A Putnam's Progress

ERNEST LEPORE AND BARRY LOEWER

In the medieval ages, before the arrival of scientific thinking as we
know it today, well, people could believe anything, anything could
be true. . . . But the wonderful thing that happened was that then
in the development of science in the western world was that certain
things came slowly to be known and understood. I mean, you know,
obviously all ideas in sciences are constantly being re-
vised. . . . That's the whole point. But we do at least know that the
universe has some shape and order. And that trees do not turn into
people and goddesses and there are very good reasons why they
don't and you can't believe absolutely anything.

—*My Dinner with Andre*

Hilary Putnam tells us that metaphysical realism (MR) has been the dominant view
in philosophy of every historical period (at least until Kant) (1978, 1). One can find
it in Plato, Aristotle, Descartes, and Locke. In this century Russell and early Witt-
genstein (on the usual interpretation) were explicit advocates, and currently Michael
Devitt (1981, 1984), Hartry Field (1972, 1982), Clark Glymour (1986) and David
Lewis (1986) confess to it. Putnam himself says he held it implicitly, if not explicitly
(1983, vii). Recently, however, Putnam announced that he has discovered MR is in-
coherent (1978, 124). And he wants to replace it with another kind of realism, inter-
nal realism (IR), a realism with a human face. Putnam argues that MR is committed
to a picture of the relation between thought (or language) and reality that is fun-
damentally implausible; it leads to an unacceptable view of the nature of truth and
to pernicious dichotomies. On the other hand, IR is claimed to maintain the scientific
aspects of realism while rejecting its metaphysical aspects. So, Putnam has given us
a characterization of MR, has produced a bill of indictment against it, and sketched
an alternative view, IR. We think it fair to say that Putnam has not succeeded in con-

vincing many metaphysical realists of the errors of their ways. The usual reaction to his arguments is puzzlement. It has seemed to many that Putnam has betrayed his previous realist contributions—he was even pronounced a renegade by a former student (Devitt 1984)—and adopted in its stead a soft, some would say, a quasi-mystical view. Our aim here is to relieve some of this puzzlement by constructing a "reading" of Putnam's progress from metaphysical to internal realism.

1. Scientific and Metaphysical Realism

Bas van Frassen characterizes scientific realism (SR) in this way:

> Science aims to give us, in its theories, a literally true story of what the world is like; and acceptance of a scientific theory involves the belief that it is true. (1980, 8)

Putnam would agree with this characterization and add that we sometimes have good reasons to believe that scientific theories, particularly those in the mature sciences, are true or approximately true and that the entities they posit exist. The truth of a theory is independent of our beliefs concerning it. Our present good reasons for believing a theory to be true do not guarantee that it is true or approximately true. Any theory we presently hold we may come to reject for good reasons. Furthermore, successive theories can often be viewed as better approximation to the truth. Thus, Bohr's 1934 theory of the electron is closer to the truth than his 1912 theory. Such comparisons require interpreting terms within successive theories as referring to the same entities, in this case, to electrons. One of Putnam's most important contributions to the defense of SR was to show how a causal account of reference can be used to support claims concerning intertheoretic reference. Bohr's 1912 term "electron" and his 1934 term "electron" were appropriately causally related to the same entity. Of course, this claim is made on the basis of present theory (and what we know of the history of the development of Bohr's views) and is itself defeasible. Putnam considers SR to be an overarching *empirical* hypothesis which best explains scientific practice and success (1978, 123).

We will not attempt to assess SR or Putnam's arguments for it.[1] Rather, we want to show how naturally it seems to fit into a more general philosophical view concerning the relationship between language and the world, namely, metaphysical realism. Putnam's official characterization is (1981, 49):

(1) The world consists of a *fixed* totality of mind-independent objects and properties

(2) Truth involves some sort of correspondence relation between words or thought-signs and external things and sets of things.

(3) There is exactly one true and complete description of the way the world is (though we may never have a language capable of expressing it or may never know it).

MR appears to lend philosophical support to SR by providing an account of truth and reference that applies to all theories. It allows for the possibility that terms in different theories refer to the same theory-independent entities. By characterizing truth nonepistemically (in terms of correspondence) it allows for the possibility that even our best theories might yet turn out to be false. And it seems to provide a way of making sense of the idea that successive theories converge on the one true description of reality.

