Chapter 6
The Metaphysics of the Model-Theoretic Arguments

Kate Hodesdon

Abstract This paper presents a exposition of Putnam's model theoretic arguments in the context of his broader philosophical position. I argue that Putnam used the arguments not just to undermine metaphysical realism, but to reveal that the philosophical debate between metaphysical realism and internal realism is dialectically problematic in that the metaphysical realist defence cannot "count against" (Putnam in Philosophical Topics: The philosophy of Hilary Putnam 20(1):355, 1992c) the converse position. Putnam's response is that this is a debate that we should simply undercut.

Putnam's model theoretic arguments have posed challenges of interpretation since their publication. In this article I shall make two claims about the arguments that add to this debate. One is to clarify the arguments' target: while it is clear that the arguments are designed to refute the position of metaphysical realism, it is less clear just which hypothesis is at stake. I present a thesis that Putnam takes to be constitutive of metaphysical realism and targets with the model-theoretic arguments. This is the posit of *epistemic humility*, which states that it is possible that there is an aspect of the world that is epistemically inaccessible as a matter of principle. The second aim of this paper is to suggest a new direction in which to seek justification for the notorious 'just more theory' response that Putnam gives to critics of the model-theoretic arguments.

Metaphysical or "external" realism and Putnam's own internal realism are broad positions, comprising two "philosophical temperaments" (1981, p. 49) or "tendencies" (1980, p. 474). They are also foundational: Putnam hints that their consequences affect almost every area of philosophy (1981, p. 49), particularly scepticism ("the question of 'Brains in a Vat' would not be of interest if it were not for the sharp way in which it brings out the difference between these philosophical perspectives"*; Ibid.*) The distinction between the two positions is inspired by Kant, with internal realism representing Kant's own position (1992d, p. 114, 1987, pp. 36–37), although the relationship between Kant and Putnam's views is complicated. Metaphysical real-

K. Hodesdon (✉)
Department of Philosophy, University of Bristol, Bristol, UK
e-mail: kate@hodesdon.com

G. Hellman and R. T. Cook (eds.), *Hilary Putnam on Logic and Mathematics*,
Outstanding Contributions to Logic 9, https://doi.org/10.1007/978-3-319-96274-0_6

ism, on the other hand, is a traditional position in philosophy. Putnam defines it as committed to the following three broad claims about the world and the way that we refer to it:

> [T]he world consists of some fixed totality of mind-independent objects. There is exactly one true and complete description of 'the way the world is'. Truth involves some sort of correspondence relation between words or thought-signs and external things and sets of things. I shall call this perspective the externalist perspective, because its favorite point of view is a God's Eye point of view. (Putnam 1981, p. 49)[1]

The first claim to which metaphysical realism is committed is that *ontology is* "*fixed once and for all*" (1989, p. 352), independently of what our theories of the world might be, or indeed what our conceptualisation of the world might be (1982). The members of this ontology are philosophically privileged: to talk of the world in terms of them is to speak of things as they are *in themselves* (Putnam 1981, p. 50, 1995a, p. 303).

According to the second claim, there is just one true and complete theory of the world. Of course, humans speak a number of different, but inter-translatable, languages, so the uniqueness of this one theory must be due to the concepts that it employs relative to a language, rather than the particular words it uses. Specifically, the uniquely true theory of the world is the one that describes the fixed totality of objects singled out in the first claim. As Putnam explains, if there is just one true theory of the world, then there is automatically one privileged ontology: the one in terms of which the theory is given.[2] This is the non-perspectival ontology of things in themselves.

> One true theory requires a *ready-made* world—the world itself has to have a "built-in" structure since otherwise theories with different structures might correctly "copy" the world (from different perspectives) and truth would lose its Absolute (non-perspectival) character. (Putnam 1982, p. 147).

The third posit of metaphysical realism tells us what it means for a theory to be true: it must *correspond correctly with the world*. The nature of the correspondence relation is not specified, but it is intended to be *unique*, and to yield a bivalent semantics (1989, p. 352, 1991, p. 110). As an example of a reference relation, Putnam typically uses the thesis that reference is causally mediated. Putnam has argued specifically

---

[1] While Putnam's characterization of his own position, 'internal realism', shifted during the period that he endorsed it, Putnam used 'metaphysical realism' to capture more or less the same thesis throughout. See, for instance, (1989, p. 352, 1999, p. 18, n. 41).

[2] The claim is held by at least one contemporary metaphysician. In his recent book, *Writing the Book of the World*, Ted Sider argues for precisely this metaphysical realist thesis. He claims that not only is there one true account of the world as it is in itself, but that there is precisely one correct language for properly describing it.

> In order to perfectly describe the world, it is not enough to speak truly. One must also use the right concepts—including the right logical concepts. One must use concepts that 'carve at the joints', that give the world's *structure*. There is an objectively correct way to 'write the book of the world'.

against causal theories of reference (1982, 1989, pp. 358–360, 1992b, pp. 61–79), but in the model-theoretic arguments, a causal chain theory functions as a toy example of a physical constraint on reference: what goes for a correspondence based on causal chains goes equally for a correspondence based on any other relation. A correspondence theory of truth rules out other norms of rational inquiry, such as verifiability, which was a central part of Putnam's own position, internal realism, until he abandoned it in 1990.

Together, these three posits of metaphysical realism entail that there is precisely one theory (up to notational variance) which must correspond in the correct way, to the correct objects in order to be true.

