Putnam's Model-Theoretic Argument Reconstructed

Igor Douven*

Department of Philosophy, Utrecht University

E-mail: `igor.douven@phil.uu.nl`

May 31, 1999

Abstract

Among those addressed by Putnam's model-theoretic argument it is common opinion that the argument is invalid because question-begging. If the standard analysis of the argument is along the right lines, then what has been called the 'just more theory move' is to be held responsible for this. In the present paper, an alternative reading of Putnam's argument is offered that makes the 'just more theory move' come out perfectly legitimate, and the argument as a whole valid.

1   Introduction

Metaphysical realism has been the main subject of the critical part of Putnam's work for over twenty years now. At the heart of this doctrine, as presented by Putnam, is a thesis which we shall here call Methodological Fallibilism (MF). What it says is that *even an epistemically ideal theory may fall short of truth*, where the notion of truth involved is that of Correspondence Truth (CT).[1] Note that MF is not merely about epistemically ideal theories. The claim is that *no* empirical theory, no matter how good our evidence for it (i.e., not even if it should be epistemically ideal), is guaranteed to be true. A further characteristic of metaphysical realism is a view of semantics sometimes called Semantic Naturalism (SN). Its credo is that 'semantics is an empirical science like any other' (Devitt 1996:48). One consequence of this view is that theories of reference (theories of the correspondence that, according to CT, has to obtain between reality and a sentence for that sentence to be true) are empirical theories, and are thus to be accepted

---

[1] As spelled out, for instance, in Devitt (1991:29). Some passages in Putnam's writings suggest that he thinks MF to be a mere consequence of the realist's view on truth (cf. for instance Putnam 1978:125). But from the fact that the metaphysical realist separates truth from any epistemic notion (like verification or rational acceptability), and in that sense conceives of truth as 'radically non-epistemic' (*ibid.*), nothing follows about the (possible or necessary) truth or falsity of (ideal) empirical theories, at least not as long as we are not informed about the metaphysical realist's view on the relation between truth and evidence. That for the realist an ideally confirmed theory is not true by definition (as it is on some pragmatist conceptions of truth), and hence not necessarily true for that reason, does not mean that a realist could not, on other grounds, consistently come to the conclusion that an ideal empirical theory must be true. As it happens to be, methodological considerations—such as the underdetermination-problem (cf. Douven and Horsten 1998, Earman 1993, van Fraassen 1980)—suggest that this is not so, and thus motivate MF. But that is another matter.

or rejected on the basis of exactly the same criteria that govern theory choice in what we more traditionally think of as the natural sciences.

Over the years Putnam has advanced several arguments against metaphysical realism (see Sosa 1993 for an overview). Of these the so called model-theoretic argument is the historically first—it dates back to 1976. The argument is claimed to be a *reductio*; it allegedly derives the negation of MF—'some empirical theories are necessarily true'— from the conjunction of MF, CT and SN. Although its 'core' is simple and uncontested, the model-theoretic argument has continued to exercise the philosophical minds ever since its first presentation, for what is not so simple is to adjudicate how effective it is as an argument against realism.

Much seems to hinge here on the legitimacy of a certain crucial move in Putnam's argument. Already in his first presentation of the model-theoretic argument Putnam considered, and quickly dismissed, a possible objection he expected his opponents to make. Commentators have called Putnam's dismissal—the 'just more theory move', as it has come to be called—'completely wrongheaded' (Van Cleve 1992:349), 'foul' and 'illicit' (Hale and Wright 1997:440f), and have argued that the move is 'simply begging the question against the realist' (Devitt 1991:227).[2] In the present paper I aim to show that, at least on one construal of the model-theoretic argument, Putnam's critics are wrong; on that construal the 'just more theory move' is perfectly legitimate, and the argument as a whole valid and, if we lay to rest some (what seem to be justified) doubts concerning one of its premises, even sound. There is, as far as I can see, at most circumstantial evidence for the claim that the construal of the model-theoretic argument that will be presented captures what Putnam 'really had in mind'. However, I urge that the question to what extent it accords with Putnam's intentions, if at all, should not concern us too much, as I think that, anyhow, on the proposed reading Putnam's argument makes much better sense than on the more common way of interpreting it.