But according to Putnam, there are important differences between MR and SR. He says that MR "is supposed to apply to all theories at once . . . and THE WORLD is supposed to be independent of any particular representation of it" (1978, 125). On the other hand, scientific-realist claims, for example, that Bohr's two uses of "electron" referred to the same phenomena, are made within a particular theory, present-day quantum theory. Putnam's alleged discovery is that MR, instead of supporting SR, actually undermines it. As we will see, the heart of his argument is that the reference and correspondence relations invoked by MR cannot be placed with a scientific account of the world.

2. THE ARGUMENTS AGAINST MR

In this section we discuss some of Putnam's arguments against MR. He observes that

> the most important consequence of metaphysical realism is that *truth* is supposed to be *radically non-epistemic* — we might be "brains in a vat" and so the theory that is "ideal" from the point of view of operational utility, inner beauty and elegance, "plausibility", simplicitly, "conservativism", etc., *might be false*. "Verified" (in any operational sense) does not imply "true", on the metaphysical realist picture, even in the ideal limit. (1978, 125)

Putnam's claim is that the MR's (1), (2), and (3) entail:

(4) Truth is radically nonepistemic.

(5) It is possible that we are all brains in a vat even though we believe we are not.

(6) Even an epistemically ideal theory may be false.

(5) and (6) are Putnam's glosses on (4), and he appears to take (5) to imply (6). The argument proceeds:

> I assume THE WORLD has (or can be broken into) infinitely many pieces. I also assume T_1 *says* there are infinitely many things (so in *this* respect T_1 *is* "objectively right" about THE WORLD). Now T_1 is *consistent* (by hypothesis) and has (only) infinite models. So by the completeness theorem (in its model theoretic form), T_1 has a model of every infinite cardinality. Pick a model M of the same cardinality as THE WORLD. Map the individuals of M one-to-one into the pieces of THE WORLD, and use the mapping to define relations of

M directly in THE WORLD. The result is a satisfaction relation SAT – a "correspondence" between the terms of L and sets of pieces of THE WORLD – such that the theory T_1 comes out *true* – true of THE WORLD – provided we just interpret "true" as TRUE (SAT). So what becomes of the claim that even the ideal theory T_1 might *really* be false? (1978, 125-26)

The structure of the argument is this:

    (i) MR implies (6).

    (ii) But an epistemically ideal theory has an interpretation SAT which makes it true.

    (iii) So, the epistemically ideal theory is true.

    (iv) So, MR is false.

There are a number of places where one might take issue with the argument. It is not obvious that the concept of an ideal theory is clear.[2] And it may be objected that although "Even an ideal theory is false" is consistent with MR, it is not obvious that MR *entails* it is possible that an ideal theory is false.[3] But the place at which most will balk at Putnam's argument is the move from (ii) to (iii). If someone told us that a theory, for example, phlogiston theory, is true because it is consistent and so has a model in THE WORLD, we would think him crazy. We would point out the theory's defects – it makes false predictions. The case of an ideal theory is different since it has no epistemic defects, for example, fails to predict an observation (Putnam 1978, 126-27; 1981, 45-48, 55; 1983, 8-15). But the MR still may object that SAT is not the correct interpretation of the ideal theory; the correct interpretation SAT* is the one determined by the reference and correspondence relations between the terms of T_1 and THE WORLD. Interpreted by SAT* T_1 may turn out to be false. Now Putnam asks this question: What makes SAT* *the correct interpretation*? This is the key question. The force of the argument is to claim that the MR must answer it if he is to defend his view. We will consider answers he might give shortly. First, we want to consider a different response to the argument that seems to avoid the need for answering the question.

Suppose that an MR argues as follows: I cannot say what makes a particular reference relation the correct one, but I can show that whatever does allows for the possibility that even an ideal theory is false. Here is how (1978, 125):

    (i) It is possible that I am a brain in a vat experiencing the stimuli I experience even though I believe I am not a brain in a vat.

    (ii) The ideal theory that is epistemically ideal for a brain in a vat that experiences the same stimuli that I experience is identical to the epistemically ideal theory for me.

