

## **Internal to what? Contemporary Naturalism and Putnam's Model Theoretic Argument**

In Reason, Truth, and History (RTH), Hilary Putnam deploys the Model-Theoretic Argument (MTA) against metaphysical realism, arguing for an "internalist" alternative and applying his thoughts to a wide range of philosophical problems. This paper examines some of Putnam's ideas from the point of view of contemporary naturalism. Naturalism shares with internalism some central elements, such as rejection of a God's Eye Point of View, yet there are deep methodological differences. Here, I discuss some of these differences through consideration of various matters such as theories of reference and truth, the existence of mathematical objects, and brain-in-a-vat type skepticism. I argue that though the internalist and naturalist share an interest in "our" methods, one point of divergence is over what this comes to, with particular disagreement over the question of whether the methods we associate with science have special epistemic status. Toward the end I explore some practical implications of this difference.

### 1. Internalism and naturalism

In a (1993) paper, Putnam expresses his pleasure that "most" of the readers of RTH have correctly interpreted its argument as a *reductio*. The position being reduced to absurdity is, of course, metaphysical realism. On the metaphysical realist perspective, "the 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 between words or thought-signs and external things and sets of things" (RTH, p. 49). Putnam dubs this perspective "the externalist perspective," because "its favorite point of view is a God's Eye point of view" (RTH, p. 49).

One of Putnam's strategies for reducing metaphysical realism to absurdity is through the model-theoretic argument. That argument assumes that there is a reference relation R attaching words to things, and draws the conclusion that even while keeping the truth-values of all of our sentences fixed, there are infinitely many relations that R could be. From this, Putnam says, it follows that "there are always infinitely many different interpretations of the predicates of a language which assign the 'correct' truth-values to the sentences in all possible worlds, no matter how those 'correct' truth-values are singled out" -- a conclusion, he suggests, that we cannot accept" (RTH, p. 35).

This problem with reference purportedly dooms the correspondence theory of truth, because there is no way to fix a particular relation between our sentences, thoughts, and theories on the one hand and the world of mind-independent objects on the other. In response to the idea that truth might involve some sort of abstract mapping of concepts onto things, Putnam says, "The trouble with this suggestion is not that correspondence between words or concepts and other entities don't exist, but that too many correspondences exist. To pick out just one ... we would have already to have referential access to mind-independent things" (RTH, pp. 72-73).

But his solution is not a turn to idealism or relativism. One of Putnam's central claims in the book is that to assume a rejection of metaphysical realism leads one into forms of anti-realism is a mistake -- this is a false dilemma. The way out is instead to reject metaphysical realism in favor of internalism. Internalism holds that "what objects does the world consist of is a question that it only makes sense to ask within a theory or description" (RTH, p. 49). "'Truth,' in an internalist view, is some sort of (idealized) rational acceptability -- some sort of ideal coherence of our beliefs with each other and with our experiences *as those experiences are*

*themselves represented in our belief system* -- and not correspondence with mind-independent or discourse-independent 'states of affairs'" (RTH, p. 50). Putnam emphasizes that the kind of rationality central to internalism is a part of our conception of "human flourishing": because the fact-value distinction is untenable, we must evaluate our beliefs according to a mix of criteria relying on various kinds of values.

The suggestion -- though this is not made quite explicit in RTH and I return to it below -- is that from within the internalist view, the absurdity of the model-theoretic conclusion does not arise: from within our theory of the world and from within our language, we can say that 'cat' refers to cats and not to cherries, because we use 'cat' in sentences like 'cats rule the internet.'" In the rest of RTH, Putnam extends his discussion beyond reference and truth to a wide range of philosophical problems.

The next few sections of this paper consider how various aspects of Putnam's argument look from the point of view of some strands of contemporary naturalism. Several factors prompt such an investigation. Some contemporary naturalists take up themes closely related to Putnam's, even while arriving at different conclusions. For example, with respect to commonalities, naturalists often share Putnam's doubts about metaphysics, single true descriptions, and a God's Eye point of view on the world. Penelope Maddy's (2007) *Second Philosophy* emphasizes the bottom-up nature of inquiry, endorsing the use of scientific methods from a range of disciplines to make headway on philosophical problems. Mark Wilson (2006) emphasizes the multiplicity and complexity of scientific descriptions, pointing out that general conclusions about language, concepts, and correspondences to reality are never simple and straightforward; he and Maddy both point out there may be multiple equally good ways of describing the world (see, e. g.,

Maddy 2007, p. 106). John Burgess (2004) rejects Realism-with-a-capital-R, which he associates with Ultimate Metaphysical Reality and "what God was saying to Himself when He was creating the universe" in favor of what he calls "anti-anti-realism;" Maddy, too, proposes distinguishing Thin Realism from Robust Realism in the context of mathematical truth (2005; 2011a).

