to describe models, then she could insist that intended models must be well-founded, and this would be enough to *rule out* the models generated by Putnam's main theorem. See pp. 401–403 (also 405 and 407) of Bellotti's paper for his discussion of this issue. See also the somewhat briefer discussion on pp. 344–345 and p. 349, n. 18 of my earlier paper.

more plausible than the argument discussed in section 1 (and essentially said this much in section II.2 of my original paper), I don't think that it's the argument Putnam actually gives. Further, I think that focusing too heavily on the hypothetical argument would lead us to miss some of the broader, more structural, features of Putnam's original argument. Hence, for the purposes of evaluating Putnam's own argument, the hypothetical argument should (at least temporarily) be set aside. When we do so—and when we focus resolutely on the details of Putnam's *own* argument—then I think we'll find that the criticisms presented in section 1 still go through.

At the end of the day, this assessment of the hypothetical argument mirrors my overall response to Bellotti's paper. From a purely philosophical perspective, there's a lot to like in Bellotti's paper. His hypothetical version of the model-theoretic argument is more plausible than Putnam's original version was, and his suggestions as to how new set-theoretic principles could be motivated are compelling. When these two parts of Bellotti's paper are put together, they provide the best positive development of the mathematical side of Putnam's Löwenheim-Skolem argument that I've seen in the literature. From a more textual perspective, however, I find Bellotti's paper less persuasive. Insofar as it eschews the aim of providing an outright *proof* that there exist non-standard models which satisfy *all* of our theoretical constraints, I think Bellotti's argument aims at a substantially weaker conclusion than Putnam's original argument did. Given the notoriety of Putnam's argument—and given that this notoriety stems largely from the role that model theory plays in the argument—I think it's important to be clear as to what Putnam's model theory can and cannot do for him. In particular, I think it's important to understand why Putnam's argument does not—and indeed *cannot*—provide a genuine proof that his various non-standard models really exist. Explaining this was the point of (section 2 of) my original paper, and, as far as I can see, this point survives all of Bellotti's criticisms.

3 Mathematics and Philosophy

So far, this paper has focused fairly tightly on the mathematical side of Putnam's argument—on whether or not Putnam can really prove certain key theorems. While this mathematical focus was necessary for the purposes of responding to Bellotti, it's something I have mixed feelings about. In this section, I want to make a few comments about the advantages and disadvantages of this kind of tight mathematical focus.

On the disadvantages side, I don't think the mathematical issues discussed in the last few sections lie at the philosophical heart of Putnam's argument. There are three problems here. First, and from a purely structural perspective, the mathematical side of Putnam's argument only rules out one possible way of fixing the intended model(s) of set theory. By displaying non-standard models of our theoretical and operational constraints, Putnam eliminates the possibility that realists could use a simple form of first-order "implicit definition" to fix their intended model(s). Other parts of Putnam's argument then rule out other methods of fixing intended models. Since most of the model-fixing methods that realists have actually proposed—set theoretic "perception," second-order definition, etc.—fall under these other cases, it's the other parts of

Putnam's argument which do most of the actual work. Hence, as interesting as Putnam's mathematics may be, it's not where the real argument takes place.[22]

Second, and as I've already mentioned above, Putnam's mathematics is strong enough to raise the real *possibility* of semantic indeterminacy. If we accept the premise that any structure which satisfies a first-order formalization of our theoretical and operational constraints counts as an "intended model" for our language, then realism is in a precarious position (whatever Putnam may or may not be able to prove). Given this, the heart of Putnam's argument has to lie in his defense of this initial premise. Again, therefore, examining the philosophical point of Putnam's argument should lead us away from an overly tight focus on the details of his mathematics.

Finally, although I think the mathematical side of Putnam's argument involves some genuinely complex issues—e.g., those discussed in section 2—I think the issues involved in other parts of his argument are more fundamental. It's in these other parts, after all, where we ask the big conceptual questions: what makes some models "intended" and others "unintended"? how do we distinguish between "standard" and "non-standard" models of a mathematical theory? what role does higher-order logic play in our understanding of mathematics? In my view, these are the most interesting questions raised by Putnam's argument, and they are not, in the final analysis, questions which turn on the specific details of Putnam's proofs.[23]

All that being said, I still think there are some good reasons for continuing to pay attention to the mathematical details of Putnam's argument, some "advantages" if you will. These go beyond any intrinsic interest these details may have for us and/or any role they may play in provoking us to ask new philosophical questions (e.g., the questions about new axioms or indefinite extensibility raised in the last section). Instead, the most important reason for focusing on Putnam's mathematics involves the role this mathematics plays in *promoting* Putnam's argument. Much of the excitement surrounding Putnam's argument stems, less from his conclusion that semantic realism is untenable, then from his claim that basic theorems of model theory *show* that semantic realism is untenable. (This, after all, is what makes his argument a *model-theoretic* argument against realism.) Given this, I think we need to be clear about just what Putnam's model theory really does for him—about the assumptions he needs to prove various theorems and about the philosophical consequences of making those assumptions.