> For the externalist philosopher [. . .] the truth of a theory does not consist in its fitting the world as the world presents itself to some observer or observers (truth is not 'relational' in this sense), but in its corresponding to the world as it is *in itself*. (Putnam 1981, p. 50, emphasis added)

The three posits also entail a deep and seemingly inescapable problem for the metaphysical realist. The problem is that *our* theories of the world are unavoidably perspectival. The describe objects that are informed by what Putnam calls our "epistemological position" in the world, which is determined both by our sensory experiences and by the conceptual schemata we employ. But the "fixed totality" of things in themselves, on the other hand, may be epistemically inaccessible to us. Consequently, it may be the case that even our best theory of the world—and Putnam is careful to spell out exactly what we mean by 'best theory' (1980, p. 473)—fails to describe the privileged totality of objects, and represents only appearances. In this case, our best theory of the world is false.

6.1 Epistemic Humility

We have seen that metaphysical realism amounts to a traditional epistemological duality thesis that posits an underlying noumenal world of things in themselves, which may possibly be very different from the appearances that our theories describe.[3] In this way, the epistemic duality thesis is coupled with a form of *epistemic humility*, which says that it is possible that as a matter of principle, we know nothing about a certain aspect of reality. Indeed, Putnam tells us that "the sharp distinction between what really is the case and what one judges to be the case is *precisely what constitutes metaphysical realism*" (Putnam 1981, p. 50, emphasis added). Elsewhere he writes:

> The realist—or, at least, the hard-core metaphysical realist— [. . .] wishes it to be the case that what, e.g., electrons are should be distinct (and possibly different from) what we believe

[3]The duality is also presented by Putnam's differentiation in "Brains in a Vat" between "vats *themselves"* and"*vats-in-the-image*" (1981, Ch. 1).

> them to be or even what we would believe them to be given the best experiments and the epistemically best theory. (Putnam 1980, p. 472)

It is not sufficient for epistemic humility just that we are, as a matter of fact, ignorant of some aspect of the world. After all, irrespective of any stance on things in themselves, we are almost certainly ignorant about the distant past, and remote regions of space. Epistemic humility says that this ignorance is *as a matter of principle*, which is to say that even in the most epistemically ideal conditions we could remain ignorant of the aspect of reality in question.

The metaphysical realist's epistemic duality between things as they are for us and things in themselves, together with humility about the latter, is a traditional thesis. It is tacitly assumed by many philosophers, but its most explicit defence comes from Kant. There are (at least) two ways of understanding the distinction in Kant: as demarcating between *two kinds of object* in terms of how we know them, or between *two ways of knowing* any given object. Pippin has called the first of these interpretations the "two worlds" or "two realms" view (1982, p. 196 ff.), and it is this that Putnam seems to have in mind for the metaphysical realist's predicament. Another place that epistemic humility finds expression—this time postdating Putnam's arguments against it—is in the work of Lewis (2009). Lewis argued, on grounds entirely different from Kant's, for an epistemic humility thesis regarding the identities of fundamental *properties*. Lewis named the position Ramseyan Humility, after its Ramseyan justification. Specifically, Ramsey argued that scientific theories describe only the nomological structure of the world and the roles of the fundamental properties within it. However, there are multiple ways in which the properties may realize the roles introduced by a theory. And quidditism, which Lewis also defends (2009, §4), says that a permutation of properties between their theoretical roles yields a distinct, but qualitatively identical, possibility. A permutation of two fundamental properties is the result of replacing all occurrences in space and time of one of them with those of the other, and vice versa.[4] From Ramsey's thesis and quidditism, Ramseyan humility follows: since we cannot distinguish between the distinct possibilities that result from such permutations, we are ignorant of the identities of the properties that may be permuted.[5]

However, there is one significant respect in which metaphysical realism differs from most other forms of epistemic humility, most notably the reading of Kant that Langton defends as 'Kantian humility' (1998). This is the centrality of scepticism. Putnam has often remarked that metaphysical realism can be *characterized* by its tolerance of scepticism (1980, p. 473, 1981, Ch. 1, Ch. 3, 1989, pp. 352–354). But, at least according to Langton, Kantian humility does not lead to scepticism: "it allows plenty of knowledge of the real world, but denies knowledge of things as they are in themselves." (2004, p. 134) Likewise, Lewis's easy acceptance of the predicament

[4]Of course, the properties must be both of the same logical "category" in order to be permuted—both monadic, or both relational magnitudes, for instance—but Lewis thinks that there are such multiply-instantiated categories of properties.

[5]Since Lewis does not believe haecceitism, which is the thesis analogous to quidditism concerning individuals instead of properties, his argument does not establish ignorance of things.

Ramseyan humility places us in indicates that, for him, it falls short of scepticism: "Who ever promised me that I was capable in principle of knowing everything?" (2009, p. 211).

6.2 The Model-Theoretic Picture

The epistemic duality of metaphysical realism makes it possible for Putnam to use model theory to give an analogy for the metaphysical realists thesis, and then, by carrying over results from model theory to the metaphysical realist account of the world, to undermine it. This is the job of the model-theoretic arguments.

In a model, we have a domain of objects, $\mathcal{D}$, an object language, $\mathcal{L}$, and an assignment function, π, that assigns terms of $\mathcal{L}$ to objects and sets of objects of $\mathcal{D}$. Models in this sense are the basis of the branch of mathematical logic known as model theory, but models in the more general sense, including physical models, like replicas, are constructed in a similar way. What is crucial for the analogy with metaphysical realism is that an interpreted model does not only introduce a correspondence, π, between names and objects; it also determines a correspondence between *two domains of objects*. This is by virtue of the fact that its object language, $\mathcal{L}$, is already at least partially interpreted: it includes terms that are meaningful and which, in the metalanguage, refer to objects and sets of objects that are generally independent of the model. If the object language did not have such an interpretation—that is, if it were just meaningless symbols—then the model could not serve its purpose of *representing* the particular state of affairs described in $\mathcal{L}$ that it was designed to portray. As Putnam has said, a model is only a model in so far as it represents *something* by somebody (1981, p. 5). For an example of the two domains in practice, consider a set-theoretic model of arithmetic. The model's assignment function will induce a correspondence between the *sets* in the domain of the model (say, the von Neumann ordinals) and the *natural numbers* described in the language of arithmetic.