Of course, these naturalists also diverge from Putnam in important ways. Maddy takes pains to show that meaningful talk about how words refer to things does not require the kind of unifying "theories" of reference or truth that Putnam is concerned about (2007, Part II). Wilson explains that the multiplicity and complexity of scientific descriptions do not, and should not, prevent us from analyzing how our language and concepts do and do not connect up with the world (2006; see also 2000). Burgess suggests that the form of "realism" we ought to adopt "amounts to little more than a willingness to repeat in one's philosophical moments what one says in one's scientific moments, not taking it back, explaining it away, or otherwise apologizing for it" (2005, p. 19) -- a realism that privileges science, not the rationality associated with human flourishing. And Maddy wholeheartedly rejects the association of justification with idealized rational acceptability, partly on the familiar ground that "a theory's being true is different from its being reasonable for us to believe it" (2007, p. 100).

To keep this investigation manageable, in this paper I will focus on Maddy's Second Philosophy -- comparing its approach and conclusions to those of Putnam, exploring relations to some of Putnam's later work, diagnosing differences and discussing them. As an approach, Second Philosophy is focused on epistemological and methodological questions; though conclusions about existence may follow, it is not, in first instance, a doctrine about ontology or the nature of reality. Second Philosophy does not look to philosophy to provide epistemological

support for scientific methods, but rather begins inquiry from within the scientific methods we already use and trust. As Maddy (2011, p. 121) puts it:

"The Second Philosopher is actually a quite mundane and familiar figure. She begins her investigations of the world with perception and common sense, gradually refines her observations, devises experiments, formulates and tests theories, always striving to improve her beliefs and her methods as she goes along; at some points in her investigation of the world, she addresses (her versions of) traditional philosophical questions; and the result is Second Philosophy."

Whether encountering questions about traditional philosophical problems, questions about scientific practice, or questions about other areas of inquiry, the Second Philosopher adopts the same approach: attempting to discern whether there is good evidence for the claims in question.

## 2. Reference and truth

First I consider truth and reference. As is well-known, some of the immediate response to Putnam's arguments centered on the possibility of appeal to causation -- and causal theories of reference -- as a way of pinning down the relationship between words and the things they refer to. Roughly speaking, in a causal theory of reference, 'cat' refers to cats because there is a causal chain, of some appropriate type, linking our use of the word 'cat' to actual cats. If this is right, it would allow us to explain why 'cat' refers always and only to cats and never to cherries. One of Putnam's responses to this line of thought was to claim that this strategy cannot achieve its ends, because appeal to a causal theory must be articulated in words, which are themselves subject to reinterpretation: when we try to articulate the idea that there is an appropriate causal chain, we must make use of the word "cause;" but according to Putnam's arguments, this word can be multiply interpreted: in particular, we may interpret 'cause' as referring to cause\*, where 'cat' is

linked by a causal\* chain to cherries and not cats (Lewis 1984; see also Devitt 1983, Douven (1999) and Devitt 1997).

David Lewis and others challenged Putnam on grounds that this this response misinterpreted the causal theorists' proposal: the idea was not to say that 'cat' and cats were linked by the referent of our word 'cause'; rather the idea was that they are linked by actual causes -- whether we know about these and are able to articulate the connection or not (Lewis 1984). Michael Devitt (1997), endorsing this line of thought, points out that the causal theorist's response to further questions about how causation is fixed should always appeal back to causation itself, "until the cows come home." (1997, p. 114). Putnam has suggested this is begging the question in favor of metaphysical realism, since the response appeals to actual causes as distinguished from the referent of the word 'cause.' At least one commenter has described the situation as "an impasse" (Sosa 1993, p. 606).

But one of the most striking developments in philosophical thinking since the writing of RTH has been the rise of the disquotational views of truth and reference, and the relative waning of interest in causal theories. Such a move is common among self-professed naturalists including Quine, Maddy, Burgess, and Hartry Field. This prompts us to ask to what extent and how naturalistic forms of disquotationalism diverge from or challenge Putnam's own understanding of reference and truth. As an entry into this question, let's take a look first at how the Second Philosopher understands truth and reference. In *Second Philosophy*, Maddy shows that the Second Philosopher will, like Putnam, eschew anything like a magical theory. For her, the question of reference and truth is not focused on the metaphysical realist project of attaching words systematically to things to find One True Description of the World; instead, the second

philosopher approaches questions of reference and truth just as she approaches any other question: by examining evidence and assessing what set of beliefs best explains the phenomena.

In one attempt at this project, Maddy argues that a disquotational theory of reference and truth, in which truth and reference are characterized by T- and R-sentences ("snow is white' is true iff snow is white" etc., and "'cat' refers to cats", etc.) represents an apt default position: from this default, we would investigate whether richer theories, such as causal theories, are needed to explain our practices. In an investigation of this latter issue, Maddy argues that although the T- and R-sentences are indeed all that is needed to characterize truth and reference, this is compatible with local investigations into particular terms and the worldly connections to which they refer. For example, while "'phlogiston' refers to phlogiston" follows immediately from the T- and R-sentences, a naturalistic inquiry would reveal that in the past, some uses of the term "phlogiston" correlated with certain worldly conditions -- e. g., a belief that "this air is dephlogisticated" would correlate with the air being rich in oxygen) (2007, pp. 155-156). The worldly conditions involved in these correlations are called "indication relations"; investigations into indication relations are necessary to explain important phenomena, e. g., why dephlogisticated air was useful for those with difficulty breathing. Explanations of successful reasoning can reveal information about indication relations and how they function.