2   The Model-Theoretic Argument

The model-theoretic core of Putnam's argument is really quite simple, and clearly correct. Here it is: Let T_I be an epistemically ideal theory (i.e. a theory possessing every conceivable virtue except, perhaps, truth). Being ideal, it will, among other things, be consistent. It then follows from some basic model-theory that T_I has models of every infinite cardinality. Choose from these some model M of the same cardinality as the world, and establish a satisfaction relation SAT by mapping M 1-1 onto the world. If T_I's language is then interpreted according to SAT, T_I comes out true of the world, i.e., T_I is true-on-SAT, or TRUE(SAT).

It should be noticed that, apart from consistency, T_I's idealness plays no role in this argument.[3] Hence the argument is really about consistent theories. The conclusion is restricted, however, by two assumptions that were tacitly made, namely, first, that T_I has at least one infinite model, and secondly, that the world is infinite. But if we agree to consider in our modal reasoning about theories only infinite worlds (so we don't examine what properties theories have in worlds which already for cardinality considerations couldn't be the actual one) and if, moreover, we restrict our attention to theories with at

---

[2]The objection that Putnam's argument ultimately begs the question has also been raised by (among many others) Brueckner (1984), Hansen (1987) and Heller (1988).

[3]As was already noted in Lewis' (1984), which also contains a lucid discussion of the question why Putnam himself restricts the model-theoretic argument to ideal theories (230f).

least one infinite model (and note that all 'real' empirical theories are of this sort), then the following can be taken to be the proper conclusion of the argument above:

(1) Any consistent empirical theory is TRUE(SAT) of any world on some satisfaction relation SAT (in fact, on indefinitely many such relations).

Prima facie this conclusion appears to be quite harmless for the metaphysical realist. That any empirical theory is guaranteed to be TRUE(SAT) for some SAT does not seem to contradict the realist claim MF that no empirical theory is guaranteed to be *true*, where, according to the realist, a theory is true if it is true on the 'intended' interpretation of the language in which it is formulated. The realist acknowledges that there is an abundance of correspondence-relations between our vocabulary and reality, but he insists that only one of those is intended, only one of those is reference. So what is the point of the argument?

The question is not readily answered. I reserve my own answer for a later section, and will here present the common interpretation of how Putnam's further moves bring the above considerations to bear against the realist. On this interpretation, Putnam's point is that, roughly put, for the metaphysical realist there is no way to discriminate between truth-on-some-interpretation and truth *simpliciter*. For, Putnam (on this reading) argues, from that perspective there can be but one distinguishing feature—call it F—between the possible interpretations of a theory's language: some make the theory come out true, others make it false. So the realist has only two options in giving substance to the notion of intendedness: he can define 'intendedness' with respect to interpretations to mean being a member of the class of F-interpretations (i.e., of the class of interpretations that make the theory true) or alternatively as being a member of the class of non-F-interpretations. Since we certainly want our theories to be true, the former seems to be preferable as a definition. But then it indeed follows that any consistent empirical theory must be true, contrary to MF. (Of course if the latter definition were chosen, it would follow that every empirical theory must be false, which the realist—as, presumably, anyone else—regards to be just as absurd as the conclusion that every such theory must be true.)

In this reasoning two premises are crucial. We state these more explicitly:

(2) If there isn't some (set of) feature(s) F^*, other than F, distinguishing among all the existing correspondence-relations between the words of a theory's language and (sets of) things in the world the putative unique reference-determining correspondence-relation, then a theory is true exactly if it is TRUE(SAT) on some correspondence relation SAT.

(3) There is no such feature F^* which complies with the metaphysical realist's standards.