Thus, an interpreted model picks out a correspondence between two discrete sets of objects: the domain elements (which are doing the representing) and those named by the object language (which are being represented). In some cases, these two domains will be structurally similar. For instance, scale models preserve the same relative distances between domain elements and the objects they represent. On the other hand, a Rutherfordian model of the atom made of plastic beads will almost certainly misrepresent the relative distances between subatomic particles, but will preserve their relative *positions* as inside or outside the nucleus. The presence of structural similarities like these are what make a model *apt* as a representation. However, although there may be such structural similarities,[6] in general the domain

[6] I do not want to claim that *all* representation trades on shared structure between the represented objects and the representing objects (domain elements). In particular, the Löwenheim–Skolem theorems provide plausible counterexamples.

elements of a model will have quite different properties from those of the objects being represented. To use Putnam's example from "Models and Reality", in a model of set theory a set that is non-constructible from the perspective external to the model may well represent a constructible set in the model.

This two-level ontology associated with a model is analogous to the metaphysical realist's "two realms" of things for us and things in themselves. The analogy justifies Putnam's application of theorems from model theory that show that there is no unique way to model a theory, to the "models" whose domain elements are the things in themselves, and whose objects are things for us. Just as there is no unique way to model a theory by pairing up domain elements with the objects they represent, neither is there a unique mapping from things for us onto things in themselves. But, just such a mapping is required by the single unique reference relation posited by the metaphysical realist's correspondence theory of truth. Therefore, by analogy, we can infer that there is no such unique reference relation to things in themselves. In this way the model-theoretic arguments establish the conditional thesis that, "assuming a world of mind-independent, discourse-independent entities [. . .] there are [. . .] many different 'correspondences' which represent possible or candidate reference relations (infinitely many, in fact, if there are infinitely many things in the universe)." (Putnam 1981, p. 47) Consequently, and as I shall claim, the model-theoretic arguments lead Putnam to reject the notion of an epistemic duality as well as the notion of things in themselves.

6.2.1 Reductio ad Absurdum

We have seen that model theory provides a suitable language for talking, by analogy, about metaphysical realism. Let us now turn to the model-theoretic arguments in closer detail to see how model theory undermines the position.

The arguments are intended to be a *reductio ad absurdum* of metaphysical realism (Putnam 1993, pp. 280–281, 1995a, p. 303). But what is the absurd conclusion that they establish? Perhaps the most obvious candidate is just what the permutation argument (1981, Ch. 2) establishes: that *reference is radically underdetermined*. Although this may be the most straightforward form for a *reductio* to take, the reading doesn't sit well with the characterization of the metaphysical realist just given. To really appreciate the problem with metaphysical realism, we shall have to turn to the other model-theoretic argument: the 'Skolemite' argument.

There are two reasons why the *reductio* of metaphysical realism requires more than simply showing that it incurs radical indeterminacy of reference. The first problem is that in order to derive referential indeterminacy from metaphysical realist posits *by contradiction,* the metaphysical realist must be implicitly or explicitly committed to the converse: not necessarily that reference *is* determinate, but at least that it is not radically underdetermined. However, as we have seen, the metaphysical realist already lives with the threat of scepticism, and so believes that it is possible that her

theories are radically false of the world. Therefore, while the metaphysical realist might *hope* that reference is fixed via causal chains, or something similar, the fixity of reference cannot be an assumption of her position, since it is inconsistent with her scepticism about knowledge. For all the metaphysical realist knows, she might be a brain in a vat. If she were a brain in a vat, she would still believe that causal chains link her talk of trees with trees *themselves*, even when, in reality causal chains link her talk only to trees "in the image". If the model-theoretic arguments show that the metaphysical realist has no guarantee that her theories correctly refer, and consequently, that she has no guarantee that they are true of reality, then this just supplies her with *one more* sceptical hypothesis. It is simply more grist for her mill.

There is a second problem with understanding Putnam's *reductio* of metaphysical realism to have even the *logical form* of a proof by contradiction. The problem is how to reconcile a proof by contradiction with Putnam's frequent claims that the model-theoretic arguments reveal metaphysical realism to be *incoherent*, or *unintelligible* (Putnam 1978, p. 126, 1980, p. 474, 1992a, p. 85, n. p. 173, 1992c, p. 355). Generally, a *reductio* only licenses us to conclude that at least one of the assumed premises is false. We could interpret the claim of metaphysical realism's incoherence as meaning only that its posits cannot all be true together. But there are other forms of absurdity that are not outright logical contradictions. And, as Putnam later claimed, his charge was never that metaphysical realism is logically inconsistent. Instead, he says it is incoherent.

> What is consistent or not is a matter of pure logic; what is coherent, or intelligible, or makes sense to us, and what is incoherent, or unintelligible, or empty, is something to be determined not by logic but by philosophical argument. (Putnam 1989, p. 354)

For the remainder of this paper, I will focus attention on Putnam's objection that metaphysical realism is incoherent in the sense that it is "empty" (1995a, p. 303):

> [M]etaphysical realism cannot even be intelligibly stated [. . .] attempts at clear formulation never succeed in capturing the content of 'metaphysical realism' because there is no real content there to be captured. (1992c, p. 353)

The emptiness claim is justified by the model-theoretic arguments—specifically, by the Skolemite argument of "Models and Reality", together with the 'just more theory' rejoinder. The argument rests on what is taken to be an extension of Skolem's historical 'paradox', that all first-order formal theories, including those like the theory of real analysis that we take to deal with non-denumerably many objects, have denumerable models. While this theorem is no longer considered paradoxical, it does indicate something about the inability of such theories to pin down (even the cardinality of) their models. Putnam's Skolemite argument shows that something similar is true of the theory consisting of our total science:

> [E]ven a *formalization of total science* (if one could construct such a thing), or even a *formalization of all our beliefs* (whether they count as "science" or not), could not rule out [. . .] *unintended* interpretations (Putnam 1980, p. 466; see also Putnam 1989, p. 353)

The Skolemite argument requires the notion of an epistemically ideal theory, T_I. This is a theory (together with an interpretation) that meets operational and theoretical constraints, defined as follows.