As Field says about indication relations, "It is a fact about me that I am a pretty good barometer of whether there is rain falling on my head at that moment: when there is rain falling on my head, I tend to believe "There is rain falling on my head"; conversely, when I do believe this sentence, usually there is rain falling on my head. This is simply a correlation, there to be

observed; and a [disquotationalist] is as free to take note of it as is anyone else, and as free as anyone else to deem it an ingredient of what he calls content" (1994, p. 254).

From this point of view, we don't need metaphysical theories of reference and truth, or unified theories such as "causal theories," in order to meaningfully investigate the relationships that certain bits of language bear to certain bits of the world. Instead, truth and reference are taken care of by the disquotation schemas, and we use ordinary tools of observation and reasoning to give reasons for beliefs that language is functioning as it is in certain contexts.<sup>1</sup> Of course, nothing in this approach suggests there would be One True Description; as Maddy emphasizes, there are often multiple apt ways of describing the same thing (2007, p. 106). Neither is a unified "theory" of reference and truth proposed; instead, semantic relationships are analyzed in a local piecemeal way (see also Wilson 2006).

Let's investigate what happens if we try to apply this Second Philosophical approach to disquotationalism to some specifics of the MTA. We conclude first that 'cat' refers to cats and 'cherry' to cherries simply follows from the T- and R-schemas, and then see there is a very high degree of correlation between our sentences involving 'cat' and states of affairs involving cats and "cherry" and states of affairs involving cherries. These correlations will allow us to draw inferences about the indication relations of these sentences; these inferences can be drawn from

---

<sup>1</sup> Though he expresses the results in less deflationary terms Mark Wilson also stresses the importance of investigating the way words connect up to the world. To explain the success of our methods, he says, we must take up the "correlational point of view," asking when, and in what ways, our theories connect up with the world. He imagines bad guys from the "land of gavagai" who, weary of MTV and the Home Shopping Channel, decide to shoot down our communications satellites. Their success doing so prompts us to ask about their methods: "how do these linguistic moves *correspond* to the worldly conditions against which the calculations proceed?" (Wilson 2000, p. 373)

explanations of facts like "There are fewer rats in the barn" and "The sailors didn't get scurvy," where 'rat' denotes rats, 'scurvy' denotes scurvy, and so on.

In RTH, Putnam does discuss something like disquotationalism: he describes the "equivalence principle" associated with the T-schema as "philosophically neutral," adding that "On *any* theory of truth, 'snow is white' is equivalent to "'snow is white' is true" (p. 129).

Contemporary disquotationalists are likely to agree: indeed, the modesty Putnam refers to is part of what motivates the Second Philosopher's idea that it is an appropriate default, from which we ask whether it fails to explain things we need to explain. But if that is so, we may wonder: why isn't the right response to the model-theoretic argument simply to point out that of course 'cat' refers to cats -- this is part of what reference is?

It might seem the answer to this question is that in context, the response is a non-starter, since the metaphysical realist cannot endorse this answer. But this is not at all obvious. Why couldn't a metaphysical realist insist that reference and truth are given by the schemas, and then give separate analyses of metaphysics, belief, etc. that buttress metaphysical realism? The real answer to why the T- and R-schemas are of no use here is, I believe, is more subtle, and has to do with assumptions about what the aims of a theory of reference and truth are. Such an answer is hinted at in Putnam's claim that the "neutrality" of the equivalence principle is the very difficulty, since "... the problem is not that we don't understand 'snow is white,' the problem is that we don't understand what it is to understand 'snow is white.'" This is the philosophical problem. About this [the T-schema] says nothing" (RTH, p. 129).

This remark suggests that when Putnam aims for an analysis of reference and truth, he is understanding the scope of the problem in a particular way: that the answers to questions about

truth and reference must illuminate, in some way, broader epistemological issues. This suggestion is borne out in Putnam's response to Field's remarks on RTH at the time. There, Putnam says that the idea that "a realist can also be a disquotationalist" is "the most substantive -- but also the most amazing -- idea in Field's paper" -- on grounds that "the problem of truth reappears when we ask what it is for an assertion to be correct and what it is for it to be incorrect" (1982, pp. 575-576). Putnam understands questions of reference and truth to be intimately connected to questions of justification and assertibility -- so intimately connected, indeed, that an account of reference and truth that is silent on these matters conveys, in its very silence, a form of skepticism about them. But many contemporary disquotationalists do not understand the scope this way, taking pains to separate out analysis of truth and reference from analysis of broader epistemological issues -- just as the Second Philosopher separates out truth and reference relations from indication relations and other matters.<sup>2</sup>