Both in Putnam's presentations of the model-theoretic argument and in the responses to it of his realist opponents, (3) has received most of the attention. We will see further on that there is an interpretation of the model-theoretic argument so that (3) is not amongst the argument's premises. However, (2), or something close to it, must be amongst the premises on any interpretation. At least it is hard to see how else we are to arrive at a conclusion about the truth *simpliciter* of theories from the model-theoretic point concerning the truth-on-SAT (on some SAT) of theories. Putnam himself says surprisingly

little about this premise.[4] Perhaps he takes it to be so obvious as to need no mention nor comments. I surmise that Putnam regards (2) to be a conceptual truth, at least from the point of view of the correspondence theorist, following from the conception of truth as correspondence to reality. It indeed seems that, if there is no feature F^* different from F, then TRUE(SAT) on some SAT is the only way to capture the idea of correspondence to reality—and under the presupposition it seems to *fully* capture that idea. One may still doubt whether the realist is committed to (2). However, I will here grant Putnam that (2) is a consequence of CT. Strictly speaking, then, my claim is a conditional one, viz. that good sense can be made of Putnam's argument provided (2) follows from CT and is thus legitimate as a premise in an argument against realism. This does not weaken my case too much, I think, given that (2), or some thesis equivalent to it, is also among the premises of the argument as commonly read, and that realist complaints about that are rare.

3   Just More Theory?

(3) has stunned many of Putnam's realist opponents. Typically naturalistically inclined realists endorse some version of the Causal Theory of Reference (CTR) according to which reference is constituted causally (cf. e.g. Devitt 1981, 1996, Glymour 1982). In other words, they claim that there is an acceptable feature F^*—causality—sufficient to single out reference among the welter of word/world relations. The realist answer to Putnam's model-theoretic argument is then obviously this: since there is no guarantee that SAT coincides with the causal relations constituting reference, there is no guarantee that T_I is true (instead of just TRUE(SAT)).

But Putnam had anticipated this response. According to him invoking CTR cannot block the model-theoretic argument, since CTR is 'just more theory'. His argument for this claim could be more detailed, but it is commonly read as follows: For CTR to be a reasonable candidate theory of reference in the first place, it must be compatible with the epistemically ideal theory T_I (given the constraints on T_I, CTR might even be assumed to be part of that theory). So CTR can be consistently added to T_I. But now the model-theoretic argument, being as general as it is, can simply be repeated, this time for $T_I \cup CTR$. Quite evidently—the argument on this reading continues—if we assumed that the problem the realist faces had already been solved for the part of the language in which CTR is couched, then this theory might plausibly be held to solve the problem for T_I (as well as for any other theory to which the model-theoretic argument applies), but 'how "causes" can uniquely refer is as much of a puzzle as how "cat" can, on the metaphysical realist picture'.[5] In other words, CTR is part of the problem not the solution sought. To assume otherwise is to beg the issue.

This 'just more theory move' is the most controversial part of Putnam's argument; we shall shortly discuss its legitimacy. Before doing so, however, we should briefly consider Putnam's claim that the model-theoretic argument only works against realism because SN is part of that position, i.e., that there is an effective response to the argument open to an advocate of CT and MF but explicitly not SN. This claim will be seen to give independent support to my own reconstruction of the argument as opposed to the common reading of it.

---

[4]I discuss this premise (or rather a principle closely related to it) in greater detail in my (1999). Cf. also McGowan (1996) and van Fraassen (1997a:86), (1997b).

[5]Putnam (1978:126). See also Putnam (1983:17f), (1990:85f).

The response Putnam sketches is to rely on what he calls a 'magical theory of reference'. Such a theory at bottom comes to a denial that reference is anything like a natural relation, that anything informative from a scientific perspective can be said about it.[6] So it could for instance be claimed to be a 'surd, metaphysical truth' (1981:46) that some one word/world relation is reference ('it just happens to be so'). Or some more substantive explanation of reference in terms of supernatural powers (noetic rays, intuition, direct grasping, what have you) might be tried. But how the details of such a magical theory of reference are filled in does not matter; all that matters is that SN and the standards of acceptability it imposes on theories of reference are given up.