Operational constraints constrain the theory from contradicting facts that can be confirmed observationally. The constraints are imposed by the stipulation that, first, the theory T_I be given in a language containing an observational vocabulary sufficient to name every one of the countably many things or events we could possibly observe. (1980, p. 472) Second, T_I is given a partial interpretation $\mathcal{OP}$ such that, for all terms $\bar{t}$ in the language of T_I that denote observable things and events, and for all predicate and relational symbols R in the language of T_I that denote observable properties or relations, if the object denoted by $\bar{t}$ *can be observed to have* the property denoted by R then the valuation $\mathcal{OP}$ makes $R\left(\bar{t}\right)$ true. This guarantees that T_I "does not lead to any false predictions" (1980, p. 473) about observable events. In sum, operational constraints ensure the following conditions:

> [I]f 'there is a cow in front of me at such-and-such a time' belongs to T_I, then 'there is a cow in front of me at such-and-such a time' will certainly *seem* to be true—it will be 'exactly as if' there were a cow in front of me at that time. [. . .]
>
> On the other hand, if 'there is a cow in front of me at such-and-such a time' is *operationally* 'false' (falsified) then 'there is a cow in front of me at such-and-such a time' is [false in the model]. (Putnam 1978, p. 126)

Theoretical constraints are extra-empirical constraints. They ensure that T_I possesses all epistemic virtues that would make it rational for scientists to accept T_I in the limit of human inquiry. Of course, these virtues include *consistency* (*Ibid.* p. 473)—along with "simplicity, elegance, subjective plausibility" (1989, p. 35). It is important that these virtues capture what is "*epistemically* ideal for *humans*" (1980, p. 472); although they are relativized to an ideal limit of inquiry, they pin down a theory that *we* would accept as best, given our epistemological position.

Operational and theoretical constraints are the best yardstick that we can reliably use to measure a theory's success. A theory that meets them will make no predictions that are falsified by what we can observe, since, by virtue of the operational constraints, all atomic sentences describing observable things and events will be theorems of the theory. But the constraints do not guarantee that a theory is true, in the sense of the correspondence theory of truth. As Putnam explains, the Skolemite argument was intended to put pressure on the notion that a theory could meet operational and theoretical constraints yet fail to be true:

> Example: an ideal theory might say that there are intelligent extraterrestrials somewhere in space-time, although in fact there aren't any. There might be overwhelming evidence that there are intelligent extraterrestrials (somewhere, some time), evidence for laws according to which the probability that such never did, don't, and never will exist is less than one in a trillion, let us say (which would certainly justify believing that intelligent extraterrestrials exist in spacetime), when, in fact, ours is a universe in which the one in a trillion chance that they don't exist is realized. This is an example of the way in which "correspondence truth" can differ from even idealized verifiability. *The purpose of the model-theoretic argument was to cast doubt on the very intelligibility of this very plausible set of beliefs.* (Putnam 2012, p. 75, emphasis added)

We are now in a position to state Putnam's Skolemite argument. Let T_I be an epistemically ideal theory.

6.3 The Skolemite Argument

P1. It is possible that T_I is *false* of reality (1980, p. 473). This is just the metaphysical realist's thesis of epistemic humility.
P2. Let T_I be false. This assumption is justified by **P1**.
P3. T_I has models by the completeness theorem. (*Ibid.*)
P4. Let $\mathcal{M}$ be a model of T_I whose domain contains the countably-many macroscopically observable things and events, and whose interpretation agrees with the partial interpretation $\mathcal{OP}$.
P5. A theory is *true* if it is true in the intended model. (*Ibid.* p. 474)
P6. Since T_I meets *all* the constraints that we can impose on a theory, the model $\mathcal{M}$ is an "intended" model.
C1. T_I is true. (*Ibid.* p. 474)
C2. Since **C1** contradicts **P2**, **P2** must be false. But **P2** follows from **P1**, so **P1** must be false. This contradicts epistemic humility..

P1 is true by definition of metaphysical realism, and **P2** follows from it. **P3** is also uncontroversial, as a theorem of first-order logic. **P4** can also be justified without much trouble, since it asserts the existence of a model that we can directly construct. However, the justification for the remaining premises has been the subject of much debate—for a sample, see Douven (1999), Bays (2001, 2008), Hale and Wright (1997).

P5 implies that truth can be equated with truth in a (specific) model. This equivalence is supported by the analogy discussed earlier: that the metaphysical realist's world picture of a uniquely privileged ontology, one true theory about it, and one correspondence relation making the theory true can be thought of as a single model.

P6 appears to equivocate on the notion of "intended", and so gives the metaphysical realist some leeway to reject the argument's conclusion. While the metaphysical realist will insist that the previously-described model is the one she "intends", **P6** asserts that $\mathcal{M}$, which is an arbitrary model of the epistemically ideal theory T_I, is in fact intended. The reason why, for Putnam, $\mathcal{M}$ *is* an intended model is that the theory it makes true satisfies operational and theoretical constraints. He asks, rhetorically, "what else could single out a model as 'intended' than this?" (*Ibid.* p. 473). However, for the metaphysical realist, the model $\mathcal{M}$ has to be unintended because it satisfies T_I. For, T_I is false. The point of contention therefore comes down to rival theories of truth: to the question whether a theory which is as epistemically ideal as we can possibly measure may be false, by virtue of its failure to describe some epistemically inaccessible part of reality.