Disagreements about the nature of truth and reference that are partly disagreements about the way concepts of truth and reference are related to other concepts are not uncommon: elsewhere I have used the idea of the "scope" of reference and truth to diagnose these disagreements more specifically (Marino 2006 and 2010). As I interpret him, when Putnam inquires into an analysis of the reference of words like 'cat,' the reason the disquotational "'cat' refers to cats" is not a solution to his problem is that he wants not merely any explanation of the relationship between the thing and the word, but also an explanation that would shed light on epistemological questions having to do with assertibility and justification, particularly those questions bearing on the matter of whether, and how, certain words and sentences manage to

---

<sup>2</sup> Leeds (2007) also argues a disquotationalist can be a realist.

*represent* things in the world. I've used the term "scope" disquotationalism to indicate the view that disquotationalism does not entail any particular conclusions about such epistemological questions precisely because they are epistemological problems. For scope disquotationalists, the R- and T-schemas tell us all we need to know about reference and truth, and questions about understanding, assertibility, and justifications are part of another area of inquiry. Likewise, as we saw in the discussion of indication relations above, analysis of how words and sentences connect up to things in the real world is possible and fruitful: it just doesn't tell us about reference or truth *per se*.<sup>3</sup> For the scope deflationist, of course disquotationalism is compatible with realism: reference and truth are disquotational, while realism is a doctrine in epistemology and metaphysics.

One reason appreciation of this difference matters is that when we encounter the model-theoretic argument, we'd like to know more about the kinds of realism it is a *reductio* of. Putnam's definition of metaphysical realism incorporates a few elements that make it a very narrow definition, inviting straw-man charges: surely many who would style themselves "realists" would nonetheless shy away from claims about One True Description and a God's Eye View of the world. If what I have said about scope is correct, we cannot adjudicate any difference in view between an internalist and any other -ist with respect to any substantive issues, including issues concerning representation and its relation to realism, by looking narrowly at what they say about truth and reference themselves: a broad investigation into epistemological matters is essential.

---

<sup>3</sup> Maddy (2007, p. 138) puts it this way: "Word–world connections aren't eliminated, their description just isn't to be found under the heading of truth or reference; we might say they turn up in the local epistemology."

This applies as well, of course, to any analysis of the difference between internalism and Second Philosophy. To understand their differences with respect to the questions of representation, we should look not just at their claims about reference and truth but more generally at how they analyze the relations between language and the world. We've seen already some of the Second Philosopher's views on such things, which involve using and refining the empirical methods we have, doing local and piecemeal analysis, and using the concept of "indication relations" to explicate word-world connections that go beyond the R- and T-schemas. How does the internalist understand the proper analysis of relations between language and the world?

It is striking that RTH has so little direct discussion of these matters. In more recent writings, Putnam has tried to address them, describing in a 1993 paper on Quine and ontological relativity how *he* thinks the absurd conclusion of the model-theoretic argument is properly avoided. Though Putnam's broader views between the time of writing RTH and 1993 changed from "internalism" to "direct realism," he glosses his discussion of reference in 1993 as a refinement of his ideas in RTH. In that later paper, he appeals to the idea that "meaning is use," distinguishing two versions of this. The first is associated with what he calls "socio-functionalism": the idea is that the use of a word, together with facts about the brain and facts about the environment, would enable drawing conclusions about what a word means. The second, "naive" interpretation, is associated with Wittengstein, Strawson, Austin, and James; here we simply acknowledge that "the use of words in a language game cannot, in general, be described without employing the vocabulary of that very game" (1993, p, 182). From within our

language, we state facts about the relations between words and things in a way that avoids the semantic indeterminacy of the model-theoretic argument.

Putnam describes himself as having been somewhat in the grip of the socio-functional view when writing RTH, but says in this paper that it's the naive view we want. The difficulty, he thinks, is making the naive view plausible, given recent work in brain science and cognitive science: in one way our focus on these has disrupted our ability to simply use our language as if we were not puzzled by the way it works. But once we get over the desire for distance from our language, and resign ourselves to simply using it, the answer becomes clear: "The answer to Quine's [ontological relativity] argument," Putnam says, "seems to me ... to be as simple as this: when we use the word 'Tabitha', we can refer to Tabitha and not to the whole cosmos minus Tabitha, because, after all, we can see the cat, and pet her, and many other things, and we can hardly see or pet the whole cosmos minus Tabitha" (1993, p. 183).

Setting aside the scope differences, and focusing on the substantive issue of how the name can connect up with the actual cat, we see certain similarities between this explanation and that of the Second Philosopher. The latter explanation begins with "'Tabitha' refers to Tabitha," goes on to explore the high degree of correlations between sentences involving "Tabitha" and the cat herself, noting along with Putnam that ordinary observable facts will bear on the matter. The Second Philosopher agrees that we can and ought to carry out, using our own language, further investigation into how our language functions; indeed, the metaphor of regaining the naivete of Austin is characteristic of Maddy's more recent work (e. g., 2011b).

