Putnam's writings on realism(1) have stirred up a rash of responses which raise questions both about Putnam's argument against metaphysical realism, and about internal realism, his positive view. A number of papers question Putnam's "brain in a vat" argument by trying to show that Putnam equivocates. Others argue that Putnam relies too heavily on formal methods(2), or on an overly restricted formalism(3). Internal realism has been accused of being incapable of giving a complete account of truth(4), and of being at least as problematic as metaphysical realism(5). I will argue that all of these responses fail to come to terms with Putnam's position through a common failing. I believe that isolating this failing will help to focus the debate.

According to Putnam, although realism in general is an empirical theory which stands or falls according to the evidence, the particular model embodied by metaphysical realism is incoherent(6). This model involves 1) a correspondence theory of truth, 2) bivalence of properties, and 3) the existence of a world independent of our representations. Its important consequences are that truth is radically non-epistemic, and that our best possible theory might be false. Putnam's arguments are directed primarily at these consequences, but his most striking conclusion is the rejection of a mind-independent world. This strong ontological conclusion comes mainly from criticism of the metaphysical realist conception of reference.

Demopoulos(7) has pointed out the banality of the fact that our ability to pick out facts depends on our language and conceptualizations. To this extent, at least, metaphysical realists can agree with Putnam. Demopoulos suggests that the realist believes that sentences are verified because the truth conditions of certain primitive sentences are satisfied, whereas the anti-realist rejects this condition, though he also notes that the notion of verification is itself infected by the contrasting positions, making the debate very precarious. This is why it is particularly important to get the basis of Putnam's position clear.

Putnam's main argument(8) relies on the assumption that referring is something that we do. This means more than that we do something that results in reference; it at least requires that any distinctions in reference are distinctions that we can have reasons for making. If metaphysical realism is correct, then it is possible for the ideal theory (the best theory we could possibly create, satisfying all of our methodological requirements) to be false. Assuming that the ideal theory must accurately predict all sense data, and that we can have at best a denumerable number of real-valued measurements, if we assume that the ideal theory is first order, or at least elementary, the Lowenheim-Skolem theorem implies that the ideal theory will be satisfied by non-isomorphic "intended models", where the "intended models" are just all the models which satisfy our ideal theoretical and methodological requirements(9). This idealized situation represents the extreme limit of what it is possible for us to do, and yet does not allow us, if metaphysical realism is correct, to distinguish between the "correct" interpretation and all the rest. Thus we can have no reason for distinguishing the "correct" interpretation from the others, and the central assumption of metaphysical realism, that such a distinction is possible, is incoherent.

This model-theoretic argument is backed up by the "brain in a vat" argument(10). A realist might argue that although we cannot fix the reference of our terms internally, their reference might be correctly fixed by external causal relations. Putnam asks us to consider that we are, and always have been, causally
isolated brains in vats with exactly the same phenomenology as if we were normal brains. If causal relations determine reference, both normal brains and the vat brains are correct in claiming that they are not brains in vats, since the vat brains, being causally isolated from the outside world, do not have the correct causal relations to refer to external vats, but only vats in the vat image. A brain in a vat is not in a vat in the vat image. Thus, the ideal vat theory and the ideal non-vat theory are equally true, contrary to what metaphysical realism would maintain.

The vat argument alone shows only that a causal theory of reference cannot rescue metaphysical realism. If the distinction between ideally justified and true theories is to be maintained, Putnam's model theoretic arguments show that it cannot be supported by purely internal factors. If the distinction is maintained by any sort of external causal relations (requiring that the vat brain and normal brain, despite all internal similarities, speak different languages), then a central principle of metaphysical realism, that a theory could be ideal but false, is wrong. The attempt to save metaphysical realism undermines it. Causality can be replaced in the argument by any other specific contingent relation except reference itself (or some equivalent). It seems that either reference "just happens", but not in any specifiable way, which Putnam derides as "the magical theory of reference", or else metaphysical realism is incoherent.

A number of authors(11) have argued with varying subtlety that Putnam has equivocated in the vat argument. In its simplest form, the criticism is that when the vat-brain says "I am not a brain in a vat", it is speaking vat-English, and it is true that it is not a vat-brain, but this does not mean that it is not a brain in a vat in the normal English sense. In any case, the criticism goes, the limitations on what we can say need not be the same as the limitations on the way we could be. Thus we could all be brains in vats, or at least there could be something wrong with the theory that says we aren't. Although any metaphysical realist would have to accept the basis of this criticism, it does not really touch Putnam's argument. The purported distinction between English and vat-English has no basis in anything we can do or can be done to us, according to Putnam's argument. This aspect of the argument must be addressed directly, on pain of begging the question. The distinction is required for the charge of equivocation to work, but it is exactly the cogency of this distinction which Putnam's model theory argument places in doubt.