We might well wonder why Putnam believes that operational and theoretical constraints on truth yield the only possible ways to determine reference. As Button (2013, §4.3) has convincingly shown, it is clear that Putnam considers these to be the limit of *naturalistic* constraints on reference. Button draws attention to Putnam's repeated characterisation of any constraints that go beyond them as "magical"—in other words, nonnatural. And while Putnam may have a reputation for changing his mind on central topics, he has *always* been a scientific realist (Putnam 2012, p. 52ff.),

so rules nonnatural methods of reference-fixing out of the question: "the suggestion that metaphysical realism might be *nonempirically* true is a possibility I did not—and still do not—take seriously." (1995a, p. 304).

Putnam's own positive account of reference (1975) on the other hand firmly links successful reference with our own, human practice of inquiry. To use his example, whether or not a substance can be referred to as 'gold' depends on whether it has properties privileged by *our* theory of chemical composition. Indeed, as Putnam wryly remarked: "Kripke expressed dissatisfaction with "The Meaning of Meaning" precisely on the ground that the notion of the "essence" of a natural kind I employ there is *not* independent of scientific practice" (1992c, p. 349). This, then, is why $\mathcal{M}$ is an intended model: because it makes true a theory that satisifies the limit of constraints that we can put on truth without contradicting naturalism.

6.4 Just More Theory

If Putnam is right that the metaphysical realist cannot supply a naturalistic account of reference that supports her view of truth, then his argument raises a dilemma for the metaphysical realist. Either she must abandon metaphysical realism altogether for antirealism about truth, or else bolster her realism with a nonnatural account of intentionality (1980, pp. 474–475). There is a commonly-raised objection to the first disjunct. The objection proceeds by simply stating that reference is fixed by some relation in particular, typically by causation. According to this objection, *contra* the Skolemite argument, $\mathcal{M}$ is not intended at all unless its assignment function picks out this causal relation.

The 'just more theory' reply is the response that Putnam gives to interlocutors who make this objection. The reply claims that since the causal theory of reference is part of our overall best theory of the world, the occurrences of 'cause' and other terms in the theory are subject to Skolemite reinterpretation. Thus, the model-theoretic arguments cannot be dismissed using a causal theory of reference. And the same goes for any other reference-fixing relation. As stated, it might seem that the 'just more theory' move is an objection to any reference fixing constraint whatsoever. However, the point is only supposed to be aimed at reference-fixing constraints supplied by the metaphysical realist. In particular, the ability to pick out one reference relation is incompatible with the metaphysical realist's epistemic humility. This is suggested by Putnam's rebuke that in proposing some referencing fixing constraint "the philosopher is *ignoring his own epistemological position*" (1983, p. xi, emphasis added). Recall that the metaphysical realist's epistemological position is perspectival. She cannot rule out the hypothesis that the way that the world is really—in terms of things in themselves—is radically different from her best theory of the world.

Before continuing to examine the 'just more theory' move, let us look briefly at the second disjunct of the metaphysical realist's dilemma: that she opt for a non-natural theory of reference. Lewis (1984, pp. 232–233) asked why Putnam offered

the metaphysical realist this way out, given that doing so missed an opportunity to generalize his argument fully. We have already seen that Putnam's naturalism means that he would not take non-natural constraints on reference seriously; to him, a non-natural constraint on reference does not consitute a legitimate way out of the dilemma at all. However, why permit it as an option for the metaphysical realist at all? Douven (1999, p. 490) answered this question with the claim that only naturalistic theories are vulnerable to the 'just more theory' rejoinder. We saw that the just more theory move requires us to suppose that the theory of any reference-fixing constraint offered is false. But according to Douven, the metaphysical realist is only committed to fallibilism for naturalistic theories. So, if the metaphysical realist held a theory of a non-natural referential constraint, she could simply reject the just more theory move when directed at this theory.

Button gives an alternative explanation for the restriction of the model-theoretic arguments to naturalistic theories of reference. He argues that, given metaphysical realism, the only constraints actually capable of fixing reference—and thus being more than mere theory—are non-empirical (2013, p. 31). So, the 'just more theory' move can *only* apply to empirical accounts of reference, in other words, naturalistic ones. The reason for this is that the metaphysical realist endorses what Button calls a Cartesian Principle, which says that "even an ideal theory might be radically false" (p. 10).

> [T]he external realist must accept that her attempts to constrain reference are without empirical content. Whatever her view of empirical content, her Cartesian Principle sets up a sceptical veil between herself and the world, and between her words and the world" (Button 2013, p. 53, see also p. 58)

Button adds to the model-theoretic arguments the premise that, according to metaphysical realism, beliefs have narrow content. In other words, the belief that I am seeing a cat requires a certain kind of cat-like sense data, but not necessarily any cat itself. In fact, the idea can be generalized beyond merely sensory data: the metaphysical realist has a full system of "constructions" in terms of which her theories of the world are given, that are over and above the objects themselves. In this way, Button's metaphysical realist posits something much like the epistemic duality I have described here. When the metaphysical realist makes a claim that goes beyond her constructed ontology, then she is talking about an "unconstructed world, made up of objects that are largely mind-, language- and theory-independent" (p. 37).