However, there are substantive differences. Notice that the kinds of ordinary facts that seem most relevant are different: Putnam emphasizes "seeing" and "petting," while the Second

Philosopher's inquiry into indication relations emphasizes the explanation of successful bits of reasoning from which we draw inferences. Furthermore, Putnam presents the naive view as one we'll struggle to regain; the Second Philosopher treats the fact that she must use her own language and tools in inquiry as obvious and unproblematic. Perception, for the Second Philosopher, plays no special a priori role in our explanations of successful reasoning; instead, we can explain why and when perception works by appeal to biological facts about our senses, evolution, and so on. To understand these differences, we'll need a broader investigation into the relations between internalism and naturalism; I pursue this in the next section.

### 3. Internalism and naturalism

Here I pursue the relations between internalism and naturalism through consideration of three topics. Going beyond reference and truth, how does the model-theoretic argument look from the point of view of the Second Philosopher? How do the Internalist and Second Philosophical analyses differ when applied to the examples of mathematical objects? How do the Internalist and Second Philosophical analyses differ when applied to an example from Putnam's text, of the brain-in-a-vat true believers, convinced by the Guru of Sydney?

First, let's address the question of how the model-theoretic argument itself looks from the point of view of the Second Philosopher. As Maddy explains, the Second Philosopher is unimpressed by the argument itself because she was never attracted to metaphysical realism in the first place. Of course we cannot completely step outside our own conceptualizations to

compare how our thoughts, sentences, etc compare with a raw, unconceptualized world.<sup>4</sup> The internalist's project of analyzing, from within our ways of knowing, how thoughts and language function to represent the world is a project she shares. Their disagreement is over whether there is anything useful to say, from a different epistemic level, to support and justify those ways of knowing. Putnam's "higher purpose," Maddy writes, "is fairly clear: for all the attractions of a lower-level empirical account, he wants a theory of truth that isn't restricted to the confines of our current science" (2007, p. 100). The Second Philosopher does not feel the motivation of this project. Of this other form of inquiry, Maddy says, "If the Second Philosopher were pushed this far, she might well ask on what grounds Putnam draws any conclusions in this higher context, where all her ordinary methods have been set aside, but she will not be pushed this far" (2007, p. 102). I take this to mean the rejection of metaphysical realism seems significant to Putnam only because of his desire for an account that will, from outside the use of our methods, support and justify those methods -- a desire the Second Philosopher does not share.

Of course, it is notable that the Second Philosopher mentions "the confines of our current science." In my view, the most substantive difference between the positive views of the internalist and the Second Philosopher is this: though they both believe in working within our theories and using our methods, they disagree about what "our" theories and methods are: the Second Philosopher takes our best ways of understanding the world to be scientific methods, while the internalist appeals to "flourishing."

---

<sup>4</sup> I've called this the "Comparison Problem;" it is raised often in discussions about truth and realism. As Maddy (2007) says, the Second Philosopher does not aim to step outside our own conceptualizations, so the difficulty does not arise. See discussion in Marino (2006).

In interpreting this, it is essential to remember that the Second Philosopher is a native to the scientific worldview: it's not that we demarcate "science" and "non-science" and appeal to the former; it's simply that when we consider our best methods, they turn out to be broadly scientific. For example: when asked about the reasons to believe that atoms exist, the Second Philosopher does not answer "because science says so," but rather "because of 'the range of evidence provided by the likes of Einstein and Perrin.'" (2007, p. 397). We're still appealing to "our" methods, it's just we have some consensus that our methods are most trustworthy when they're based on evidence, and as natives to the scientific context, that's where we look for evidence. The internalist appeal to "flourishing" represents a different approach we take up all our ways of knowing, broadly conceived, refine those methods as best we can, and use them.

Let me try to illuminate this difference by considering examples. First, consider inquiry into the status of mathematical statements. In RTH, with an aim to illuminating issues in ethics, Putnam suggests that an internalist will take the statements of mathematics to be true despite the fact that mathematical objects are not material objects. Of fundamental assumptions, such as proposed new axioms of set theory, Putnam says "[these] may be adopted partly because of [their] agreement with the 'intuition' of expert mathematicians and partly for [their] yield" (RTH, p. 146) -- that is, for what they allow us to prove. But intuition is not some mysterious faculty: "mathematical intuition is good when it enables us to see mathematical facts 'as they are' -- that is, as they are in the mathematical world which is constructed by human mathematical practice (including the application of mathematics to other subject matters)" (RTH, p. 146).