Let me head off a possible misunderstanding here. In focusing on the problems with Putnam's model theory—and, in particular, on Putnam's inability to really *prove* certain key theorems—it may seem that I'm taking sides in a long-running dispute concerning the burdens of proof in realist/anti-realist debates. Putnam thinks its up to the realist to explain how our language gets its intended interpretation (and, for the reasons discussed in section 1, he thinks that the realist will be unable to meet this explanatory burden). Realists, on the other hand, tend to think that the burden of proof lies with Putnam—that it's up to Putnam to show that no candidate for reference-fixing can ever be successful. At first glance, therefore, my emphasis

[22]For another formulation of this point, see p. 341 of my earlier paper. See also pp. 405 and 407 of Bellotti's paper.

[23]For more on my own approach to these questions, see section III–V of my earlier paper. See also Timothy Bays, "Two Arguments Against Realism," (in preparation).

on Putnam's inability to *prove* certain theorems may seem to presuppose the realists' position in this dispute.

I think, however, that later glances will show that this impression is mistaken. We need to distinguish between, on the one hand, the global burdens of proof in realist/anti-realist debates and, on the other hand, the more local considerations which are relevant to assessing Putnam's model-theoretic argument. In thinking about the model-theoretic argument, we want to know what Putnam's model theory actually *shows* about realism. That is, even if we agree that realists ultimately have the burden of proof in realist/anti-realist debates, we still want to know how Putnam's model theory *contributes* to those debates.[24] This is why it's so important to be clear about what Putnam can and cannot prove and, more broadly, about the role that model theory really plays in his overall argument.[25]

At then end of the day, then, I think it's still important to focus some of our attention on the purely mathematical problems in Putnam's argument. It's not that these are the *only* problems in this argument or even that they're the *deepest* problems in this argument; it's that they're the problems which are most closely connected to the things which make this argument philosophically interesting. Given its promotional significance, we need to be clear about just how Putnam's mathematics really works (or doesn't work). Only then will we be in a position to address the larger, more purely philosophical, questions that Putnam's model-theoretic argument poses.

References

Bays, Timothy. "On Putnam and his Models." *The Journal of Philosophy* XCVIII (2001): 331–50.

Bays, Timothy. "Two Arguments Against Realism." (In Preparation).

Bellotti, Luca. "Putnam and Constructibility." *Erkenntnis* 62 (2005): 395–409.

Gaifman, Haim. "Non-Standard Models in a Broader Perspective." In *Non-Standard Models of Arithmetic and Set Theory*, edited by Ali Enayat and Roman Kossak. New York: American Mathematical Society, 2004.

Garcia-Carpintero, Manuel. "The Model-Theoretic argument: Another Turn of the Screw." *Erkenntnis* 44 (1996): 305–316.

Levin, Michael. "Putnam on Reference and Constructible Sets." *British Journal for the Philosophy of Science* 48 (1997): 55–67.

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

[24]See pp. 347–348 of my previous paper for an alternate formulation of this point.

[25]The worry, of course, is that Putnam's model theory may *simply* play a promotional role—i.e., that some kind of burden-of-proof claim is doing all of the actual work in Putnam's argument, and that the model theory is simply serving as an impressive technical side-show. For reasons of space, I won't track this worry through all the twists and turns of Putnam's own discussions of the model-theoretic argument. For a more detailed discussion of these twists and turns, see sections 2–4 of Bays, "Two Arguments Against Realism"; see also the final sections of Manuel Garcia-Carpintero, "The Model-Theoretic argument: Another Turn of the Screw," *Erkenntnis* 44 (1996): 306–316.

Velleman, Daniel. "Review of Levin, "Putnam on Reference and Constructible Sets"." *Mathematical Reviews* 98c (1998): 1364.