Lewis (1984) finds Putnam's claim that magical theories of reference can pull the trick where naturalistic ones fail puzzling. For, he thinks, if the 'just more theory move' were legitimate, then it should work no matter what kind of feature F^*—whether natural or non-natural—an opponent might appeal to. As Lewis says:

> What's the good of holding up yet another sign, thus
>
> DIRECTLY GRASPS
>
> or perhaps
>
> INTUITS
>
> if it is still open to Putnam to challenge the determinate reference of the words written on the sign? ... Putnam ... has said nothing to show why the ['just more theory'] treatment applies only to natural constraints. (233)[7]

So if reference is claimed to be singled out by some supernatural power, why not add 'reference is singled out by some supernatural power', and perhaps some account of that supernatural power, to T_I and apply to that theory the same bit of model theory that was applied before to $T_I \cup CTR$? The 'unnaturalized' metaphysical realist would thereby not fare any better in the light of the model-theoretic argument than the naturalized. However, although it is true that Putnam does not say much about why he thinks his argument only affects the naturalist, we will see that, on the reconstruction of the argument to be presented, it is perfectly plausible that the distinction should be essential in Putnam's eyes.

We now turn to the evaluation of Putnam's argument, and more in particular to the evaluation of the 'just more theory move' therein. As intimated earlier, realists have protested vigorously against Putnam's dismissal of CTR as being just more theory and have claimed that the gambit is itself question-begging. Let us *pro temp* suppose that the common interpretation of Putnam's model-theoretic argument is correct, and that it correctly locates the place of the 'just more theory move' in the argument. Then it is plain how it is that Putnam begs the issue against the realist: To come from '$T_I \cup CTR$ is TRUE(SAT)' to '$T_I \cup CTR$ is true', Putnam must assume (3). But to assume (3) is to assume that CTR fails. For if CTR is true, (3) cannot be. Putnam might answer that he

---

[6]Note that, as defined here, Putnam's own (1981) 'account' of reference, according to which all that can be said about reference are (what van Fraassen 1997a, 1997b calls) pragmatic tautologies, such as that 'cat' refers to cats, counts as magical too.

[7]Cf. also Hale and Wright (1997:440), who contend that the 'just more theory move' '[i]f allowable ... will be available against any specific constraint the metaphysical realist may propose, regardless of its precise content ...'.

does *not* assume CTR to be false. Even if CTR is true—he might continue—and states a naturalistically acceptable constraint on reference, as the realist claims, I, Putnam, have just shown the constraint to be insufficient to *determine* reference. After all, we can find any number of satisfaction relations on which $T_I \cup CTR$ comes out true, can't we? But as virtually any commentator has pointed out, this would be to miss a crucial distinction, viz. that between an interpretation satisfying some *constraint C* and an interpretation satisfying a *theory stating C*. Evidently, the realist never claimed that reference is fixed by the *words* of the realist theory of reference, but by the actual reference-determining mechanism the theory is a theory *of*.[8]

What should make us suspicious about our understanding of Putnam's argument here, is that, if the standard realist analysis of the argument is along the right lines, it would make that argument so *evidently* mistaken. It would mean that Putnam simply takes (3) for granted, whereas the dialectical situation quite obviously requires him to argue for it. Indeed, the mistake would be so obvious that it is hard to imagine a philosopher as astute and logically sophisticated as Putnam committing it.

Could it be, then, that Anderson (1993) is right that the model-theoretic argument is not the argument all or most of Putnam's critics have taken it to be but that the full model-theoretic argument is an argument in two stages, namely, one stage in which Putnam argues that if (3) is the case then metaphysical realism is false, and another stage in which he seeks defend (3)? Putnam's argument for (3), on Anderson's reconstruction, is, in a nutshell, that if anything could fix reference from a naturalist perspective it is causality, but because the only notion of cause we can make sense of (physical or empirical cause) is generally overdetermined, causality won't work as a reference-fixer.