Now let's look at the Second Philosopher's approach. Maddy considers three possibilities for mathematical objects. Roughly described, these are Robust Realism, in which mathematical

objects are abstract, eternal, objectively existing things; Thin Realism, in which mathematical objects, like "posits," are things whose features are revealed by the proper practices of set theory, including definitions, axioms, and means-ends arguments for and against new axioms, and Arealism, in which we look to exactly the same aspects of mathematical practice to understand mathematics, but refrain from saying the objects in question exist (Maddy 2005 and 2011a). Rejecting Robust Realism, she argues that Thin Realism and Arealism are ultimately not very different, with the choice hinging partly on how we understand the word "exists." But what is most relevant here is the reason mathematics comes out as trustworthy at all, and for TR and AR alike, this has to do with its role in scientific practice: it is essential to science, and thus accorded special standing not appropriate for just any area of inquiry. Astrology is a practice, but we have no reason to regard it as trustworthy. Mathematics, because of its role in science, is different.<sup>5</sup>

A full comparison isn't possible here, but it's noteworthy that as realisms, Thin Realism and internalism about mathematics bear some similarities, each taking mathematical objects to have the properties accorded them by mathematical theories and each emphasizing the way mathematical assumptions are evaluated for both their fit with other mathematical facts and for their usefulness in proving certain theorems. The bigger difference isn't over what, but over why: the difference concerns *why* a person would adopt such an -ism for this particular area of inquiry: for the Second Philosopher the role mathematics plays in science gives it a special status; for the internalist this role is not directly relevant.

---

<sup>5</sup> Maddy notes, however, that "This is not a reversion to a Quinean indispensability argument, because the conclusion is only that mathematics is different from pure astrology, not that mathematics is confirmed" (2007, p. 346).

This difference comes out more starkly in consideration of the internalist and naturalist approaches to a problem like that of the Sydney Guru. In this imaginary situation presented in RTH, the thesis that we are all brains in a vat is believed by "virtually all the people in some large country, say Australia," who have "been convinced of this by the "Guru of Sydney" who "just knows" that it is true -- and who is very convincing. These believers can do ordinary science in the normal way -- their airplanes fly, and their bridges stay up, and they agree with us about all worldly empirical matters -- they just also believe we are brains in a vat. Putnam says these Australians are "crazy" in the sense of having "sick minds" (RTH, p. 132). We should, he says, try to convince them of the error of their ways, and the best way to do this is to appeal to the way their views do not exhibit the virtues associated with an appropriate representation of the world -- one that is instrumentally efficacious but also "coherent, comprehensive, and functionally simple" (RTH, p. 134). Putnam says "... having this sort of representation system is *part of our idea of human cognitive flourishing*, and hence part of our idea of total human flourishing, of Eudaimonia" (RTH, p. 134).

What would a Second Philosopher make of this thought experiment? Based on Maddy's discussions of astrology and creationism in *Second Philosophy*, I propose a multi-part answer to this question. First, she would emphatically agree with Putnam that an examination and refinement of our methods using our methods is always appropriate: we examine our scientific practices by subjecting them to all kinds of analysis, and this analysis is often important when we defend our theories against rival views.

But second, from the point of view of the Second Philosopher, this suggests that as long as they have no explanation of what makes the Guru so trustworthy, there is one straightforward

sense in which the Australians are forming their beliefs in a peculiar way: their own methods are adopted for no good reason, even by their own lights. It would seem simple to set up tests for the Guru to show that in ordinary matters he does not have extraordinary ways of knowing; if the believers have not done so this is a simple evidence failure, not a complex one requiring appeal to values and human flourishing.

Third, if the believers respond to the idea of a test by insisting that there is no possible relationship between the facts about our world and the facts about the vat-world, because there is no way, even in principle, for us to perceive or know about objects in the vat-domain, then for the SP these vat-beliefs cannot be investigated for truth at all. Considering systems of belief concerning angels in a realm entirely disconnected from ours Maddy says, "as human activities, [these] would be apt subjects for the Second Philosopher's sociological or anthropological study" (2007, p. 346), but they would not be beliefs about the world and so they would not be of interest otherwise -- that is, they are not assessable for whether they are true. The brain-in-a-vat scenario is not quite the same, since there is a degree of causal overlap -- something is causing our collective hallucinations -- but presents a similar challenge, since it is a case in which, by hypothesis, our methods cannot work. For the Second Philosopher, we cannot assess such beliefs for craziness or sickness, because we cannot evaluate them at all.

Finally, the Second Philosopher may question Putnam's claim that it is essential that we be able to explain what is "sick" about the believer's views. If these people truly agree with us about every single thing related to this world and our life in it, if we can cooperate with them, live alongside them, and carry out successful projects with them, what difference does it make that they have peculiar views about matters we have no evidence to adjudicate? As we've seen,

the Second Philosopher will want to understand the Australians' beliefs -- from the psychological anthropological, and social point of view -- but does not need a way of showing that that these beliefs are false and "crazy." Of course, the hypothetical situation in this case is very different from the one we typically encounter in the real world, in which beliefs about gurus typically lead people to infer very worldly particular things. In the hypothetical, there are no such inferences.