Where I disagree with Button on Putnam is his claim that "the empirical content of any claim is exhaustively accounted for *within* the construction system itself" (*Ibid*.) Button uses this thesis to justify the just more theory move: if all claims with empirical content can be given solely in terms of the constructions, then simply by virtue of being an empirical claim, any theory about reference will be 'just more theory' in so far as it talks only about the constructions, and not about anything beyond the veil. Certainly, we can imagine a dual ontology comprised of, on the one hand, objects in terms of which *all empirical theories* can be given, and on the other hand, whatever objects are left. Just think, as Button suggests, of Carnap's distinction between 'internal' and 'external' questions. But I am not convinced that

the metaphysical realist's dual ontology *is* like this: her ontology of constructions, or appearances, contains (roughly) things like cat-like sensory impressions or objects based on appearances of cats, but her ontology of things in themselves is supposed to be made up of cats, vats, and so on—all very much empirical things. This is, of course, unless her sceptical hypothesis is true, and we are being radically deceived—and in this case, things in themselves are indeed *who knows what* and constructions are al we have to go on. In short, I don't see how to justify the idea that empirical claims are exhausted by appearances or constructions, *without* assuming that the metaphysical realist's sceptical hypothesis is correct. While Button's exegesis of the internal realist and metaphysical realist positions is convincing, his account of the 'just more theory' move appears to assume that the metaphysical realist is *in fact* veiled off from the reality beneath appearances—which is just what the model-theoretic arguments were supposed to establish.

But if the metaphysical realist's belief that *it is possible that* she is epistemically isolated from reality doesn't justify the 'just more theory' move, what does? And what should we make of Putnam's remark, quoted earlier, that the move is justified by the metaphysical realist's epistemological position? I want to suggest that we can justify the 'just more theory' manoeuvre on the basis simply of the metaphysical realist's model-theoretic picture of the world and our theories' relationship to it, without any assumption regarding whether or not the metaphysical realist is in fact epistemically isolated from the world. But first I will flesh out the 'just more theory' move a little more. My interpretation here is guided by an account that Putnam has given in just some of his later discussions of the model-theoretic arguments, most fully in a 1992 special edition of *Philosophical Topics*, in which he responded to his critics (1992c, see also Putnam 1995a). There Putnam urged that the problem with metaphysical realism was a problem of *demarcation*: the metaphysical realist cannot articulate a theory that rules out the antirealist position that equates truth with satisfaction of operational and theoretical constraints. Putnam explained that,

> The main point [of "Models and Reality"] was that metaphysical realism cannot even be intelligibly stated. I expressed this by saying that metaphysical realism is 'incoherent'. I did not mean by that it is inconsistent in a deductive logical sense, but rather that when we try to make the very vague claims of the metaphysical realist precise, we find that they become compatible with strong forms of 'antirealism'. (Putnam 1992c, p. 353, see also 1995a, p. 303)

In this remark, Putnam cannot mean that metaphysical realism and antirealism are compatible in the sense of being jointly consistent, since clearly these theories say different things—about what makes a sentence true, for example. Their compatibility, as Putnam explains it, is due to the fact that the metaphysical realist's theory about how reference gets fixed, and consequently how sentences are made true, can *itself* be made true in the antirealist sense of truth. This is because the core posits of metaphysical realism, given at the start of this paper, satisfy operational and theoretical constraints.

This is true even of claims to which the metaphysical realist is committed that are directly contradicted by antirealism, such as the claim that reference is fixed by causal connections. As long as the claim that 'causation fixes reference' does not

contradict anything observable, and thus does not violate operational constraints, nor violate any theoretical constraints, the antirealist may accept it. It is precisely because operational and theoretical constraints couple truth to appearances that the antirealist can accept that causation fixes reference. Putnam continues,

> Granting that one can *say* [. . .] that reference is fixed by some physical relation, say, 'causal connection' [. . .], the question is whether these statements (assuming they make sense) express what the metaphysical realist 'wants to say'. After all, the claim that 'reference is fixed by causal connection' will, if true, be satisfied by all intended models. It will satisfy operational and theoretical constraints. Its truth does not, by itself, count against [the antirealist] conception of truth" (Putnam 1992c, p. 355)

Putnam's remark that the metaphysical realist's claim can be made true in a model, without *counting against* the theory external to the model, recalls the demonstration at the beginning of "Models and Reality" that V = L can be made true in a model of set theory even when "in reality" V = L is false. The remark also puts the 'just more theory' manoeuvre in a somewhat different context. It emphasizes a similarity between the treatment of the metaphysical realist's claims by her interlocutor and the way that a formal theory is treated model-theoretically, which is to say, with reference that is only fixed up to isomorphism. This suggests that the problem lies with the model-theoretic nature of the metaphysical realist relationship between a successful theory and reality. We have seen that the analogy between metaphysical realism and the way that models represent theories makes the Skolemite argument possible (it justifies **P5** in particular). It is this model-theoretic picture that permits the metaphysical realist's interlocutor to treat her words as 'mere theory' with no uniquely privileged correspondence to any particular objects or relations. If it is merely the existence of this strong similarity that justifies the 'just more theory' move, then there is no need to assume that the metaphysical realist's theories do not, in fact, correspond with things in themselves: we can remain agnostic on this matter.

Putnam's remark that "the question is whether [causal theories of reference] express what the metaphysical realist 'wants to say"' suggests that Putnam's goal is not to show that metaphysical realism is false,[7] but rather that it cannot *get an edge* on antirealism—cannot "count against it". Quite simply, when the metaphysical realist makes a claim, she can't guarantee that her interlocutor, who holds a different theory of truth, will interpret her claim as she intends. Describing this dialectic, we could say that if two opposing sides in a debate don't agree on the notion of truth then there is no hope in either side getting one up on the other. With this understanding of the justification of the 'just more theory' reply, the antirealist's treatment of metaphysical realism—simply reinterpreting its claims—is indicative that the dispute between metaphysical realist and antirealist, who each hold different theories of truth, is seriously fraught. It is not the kind of dispute that can give way to a mutual resolution.

[7] Putnam says as much in (1995a): "clear attempts at a formulation of [metaphysical realism] never succeed—because there is no real content to be captured. My aim [. . .] therefore, was not to argue for the truth of a counter-thesis (one which could be identified with the negation of metaphysical realism but rather simply to provide a reductio ad absurdum of metaphysical realism by teasing out the consequences of its own presuppositions." (p. 303).