It is true that Putnam at many places in his work argues for the aforementioned claim, but, as Devitt (1997:114n) rightly remarks, this is a separate argument against the realist and not in any sense part of the model-theoretic argument. Putnam's treatment of CTR, in the model-theoretic argument, has nothing to do with the specific content of CTR;[9] it has nothing to do with whatever problems might beset the notion of cause or with any other 'intrinsic' shortcomings CTR might have, and this is so both on the standard interpretation of Putnam's argument and on mine below.[10] Another reason to doubt the correctness of Anderson's reading of the model-theoretic argument is that, although Anderson is apparently aware that the argument is intended as a *reductio* of metaphysical realism (cf. 312), on his construal it isn't (or at least not obviously so): the argument for (3), as Anderson presents it, draws on much more than is implied by the premises representing realism (whether we take these to be P1 and P2 in Anderson's paper or the premises CT, MF and SN of our exposition).[11]

Must we then, after all, conclude that the common reading of Putnam's model-theoretic argument is correct and thus that the argument rests on an elementary mistake?

---

[8]See e.g. Alston (1996:141), Devitt (1983), (1991), (1997), Field (1994), Glymour (1982), Hacking (1983), Hale and Wright (1997), Lewis (1984). Hale and Wright (1997:441) suggest, and rightly reject, a further response Putnam might make at this point.

[9]As we saw, this was also noted by Hale and Wright (1997:439ff) and Lewis (1984).

[10]That Putnam at so many places in his work extensively argues against CTR itself might be interpreted as an indication that he himself is doubtful about the 'just more theory move'—for if that works, it seems that any argument specifically directed against CTR is otiose. However, in view of Putnam's general style of philosophizing it is more plausible to assume that he seeks to show that the realist runs into trouble at every turn.

[11]Moreover, if Putnam's argument should indeed be intended as Anderson claims it is, then I think García-Carpintero (1996:305) is right in pointing out that 'the specifically model-theoretic considerations are irrelevant, and could lead to spurious debates'.

I think not. The next section offers a more charitable rendition of the argument, one on which the argument comes out perfectly valid.

4   The Argument Reconstructed

On my reading of Putnam's argument realists are right that Putnam in it assumes CTR to be false. What I would want to deny is that Putnam is not entitled to do so. Here it will be argued that there is a way of reading Putnam's argument so that the assumption that CTR is false is legitimately made. On the proposed reading the assumption that CTR is false is made in a subargument of the full model-theoretic argument, and is discharged at the end of that subargument. Thereby we obtain a derivation of ¬MF from the uncontroversial premises MF and (1) and from the somewhat less uncontroversial (2). Important is, however, that (3) is not among the premises.

Before we go on to consider the alternative reading of Putnam's argument, a few words should be said about the argument's scope. Surely Anderson is right in pointing out that according to Putnam causality is the only candidate reference-fixer from a naturalistic perspective. Moreover, some realists think likewise (see for instance Devitt 1990:83, 1996:160). But as I see it, Putnam's argument in no way depends on this kind of pessimism about naturalistically acceptable alternatives to causality as reference-fixer. In what follows, we read CTR as asserting that reference is fixed by causality *or by any other naturalistic constraint*.[12] From the ensuing reconstruction it will emerge that Putnam in his argument can even make the stronger assumption that CTR, read in this way, is false without begging any realist issues.

The basic idea underlying the reconstruction is this: Putnam seeks to establish the falsity of MF. To argue towards a *reductio* he assumes the thesis to be true. Recall now that what this thesis expresses is that every scientific theory, no matter how good our evidence for it, may be false. Assuming this to be true, it follows that CTR, being an empirical theory, may be false. It is then obviously legitimate to go on: 'Suppose it *is* false ...' (i.e., to enter a 'necessitated subproof', as it is called in some proof systems for modal logic), as Putnam does, at least on the interpretation of his argument I favor. The following natural deduction style representation of how the argument-as-I-think-of-it goes may be helpful.[13] For ease of exposition, I assume the variable x to range over all and only (actual and potential) consistent empirical theories, so that we can formalize MF as $\forall x \diamond \neg \text{True}(x)$ (i.e., the antecedent 'if x is an empirical theory'-part of MF need

---

[12] It is not clear whether CTR (however read) can be expressed in a first-order language. This might matter, since Putnam's argument, at least in its original presentation, pertains only to first-order theories. In a later presentation (1983:23), Putnam argues that the first-order/higher-order distinction isn't really essential to his argument; although for higher-order languages the argument must be different, the same result can be obtained. Putnam may be too quick on this issue (cf. Hansen 1987 for some sensible remarks on this point). However, I will here side-step that discussion and assume that CTR is a first-order theory.