What can we learn from these examples? As I mentioned in the introduction, in my view a crucial distinction has to do with what "our" methods are: do the methods we associate with science have special epistemic status? For the Second Philosopher the answer is Yes: "our" methodology is to trust the kind of evidence we find in scientific practice and experiment in a special way. Asked about a practice like astrology, the Second Philosopher points out, along with Richard Feynman, that there are all sorts of ordinary tests that would determine whether astrological predictions work and none of them show that they do (2007, pp. 108-109). For the internalist, the answer seems to be No: because of the way facts and values are intermingled in our epistemic practices, there is no reason to treat scientific evidence as a special kind of evidence.

#### 4. Discussion and implications

Of course, there is not space here for a full discussion of the question of whether scientific methods have -- or should have -- some kind of special epistemic status. But let me use this section to raise a few considerations.

As we've seen, Putnam's appeal to the very general notion of "flourishing" as an essential part of rationality is based partly on his rejection of the fact-value distinction. Briefly, the

connection is as follows: since science requires epistemic values, Putnam says, and since epistemic values can't be separated from other values, reasoning concerning the empirical world and reasoning including normative matters cannot be separated. But there is a difference between saying that some values play some role in science and saying that all kinds of beliefs must be evaluated from the very general point of view of human flourishing.

One striking development in philosophy since the writing of RTH is a set of more nuanced reflections on the particular roles values do and should play in scientific reasoning. For example, Heather Douglas's work in this area distinguishes between appropriate and inappropriate uses of values in science (see, e.g., her 2009 book). Douglas says that there is no way to rid science of values, and nor should we want to. Values do, and should, play certain roles in science, as when value considerations prompt us to ask one research question rather than another, or when judgments about tolerable risk inform conclusions over questions like "is this drug dangerous for humans?" But there are inappropriate roles for values to play. The most obvious is allowing a scientific conclusion or process to be altered specifically to reach outcomes that fit well with value judgments and interests. In the "internal stages" of science, as when we're characterizing phenomena and interpreting evidence, values should not play any such "direct" role. Preferring a certain outcome, or regarding that outcome as socially beneficial, should not be allowed to influence the methodology of a study so that a particular outcome is more likely to be found.

As I see it, forms of naturalism like that of Second Philosophy should have no problem incorporating and making sense of the appropriate roles values can play in scientific reasoning --

Patricia Marino, pmarino@uwaterloo.ca

Second Philosophers are, after all, appealing to practice. But I'd like to raise two thoughts about the generalized appeal to flourishing and the role of values in science.

First, appealing to flourishing to understand rationality makes it difficult to explain why values should not play a direct role in the internal stages of science -- and thus difficult to explain why political and ethical commitments should inform plans but should not alter our understanding of the facts. To seek out the most coherent and comprehensive view of the world given considerations of all kinds seems to allow the possibility that judgments about ethical and political matters should appropriately influence judgments about the facts, when to allow this would be a mistake. Consider the contemporary debate over climate change. Some climate-change deniers have strong political commitments against government intervention and against regulation. If rationality means forming an overall comprehensive and coherent theory with flourishing as its goal, and science has no particularly special status, then it would seem just as rational -- perhaps more so -- for this person to change his beliefs about whether climate change is real than to acknowledge that his political opinions do not fit with the scientific facts. Obviously, this would be a terrible error. Naturalism, according science a special status, makes plain and obvious that when it comes to matters like addressing climate change, scientific methods should have a special authority and the focus should be on these forms of evidence.

Of course, it is possible to say that failure to address climate change will not tend toward flourishing. But this is largely because of the facts -- facts we must take on the basis of evidence offered. The conceptualization of the point in terms of flourishing obscures this point while naturalism makes it clear.

Even ethical and political commitments that are likely to lead to flourishing should not, simply because they are more coherent with empirical conclusions, make us more likely to believe those empirical conclusions. That great distributive inequality is plausibly antithetical to flourishing is not a reason to believe any particular facts about the causes of inequality. For example, in the middle-twentieth century, Simon Kuznets gave an argument that claimed to show that market forces tend toward decreasing inequality in the long run. The reasoning has been criticized on various grounds. But clearly an ethical commitment in favor of reduced inequality should carry no weight with respect to the question of whether the Kuznet claim is true. In fact, it's the opposite, an ethical commitment to equality should make us subject Kuznet's claim to extra-rigorous testing and skepticism, given that its falseness would be so bad for us. This is an appropriate role for values in science, addressed toward such questions as "what are the risks of accepting the claim if it is wrong?" But if rationality is having a coherent comprehensive set of beliefs and rationality is about human flourishing, it is obscure why the ethical commitment would not count in favor of the empirical one, given how well they fit together.

My second thought has to do with way naturalism makes sense of some of our current practices of disagreement in a way that internalism does not. I'd like to suggest first that it is characteristic of our time, perhaps more than in the 1970s, to recognize the possibility of multiple forms of flourishing, so that relativism in ethics and politics is a real and complex issue in ways that it is not in science.<sup>6</sup> We often respond to others very differently in cases of scientific disagreements than in ethical or political ones.