This observation has also been made by Maximilian de Gaynesford, who, in a talk at a conference at Harvard in 2010 for Putnam's 85th birthday, drew attention to Putnam's use of the term "antinomy" to describe the model-theoretic arguments. De Gaynesford presented what he called the "antinomy picture" of the model-theoretic arguments, according to which, whichever party begins the argument invariably wins it; neither is outright victorious, and neither succeeds in convincing the other.

As I understand it, Putnam's attempted resolution of the model-theoretic arguments was simply to undercut this dispute. This is seen in the origins of Putnam's own position, internal realism. When Putnam first introduced 'internal realism', in (1978), it was the name of a position intended as a form of scientific realism that could be endorsed by metaphysical realists and antirealists about truth alike (1992c, p. 352, see also 2012, pp. 55–56). It was an empirical theory (1978, p. 130) about how scientific theories referred. The position said nothing about what truth *was*, and thus could be seen as a kind of Carnapian 'internal question', circumventing the debate between metaphysical realism and antirealism. Later on, when Putnam published his own verificationist position on truth, distinct from the account based on operational and theoretical constraints (Putnam 1992c, p. 353), he found that his readers took that position to be the one named 'internal realism', and decided that "it seemed easiest to me to go along with this, as I did in *Reason, Truth and History*" (*Ibid.*).

While Putnam may have initially intended to undercut the metaphysical realist-antirealist debate by staying neutral on the topic of truth, he then moved into this debate with his defence of verificationism. Given this, it is worth identifying how Putnam's own position avoided the pitfalls of the model-theoretic arguments. Putnam's strategy for avoiding the problem himself is, in part, to deny that an epistemically ideal theory might be false, in favour of the alternative verificationist thesis that truth coincides with epistemic ideality, at least in the limit of inquiry. But there is another significant difference between Putnam's position and the metaphysical realist's: internal realism does away with the model-theoretic picture of truth by denying any kind of dichotomy between things for us and things in themselves. And, as we have seen, a dichotomy between things in themselves and things for us is essential to the model-theoretic arguments. So, whereas a distinction between appearances and things in themselves was "precisely what constitutes metaphysical realism" (Putnam 1981, p. 50), Putnam tells us that, *"[T]he adoption of internal realism is the renunciation* of the notion of the 'thing in itself'. (1987, p. 36, emphasis added). He continued:

> Internal realism says that the notion of a 'thing in itself' makes no sense; and *not* because 'we cannot know the things in themselves'. [. . .] Internal realism says that we don't know what we are talking about when we talk about 'things in themselves'. (Putnam 1987, p. 36)

6.5 The Renunciation of the Notion of the 'Thing in Itself'

Putnam defended the claim that there can be no such thing as a thing in itself throughout his internal realist period (1995a, p. 302, 1978, p. 6, 1987, p. 41, 1982, p. 163), and continued to do so even after abandoning internal realism. It is a goal of his 2005 book, *Ethics Without Ontology*, to refute the project of "describing the world as it is 'in itself"' (2005, p. 24). In this final section, I'l highlight some areas of Putnam's broader philosophical thought to which this thesis is foundational.

One place where Putnam's rejection of the dichotomy between things for us and things in themselves is particularly apparent is in his critique of ontological relativity. Ontological relativity is the thesis that "there is no absolute sense in speaking of the ontology of a theory" (1969, p. 48). It was established by Quine's *Gavagai* argument, which aimed to show that there is no possible answer to the question whether a given object is a rabbit, versus an undetached rabbit-part, and so on, since no difference in behaviour can be detected between the case in which 'rabbit' refers to one or the other.

Putnam's disagreement with Quine consists in fact that, whereas Quine would be "willing to put up with the slack" (*Ibid*. p. 45) that the *Gavagai* argument reveals between language and world, Putnam is already convinced that there is no slack. If there were, then we would have an epistemic duality of the kind that the model-theoretic arguments refute. As Putnam sees it, Quine just accepts that there is a world of objects out there, and we cannot know to which one 'rabbit' refers. Quine thus retains a realm of epistemically inaccessible objects just like those posited in the metaphysical realist's sceptical hypotheses. Consequently, as for the metaphysical realist, only a magical theory of reference would allow us to refer to these objects: "it is magical, in Quine's view, to think that science can do more than fix the structure of the world up to isomorphism" (Putnam 1989, n. 20) Ultimately, its commitment to things in themselves makes ontological relativity as untenable as metaphysical realism:

> What am I to make of the notion of an X which is a table *or* a cat *or* a black hole (or the number three *or*. . .)? An object which has *no* properties at all in itself and any property you like 'in a model' is an inconceivable *Ding an sich*. The doctrine of ontological relativity avoids the problems of medieval philosophy (the problems of classical realism) but it takes on the problems of Kantian metaphysics in their place. (Putnam 1983, p. xiii)

Putnam therefore takes the argument to show that an "alternative" is needed to the entire metaphysical picture it presupposed. Quine's *modus ponens* is Putnam's *modus tollens* (Putnam 1993, p. 280).