[13] The reader unfamiliar with the natural deduction system employed here may consult e.g. Tennant (1990) or van Dalen (1994) for a clear presentation of it.

not be expressed in the formalization of the thesis).

$$\cfrac{\cfrac{\neg\bot}{\Box\neg\bot\,(=\neg\Diamond\bot)}\,{\scriptstyle\Box I} \qquad \cfrac{\cfrac{[\forall x\Diamond\neg\mathrm{True}(x)\,(=\mathrm{MF})]^1}{\Diamond\neg\mathrm{True}(\mathrm{CTR})}\,{\scriptstyle\forall E} \qquad \cfrac{[\neg\mathrm{True}(\mathrm{CTR})]^2 \quad \cfrac{[\neg\mathrm{True}(\mathrm{CTR})]^2 \qquad \cfrac{\mathcal{D}}{\mathrm{True}(\mathrm{CTR})}}{\bot}\,{\scriptstyle\neg E}}{\Diamond\bot}\,{\scriptstyle{}_2\Diamond E}}{\bot}\,{\scriptstyle\neg E}}{\neg\mathrm{MF}}\,{\scriptstyle{}_1\,\neg I}$$

This is only a first approximation of course. But it does suffice to show that, by reflecting on what MF says, we might be able to get a sound derivation of the negation of MF from MF and from whatever premises may be hidden inside $\mathcal{D}$.

In a necessitated subproof we allow, apart from the subproof's hypothesis, only hypotheses which are necessarily true in the world in which all the argument's premises are true. We so to speak enter a possible world[14] in which a certain hypothesis—in our case $\neg\mathrm{True}(\mathrm{CTR})$—holds, in which, of course, all tautologies hold, and, most importantly, in which everything is true that according to the argument's premises is true in any possible world.[15] (1) evidently expresses a necessary truth; it can be formalized thus: $\Box\forall x\exists\mathsf{SAT}[\mathsf{TRUE}(\mathsf{SAT})(x)]$. (2) was granted to be a conceptual truth, at least from the perspective of the realist. If it is, then, from that perspective, it is also necessary, and can be formalized as: $\Box((3) \to \forall x(\exists\mathsf{SAT}[\mathsf{TRUE}(\mathsf{SAT})(x)] \leftrightarrow \mathrm{True}(x)))$. Finally, it must be noted that (3) is equivalent to the assumption of CTR's falsity (CTR understood in the 'broad' sense defined earlier).[16] Thus, $\neg\mathrm{True}(\mathrm{CTR}) \leftrightarrow (3)$ is a tautology.

We are now ready to state the subproof:

_____

[14] This should not be taken too literally, if only because the assumption that there is a possible world of a certain kind may lead to a contradiction so that there can be nothing we might have entered in our subproof.

[15] So the $\Diamond E$-rule can alternatively be stated as follows:

$$\cfrac{\Box\Gamma,\Diamond\phi \qquad \cfrac{[\phi]^1 \quad [\gamma_1]^2 \quad \cdots \quad [\gamma_n]^{n+1}}{\vdots \atop \psi}}{\Diamond\psi}\,{\scriptstyle 1,\ldots,n+1}$$

where $\Box\Gamma$ abbreviates $\Box\gamma_1,\ldots,\Box\gamma_n$. It is easily seen that for $\Diamond E$ to be sound no requirements regarding the accessibility-relation have to be made, which means that it can serve as a rule in a natural deduction presentation of any modal logic. This also holds true for the other modal rule involved in the reconstruction.