Lewis(12) accuses Putnam of equivocating on our conception of causation and causation itself. This criticism can be dealt with in much the same way as the previous one, but it does bring out the issue of the limitations of formal representations. Since Putnam explicitly relies on first order theories, and relies on the formal notions of satisfaction and interpretation, it might seem that other formalisms or informal methods might avoid his conclusion. Second order logic looks promising, since the Lowenheim-Skolem theorem fails. Henkin's work in the early 1950's, however, shows that non-standard models of second order theories are possible. Although some object (correctly, I think) that the Henkin models are not really second order, the incompleteness of second order logic raises questions of how we could distinguish real second order logic from imposters. The issue again is what can be done effectively. Merrill(13) suggests that increased structure in the interpretation will reduce or resolve ambiguities. Farrell(14) recognizes that this will not work, since the new structure is also subject to multiple interpretations. He suggests that the answer is to be found in a move to an informal notion of interpretation. Rather than going into detail as to why these approaches will not work, I would like to point out that Putnam's use of first-order logic and formal notions is inessential. The real basis of his argument is in the limits of effective procedures, which, assuming Church's thesis, affect everything that can be done, formal or informal.

Our ability to make distinctions in referring is limited by the usual human inadequacies: contingent facts of our existence. Putnam's metaphysical realist would have it that we can make distinctions which we cannot justify (have evidence), otherwise we could not have a fully justified theory which was also false. This ability cannot be solely internal, and the metaphysical realist bases it in an alleged correspondence relation between our ideas and the world. The problem with this, for Putnam, is, what makes the correspondence? If we remove all restrictions on our abilities, and allow ourselves to do anything that can be done, there are still distinctions, such as between denumerable and non-denumerable cardinality, or between circumstances which satisfy an ideal theory, which cannot be justified by any possible procedure.
If reference to one and not the other is possible, it must be something that isn't done by any recognizable procedure, i.e., it must be magical(15).

I now turn to attacks on internal realism. Martin(16) tries to show that internal realism suffers from problems which are at least as severe as those of its competitor. This approach does run some risk of doing in realism completely. He objects to the use of the ideal theory, holding that it is not determinate enough itself to serve as a basis for determinate interpretation, and complains that internal interpretation supports the same indeterminacy of interpretation as external approaches. Lurking behind these objections is a concern, first raised by Demopoulos(17), that the notion of interpretation being "soft all the way down" is very unclear. One is inclined to think that if there is a rational way to divide up the world, there must be some basis in the world for the sectioning(18). Putnam would not deny this, but holds that the nature of the sectioning depends on the sectioning capabilities of the sectioner, and these are restricted to internally grounded distinctions.

I suspect that there is some basis for the concern about the nature of the ideal theory(19), but Martin is not clear why the ideal theory must actually exist, and why vagueness in its conception is problematic. As I have pointed out above, Putnam's use of the ideal theory in his arguments against metaphysical realism uses only the existence of a clearly defined limit: the limits of effective procedures. Putnam is vague about how the ideal theory is involved in reference. He does remark that the ideal theory has problems, but not the same problems as the metaphysical realist notion of truth(20). I think that, given his assumptions, the vagueness is appropriate: the details of reference are an empirical matter, not to be decided a priori. The existence of some limit, however, is a matter of logic.

Martin's objection that internal reference is just as ambiguous as external reference misses the same central point that our ability to refer is limited to what can be done. He ignores the restriction of internal reference to the scope of what can be done, whereas Putnam's metaphysical realist places no such restrictions on external reference. There may be indeterminacies in internal reference, but they aren't of the same sort as for external reference. Inasmuch as internal indeterminacies exist, reference is indeterminate to that degree. Martin seems to be importing the metaphysical realist bias that reference must be objective and determinate.

The last objection I will consider holds that the internalist notion of truth does not capture the limits of what is possibly true. Smart(21) gives an example of a theory which tells us that something is either true or false, but we can never tell which. His example is too speculative to determine whether or not it works, but contemporary inflationary cosmologies postulate determinate regions of spacetime from which we cannot receive information. It seems that truth cannot be equivalent to justified belief. Putnam's reply to this objection, based on what he says about counterfactuals(22), would be that the truth conditions of claims about the unobservables are given by what we would find if the restrictions preventing us from observing were not there. Truth conditions, if not truth, remain internal.

How can metaphysical realism be defended? I can't give a complete answer here, but I will make several remarks. First, I think we have to distinguish between meaning and reference. If not, the way we understand a concept will determine its reference, and Putnam's arguments invoking the limits of our capabilities are compelling. Secondly, I believe we should look at cases like Kuhnian anomalies in which new evidence requires a substantial change in what we think we are talking about. I suspect that counterfactual cases in which new evidence forces a radical change in our theory of the world will cause problems for internal realism (and any other form of anti-realism). Lastly, I think we should accept the possibility that the true nature of the world might be beyond even the possibility of our understanding. Putnam, I think, implicitly assumes otherwise. The metaphysical realist needs a notion of truth which can be divorced from understanding. Why must the truth make sense?

Notes


5. Martin, "Magical Theory".

6. "Realism and Reason".


8. "Models and Reality". There is some disagreement about whether this is Putnam's main argument, but I do not see how to make sense of what he is doing otherwise.


10. Reason, Truth and History: 5-14. Many writers have tried to deal with the vat argument independently. This cannot be done.


12. David Lewis, "Putnam's Paradox". There is much more in this interesting paper.

13. "Model-Theoretic Argument Against Realism". Merrill also opts for informal interpretations in this article.

14. "Blanket Skolemism".

15. See Reason, Truth and History: 17-21 for a clear statement of the importance of the limits of what can be done.

16. "Magical Theory".

17. "Rejection of Truth-Conditional Semantics".


20. "Realism and Reason".

21. "Metaphysical Realism".