The untenability of ontological relativity is a surprising consequence of Putnam's attack on metaphysical realism, given the extent to which Quine's work in this area influenced Putnam's. Indeed, the permutation argument was devised during what Putnam described as a period of "intense interaction" (1978, p. ix) with Quine's views, and introduced as extending Quine's work in *Ontological Relativity* "in a very strong way" (1981, p. 34, see also 1992a, p. 112). And, as Putnam says, it is true that both the permutation argument and the *Gavagai* argument were designed to refute the

thesis that "words stand in some sort of one-one relation to (discourse-independent) things and sets of things." (1981, p. 41) Putnam's rejection of ontological relativity is also surprising, given his defence of a quasi-structuralist position about mathematics (1967), in which he argued that there are multiple ontologies for mathematical objects, each equally good, but suited for different purposes. This sounds remarkably like ontological relativity restricted to mathematics.[8]

Internal realism's "renunciation of the notion of the'thing in itself"' is also ostensibly hard to reconcile with Putnam's repeated claims that Kant *himself* was an internal realist (1981, p. 60, 1987, p. 43), given that Kant famously posited the existence of a noumenal domain. But this is not the interpretation of Kant that Putnam endorses. During his internal realist period Putnam defended an interpretation of Kant's transcendental realism according to which there is no bijection between the noumenal and phenomenal realms (1981, p. 61), which is to say, no correspondence between things in themselves and things for us. In fact, Putnam went on to claim that "almost all of the *Critique of Pure Reason* is compatible with a reading in which one is not at all committed to a Noumenal World, or even [. . .] to the intelligibility of thoughts about noumena (1987, p. 41). In light of this reading of Kant, the Kantian heritage of Putnam's thought is much clearer.

To conclude, while the model-theoretic arguments defy easy characterisation, they do provide a refutation of the metaphysical thesis of an epistemic duality, a concept aptly illustrated by the ontologies involved in modeling. Moreover, I believe that we can find justification for the 'just more theory' manoeuvre—the most troubling step in the arguments for many of their critics. The manouvre is justified when made against the metaphysical realist by someone already committed to the view that truth is satisfaction of operational and theoretical constraints. However, this justification also reveals the problematic nature of the debate between metaphysical realism and rival metaphysical theories.

References

Bays, T. (2001). On Putnam and his models. *The Journal of Philosophy*, *98*(7), 331–350.
Bays, T. (2008). Two arguments against realism. *The Philosophical Quarterly*, *58*, 193–213.
Braddon-Mitchell, D., & Nola, R. (Eds.). (2009). *Conceptual analysis and philosophical naturalism*. Cambridge: MIT Press.
Button, T. (2013). *The limits of realism*. Oxford: Oxford University Press.
Douven, I. (1999). Putnam's model-theoretic argument reconstructed. *Journal of Philosophy*, *96*(9), 479–490.
Gunderson, K. (Ed.). (1975). *Language, Mind, and Knowledge. Minnesota Studies in the Philosophy of Science, Vol. VII*. Minneapolis: University of Minnesota Press.
Hale, B., & Wright, C. (1997). Putnam's model-theoretic argument against metaphysical realism, pp. 427–457.
Hodesdon, K. (2014). Mathematical representation: Playing a role. *Philosophical Studies*, *168*(3), 769–782.

[8] An example that Quine also considered (1969, pp. 58–60).

Hodesdon, K. (forthcoming). Structure, symmetry and semantic glue. *Philosophy in an age of science: Themes from the philosophy of Hilary Putnam, Berger A*. Oxford: Oxford University Press.
Langton, R. (1998). *Kantian humility*. Oxford: Oxford University Press.
Langton, R. (2004). Elusive knowledge of things in themselves. *Australasian Journal of Philosophy*, *82*(1), 129–136.
Lewis, D. (2009). Ramseyan humility. *See Braddon-Mitchell and Nola*, pp. 203–223.
Lewis, D. K. (1984). Putnam's paradox. *Australasian Journal of Philosophy*, *62*(3), 221–236.
Pippin, R. (1982). Kant's theory of form. *An Essay on the Critique of Pure Reason, New Haven, London*.
Putnam, H. (1967). Mathematics without foundations. *The Journal of Philosophy*, *64*(1), 5–22.
Putnam, H. (1975). The meaning of meaning. See Gunderson, pp. 131–193.
Putnam, H. (1978). *Meaning and the moral sciences*. Boston: Routledge and Kegan Paul Ltd.
Putnam, H. (1980). Models and reality. *The Journal of Symbolic Logic*, *45*(3), 464–482.
Putnam, H. (1981). *Reason, truth, and history*. Cambridge: Cambridge University Press.
Putnam, H. (1982). Why there isn't a ready-made world. *Synthese*, *51*(2), 141–167.
Putnam, H. (1983). *Philosophical papers: Volume 3, realism and reason*. Cambridge: Cambridge University Press.
Putnam, H. (1987). *The many faces of realism: The Paul Carus lectures*. La Salle: Open Court.
Putnam, H. (1989). Model theory and the 'factuality' of semantics. See Putnam (1995b), pp. 351–375.
Putnam, H. (1991). *Representation and reality*. Cambridge: MIT Press.
Putnam, H. (1992a). *Realism with a human face*. Cambridge: Harvard University Press.
Putnam, H. (1992b). *Renewing philosophy*. Cambridge: Harvard University Press.
Putnam, H. (1992c). Replies. *Philosophical Topics: The philosophy of Hilary Putnam*, *20*(1), 347–408.
Putnam, H. (1992d). Why is a philosopher? See Putnam (1992a), pp. 105–109.
Putnam, H. (1993). Realism without absolutes. See Putnam (1995b), pp. 279–294.
Putnam, H. (1995a). The question of realism. See Putnam (1995b), pp. 295–312.
Putnam, H. (1995b). *Words and life*. Cambridge: Harvard University Press.
Putnam, H. (1999). *The Threefold cord: Mind, body and world, the John Dewey essays in philosophy*. Irvington: Columbia University Press.
Putnam, H. (2005). *Ethics without ontology*. Cambridge: Harvard University Press.
Putnam, H. (2012). In M. Di Caro, & D. Macarthur (Eds.), *Philosophy in an age of science*. Cambridge: Harvard University Press.
Quine, W. (1969). *Ontological relativity and other essays*. New York: Columbia University Press.

Kate Hodesdon is research explores how it is possible to pick out and refer to mathematical objects. She is interested in what mathematical logic itself, particularly model theory and set theory, tells us about what mathematics is all about.