[16] The same holds, from Devitt's viewpoint, for CTR understood narrowly. According to Devitt, causality is the only naturalistically acceptable candidate reference-fixer, so if CTR (narrowly understood) is false, (3) must be true, and *vice versa*.

$$
\cfrac{
  \cfrac{
    \cfrac{
      {}^{1}[\neg\mathrm{True}(\mathrm{CTR})] \quad
      \cfrac{\neg\mathrm{True}(\mathrm{CTR}) \leftrightarrow (3)}{\neg\mathrm{True}(\mathrm{CTR}) \to (3)}{\scriptstyle \wedge E}
    }{(3)}{\scriptstyle \to E} \quad
    (3) \to \forall x(\exists \mathsf{SAT}[\mathsf{TRUE}(\mathsf{SAT})(x)] \leftrightarrow \mathrm{True}(x))
  }{\forall x(\exists \mathsf{SAT}[\mathsf{TRUE}(\mathsf{SAT})(x)] \leftrightarrow \mathrm{True}(x))}{\scriptstyle \to E}
$$

$$
\cfrac{
  \cfrac{
    \cfrac{\forall x \exists \mathsf{SAT}[\mathsf{TRUE}(\mathsf{SAT})(x)]}{\exists \mathsf{SAT}[\mathsf{TRUE}(\mathsf{SAT})(\mathrm{CTR})]}{\scriptstyle \forall E} \quad
    \cfrac{\cfrac{\forall x(\exists \mathsf{SAT}[\mathsf{TRUE}(\mathsf{SAT})(x)] \leftrightarrow \mathrm{True}(x))}{\forall x(\exists \mathsf{SAT}[\mathsf{TRUE}(\mathsf{SAT})(x)] \to \mathrm{True}(x))}{\scriptstyle \wedge E}}{\exists \mathsf{SAT}[\mathsf{TRUE}(\mathsf{SAT})(\mathrm{CTR})] \to \mathrm{True}(\mathrm{CTR})}{\scriptstyle \forall E}
  }{\mathrm{True}(\mathrm{CTR})}{\scriptstyle \to E} \quad {}^{1}[\neg\mathrm{True}(\mathrm{CTR})]
}{
  \cfrac{\Diamond\neg\mathrm{True}(\mathrm{CTR}) \qquad \cfrac{}{\bot}{\scriptstyle \neg E}}{\Diamond\bot}{\scriptstyle {}_{1}\Diamond E}
}
$$

What we get when we fit this into the earlier schema is a refutation of metaphysical realism which only depends on premises acceptable to the realist. In particular, (3) is not among the argument's premises, as promised. Furthermore, the reconstruction immediately suggests an answer to Lewis' puzzle why Putnam thinks supernatural constraints are exempt from the 'just more theory treatment'. Consider the non-naturalist realist who endorses some magical, non-scientific, theory of reference MTR. From the fact that he avows MF it doesn't follow that he is committed to 'possibly, MTR is false', for MF is a thesis about *scientific* theories; it says nothing about magical theories. Hence, an account of some supernatural power, or even just the claim that reference is fixed by supernatural means, although a theory in the strict sense of that word (it *is* a set of sentences of course), is not just more theory in the sense which makes it susceptible to Putnam's model-theoretic considerations—it is not just more *empirical* theory. It is indeed the naturalist's insistence that semantics, properly pursued, is a branch of natural science that plays havoc on him; if a theory of reference is to be an empirical theory, then it is automatically within the scope of MF.

Thus the reading of Putnam's argument offered in this paper shows the argument to be valid, contrary to what most if not all realists believe, and would in fact rightly believe if the common reading were the only one possible. As a further advantage, the alternative reading solves Lewis' puzzle why Putnam thinks his argument only holds against the semantic naturalist. If my reading should not be the 'intended one', then, I think, the former constitute good reasons why Putnam *should* have intended it.

References

Alston, W. (1996) *A Realist Conception of Truth*, Ithaca/London.