---

<sup>6</sup> I do not claim there are no issues to do with relativism in science, only that there is a difference.

Consider as an example the ways that responses to the AIDS epidemic in Africa have prompted extraordinary measures to combat the illness. Some of these responses rest on empirical beliefs widely shared, but also rest on controversial ethical considerations, such as how to make trade-offs between the collective good, justice, and respect for individual autonomy. But other responses, like those of the South African president Thabo Mbeki, deny that HIV causes AIDS, thus disagreeing about scientific matters. Though not universal, it is common in our era to sharply distinguish the kinds of differences involved in these forms of disagreement and to respond to them in very different ways. To the claim that HIV causes AIDS we respond with a discussion about empirical evidence, and have no hesitation to adjudicate the matter on those grounds. To the idea that values can be prioritized in different ways our response is much more complex, varied, and nuanced.

Naturalism fits with and explains these different responses in ways the internalist approach cannot. After his discussion of the Guru of Sydney brain-in-a-vat believers, Putnam raises a parallel problem of "super-Benthamites" who always act to maximize "hedonic tone." Putnam emphasizes the parallel between the brain-in-a-vat true believers and the ultra-Benthamite true believers: each have world views that rest on a "sick conception of human flourishing" (RTH, p. 141). This parallel suggests that our ethical disagreements and our scientific ones are to be understood as deeply similar -- a conclusion that is, at the very least, seriously at odds with our current practices.

In this investigation of the views of RTH from the point of view of contemporary naturalism, I've argued 1) that claims about reference and truth cannot be adjudicated except in

Patricia Marino, pmarino@uwaterloo.ca

the context of broader epistemological considerations, 2) that when it comes to the broad issues of internalism versus naturalism, the main difference concerns whether scientific methods have a special epistemic status, and 3) that in certain ways internalism is an ill-fit both with contemporary understandings of the role of values in science and with contemporary practices having to do with disagreement.<sup>7</sup>

---

<sup>7</sup> I'm grateful to Jonathan Dewald, Heather Douglas, Tim Kenyon, and Penelope Maddy, for discussion and comments on previous drafts.

Patricia Marino, pmarino@uwaterloo.ca

### References

Burgess, John (2004), "Mathematics and Bleak House," *Philosophia Mathematica*, 12(1), 18-36.

Devitt, Michael (1983), "Realism and the Renegade Putnam: A Critical Study of Meaning and the Moral Sciences," *Noûs* 17(2), 291-301.

Devitt, Michael (1997), "On Determining Reference," in *Language and Thought*, Alex Burri, ed., (Walter de Gruyter).

Douglas, Heather (2009), *Science Policy and the Value-Free Ideal* (Pittsburgh University Press).

Leeds, Stephen (2007), "Correspondence Truth and Scientific Realism," *Synthese*, 159(1), 1-21;

Douven, Igor (1999), "Putnam's Model-Theoretic Argument Reconstructed," *The Journal of Philosophy* (1999): 479-490.

Field, Hartry (1994), "Deflationist Views of Meaning and Content," *MindNew Series*, 103(411), 249-285.

Maddy, Penelope (2005), "Mathematical Existence," *The Bulletin of Symbolic Logic*, 11(3), 351-376.

Lewis, David (1984), "Putnam's Paradox," *Australasian Journal of Philosophy*, 62(3), 221-236.

Maddy, Penelope (2007), *Second Philosophy: A Naturalistic Method* (Oxford University Press).

Maddy, Penelope (2011a), *Defending the Axioms* (Oxford University Press).

Maddy, Penelope (2011b), "Naturalism, Transcendentalism, and Therapy," in J. Smith and P. Sullivan, eds., *Transcendental Philosophy and Naturalism* (Oxford University Press), pp. 120-156.

Marino, Patricia (2006), "What Should a Correspondence Theory Be and Do?" *Philosophical Studies*, 127(3), 415-457.

Marino, Patricia (2010), "Modest Correspondence versus Representation-Friendly Deflationism" in Cory Wright and Nikolaj Pedersen, eds., *New Waves in Philosophy: Truth* (New York: Palgrave MacMillan), pp. 218-231.

Putnam, Hilary (1981) (RTH), *Reason, Truth and History* (Cambridge University Press).

Putnam, Hilary (1982), "Reply to Two Realists," *The Journal of Philosophy*, 575-577.

Patricia Marino, pmarino@uwaterloo.ca

Putnam, Hilary (1993), "Realism Without Absolutes," *International Journal of Philosophical Studies*, 1:2, 179-192.

Sosa, Ernest 1993: "Putnam's Pragmatic Realism," *Journal of Philosophy* 90: 605-26.

Wilson, Mark (2000), Inference and Correlational Truth," in Chapuis, André, and Anil Gupta, eds. *Circularity, Definition, and Truth* (Indian Council of Philosophical Research).

Wilson, Mark (2006), *Wandering Significance: An Essay on Conceptual Behavior* (Oxford University Press).