Anderson, D. (1993) 'What is the Model-Theoretic Argument?', *Journal of Philosophy* 90:311–322.

Brueckner, A. (1984) 'Putnam's Model-Theoretic Argument against Metaphysical Realism', *Analysis* 44:134–140.

Devitt, M. (1981) *Designation*, New York.

Devitt, M. (1983) 'Realism and the Renegade Putnam', *Noûs* 17:291–301.

Devitt, M. (1990) 'Meanings just ain't in the Head', in: G. Boolos (ed.) *Meaning and Method*, Cambridge, 79–104.

Devitt, M. (1991) *Realism and Truth*, Oxford.

Devitt, M. (1996) *Coming to Our Senses*, Cambridge.

Devitt, M. (1997) 'On Determining Reference', in: A. Burri (ed.) *Language and Thought*, New York, 112–121.

Douven, I. (1999) 'A Note on Global Descriptivism and Putnam's Model-Theoretic Argument', *Australasian Journal of Philosophy*, in press.

Douven, I. and L. Horsten (1998) 'Earman on Underdetermination and Empirical Indistinguishability', *Erkenntnis* 49:303–320.

Earman, J. (1993) 'Underdetermination, Realism and Reason', in: P. French, T. Uehling and H. Wettstein (eds) *Philosophy of Science* (Midwest Studies in Philosophy XVIII), Notre Dame, 19–38.

Field, H. (1994) 'Are Our Logical and Mathematical Concepts Highly Indeterminate?', in: P. French, T. Uehling and H. Wettstein (eds) *Philosophical Naturalism* (Midwest Studies in Philosophy XIX), Notre Dame, 391–429.

García-Carpintero, M. (1996) 'The Model-Theoretic Argument: Another Turn of the Screw', *Erkenntnis* 44:305–316.

Glymour, C. (1982) 'Conceptual Scheming, or Confessions of a Metaphysical Realist', *Synthese* 51:169–180.

Hacking, I. (1983) *Representing and Intervening*, Cambridge.

Hale, B. and C. Wright (1997) 'Putnam's Model-Theoretic Argument against Metaphysical Realism', in: B. Hale and C. Wright (eds) *A Companion to the Philosophy of Language*, Oxford, 427–457.

Hansen, C. (1987) 'Putnam's Indeterminacy Argument: The Skolemization of Absolutely Everything', *Philosophical Studies* 51:77–99.

Heller, M. (1988) 'Putnam, Reference and Realism', in: P. French, T. Uehling and H. Wettstein (eds) *Realism and Antirealism* (Midwest Studies in Philosophy XII), Notre Dame, 113–128.

Lewis, D. (1984) 'Putnam's Paradox', *Australasian Journal of Philosophy* 52:221–236.

McGowan, M.-K. (1996) *Realism or Non-Realism: Undecidable in Theory, Decidable in Practice*, Doctoral Dissertation, Princeton University.

Putnam, H. (1978) *Meaning and the Moral Sciences*, London.

Putnam, H. (1981) *Reason, Truth and History*, Cambridge.

Putnam, H. (1983) *Realism and Reason*, Cambridge.

Putnam, H. (1989) 'Model Theory and the 'Factuality' of Semantics', in: A. George (ed.) *Reflections on Chomsky*, Oxford, 213–232.

Putnam, H. (1990) *Realism with a Human Face*, Cambridge MA.

Sosa, E. (1993) 'Putnam's Pragmatic Realism', *Journal of Philosophy* 90:605–626.

Tennant, N. (1990) *Natural Logic*, Edinburgh.

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

van Dalen, D. (1994) *Logic and Structure*, New York.

van Fraassen, B. (1980) *The Scientific Image*, Oxford.

van Fraassen, B. (1997a) 'Elgin on Lewis's Putnam's Paradox', *Journal of Philosophy* 94:85–93.

van Fraassen, B. (1997b) 'Putnam's Paradox: Metaphysics Revamped and Evaded', in: J. Tomberlin (ed.) *Philosophical Perspectives* Vol. 11, Boston, 17–42.