    (iii) Among the statements included in the ideal theory is that I am not a brain in a vat.

    (iv) But in that case the ideal theory constructed by the brain in the vat must be false.

This response to Putnam's argument, whatever its other features, depends on the assumption that a brain in a vat can entertain the thought that it is a brain in a vat. But Putnam has argued that, given the causal theory of reference, it is not possible for a brain in a vat to even entertain the thought that it is a brain in a vat (1981, chap. 1). Putnam's argument for this conclusion is usually taken to be a new (or not so new) argument against skepticism, but critics are quick to observe that as such it is not very convincing since the argument depends on a premise that the skeptic is unlikely to allow: the causal theory of reference (Brueckner 1986, Conee 1987, van Inwagen 1988). But the MR whom we imagine making this response to Putnam's original argument accepts the causal theory of reference. If we see the brain-in-a-vat argument as a reply to that MR's response, then Putnam's puzzling article seems to make sense.

In "A Problem about Reference," Putnam constructs an argument that is related to, but slightly different from, the argument based on Godel's completeness theorem. Putnam calls this the permutation argument:

> I shall argue that even if we have constraints of whatever nature that determine the truth value of every sentence in every possible world, still the reference of individual terms remains indeterminate. (1981, 33)

Suppose that there is a language L and an assignment to each sentence of L at world w a truth value. These assignments might be determined by whether or not the sentences satisfy certain constraints, for example, are part of an ideal theory at that world. Putnam shows that there are distinct interpretations of the predicates and constants of the language that agree on the assignment of truth values at possible worlds. He concludes that this shows no view about how content is determined, which works by assigning truth conditions to whole sentences, is capable of explaining the reference relation. We seem to have the following argument:

(i) To defend MR one must specify the reference relation.
(ii) Specifying the reference relation by listing a collection of sentences even if we take those sentences to have truth conditions — functions from possible worlds to truth values — is not sufficient for specifying a unique reference relation.
(iii) So, MR cannot be defended.

A natural response to both of Putnam's "model-theoretic" arguments is to say that the interpretations of one's terms and thought-signs are fixed by their causal connections to items in the world. In fact, a number of commentators have expressed puzzlement as to why Putnam, who is one of the developers of the causal theory of reference, does not simply recognize that a causal theory of reference will succeed in specifying the "intended" interpretations (Blackburn 1984, 301; Brueckner 1984, 137; Devitt 1984, 86-87; Glymour 1982, 177; Harman 1982, 569, 573; Leeds 1978, 113; Lewis 1986):[4]

> The natural modern reply to the rhetorical question [What fixes reference?] is that all of the intended interpretations should be replaced by talk of causally

determined reference relations; roughly, our physical and social circum-
stances, and sometimes perhaps our beliefs as well, determine together a set
of links, connecting words and objects, and thus delimiting the admissible in-
terpretations of our theories. (Glymour 1982, 177)

So Glymour's view is that the admissible interpretation of our language, the interpre-
tation under which an ideal theory might turn out to be false, is determined not by
the sentences of the theory but by causal relations between bits of the world and
terms of the theory.

Putnam's response to the suggestion that a causal theory of reference will suc-
ceed in answering the key question has exasperated his critics. He says that the causal
theory of reference is just more theory and observes that the argument applies to it
as well; in particular, there are interpretations that assign "cause" some relation
cause*, which makes all the sentences in the causal theory of reference (and the rest
of the theory) come out true. Putnam thinks that the MR is begging the question by
appealing to the causal theory of reference (or any other account for that matter)
since the appeal works to single out a unique reference relation only if we assume
that "cause" refers to cause and not, say, to cause*. Glymour, Devitt et al. think that
at this point Putnam has moved from name-calling to game-playing. Their claim is
not that adding the causal theory of reference to their theory fixes reference, but
rather that causation itself fixes reference. The causal account of reference also is
supposed to apply to the terms appearing in it—so that the reference of "cause" is
causally determined to be cause instead of cause*. Devitt accuses *Putnam* of begging
the question:

The question begging is most striking in Putnam's latest response to the idea
of a causal theory of reference. He claims that "if reference is only determined
by operational and theoretical constraints then the reference of terms in that
theory of reference will themselves be inderterminate." . . . Maybe so but
if reference is determined causally, as the theory says it is, then the reference
of those terms will be determinate. He is not entitled to assume the theory is
false in order to show it false. (1984, 190)

Devitt has missed the force of Putnam's argument that there are alternative reference
relations that result in identical assignments of truth values at every possible world.
If there is a causal characterization of reference, there also are alternative characteri-
zations (in terms of cause*, cause**, and so on) that agree on the truth values of
statements at all possible worlds. The question Putnam asks the MR is why say that
the account of reference in terms of cause rather than in terms of cause* is the correct
one? On MR both accounts cannot be correct since presumably that x causes y is
a different (physical) fact than that x causes* y (even though the sentences are true
in the same worlds) and the MR says that x refers to y is a particular fact in the world.
So the challenge to the MR is to give some reason for holding one identification of
reference rather than another. The problem that confronts the naturalistically

minded MR (as those who put forward causal accounts claim to be) is that from the point of view of scientific explanation, the different identifications of reference are equally acceptable. That is, if the causal theory assigns truth conditions in such a way as to meet whatever standards placed on the theory, so will the causal* theory. Putnam is not, as Devitt says (1984, 189), like a small child delighting in discovering that there is no end to questions. Instead, he has asked a question — Why identify reference with cause rather than cause*? — that the naturalistic MR cannot answer.[5] Further, this question is not unanswerable because the MR lacks some added piece of information. It is in principle unanswerable as long as the MR keeps his empiricist (naturalist) credentials. If one were to give up naturalism, one could reply to Putnam that there simply is a brute metaphysical fact that reference is identified with causation and not causation*. Here name-calling — it is a magical, mysterious theory of reference — would be warranted (1981, 3). For this would be to suppose that an *intrinsic* property of an object can determine its relation to a particular object external to it.

We can now summarize our discussion of the model-theoretic arguments as follows: *None proves that MR is false.* Rather, they are vivid ways of putting to the MR the question what on his view determines the reference of representations? That Putnam also sees his arguments in this light is shown by the following remark:

> The only paper in this book which makes use of technical logic "Models and Reality" is not an attempt to solve the problem [how correspondence is fixed] but rather a verification that the problem really exists. (1983, ix)

3. PUTNAM AND KRIPKENSTEIN

The MR's picture of the relation between language and the world is this:
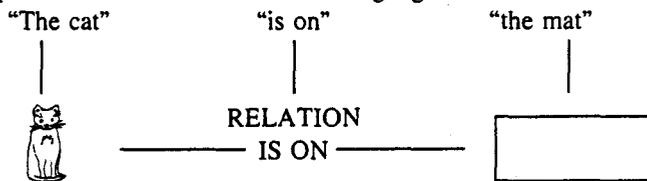
The vertical lines indicate reference relations between terms and parts of the world. The important feature is that the holding of these relations are themselves facts — facts that must obtain for "The cat is on the mat" itself to succeed in stating a fact. Putnam's question to the MR, then, is: What fixes these reference relations? Or, in what do these facts of reference consist? For an MR who is also a naturalist it is important that in answering this question he makes use only of naturalistic properties and facts. To complete Putnam's argument against MR, one would have to show that the MR cannot provide an answer to this question consistent with his naturalistic scruples. Although Putnam does not do this, it seems to us that the situation is exactly this: Reference and meaning are *not* naturalistically reducible. Kripke argues for this claim in his discussion of Wittgenstein on rule-following (1981).

The question he asks is similar to the one we have found in Putnam. What fact makes it the case that my use of "plus" refers to plus (or means that plus — in Kripke)? Kripke observes (as does Putnam) that citing images in my mind when I use "plus" or my intentions to use "plus" in a particular way cannot provide a satisfactory answer, since images and intentions are themselves open to interpretation. The heart of Kripke's objection is that if there were a fact that my use of "plus" means plus, then that fact would have to possess a certain normativity. It would have to make it the case that if I answer "5" to the question "What is 68 plus 57?" I would be *wrong*. Kripke argues that no "natural" fact possesses this normativity. For example, the fact that I have dispositions to answer various questions involving "plus" is not the appropriate kind of fact, since the dispositions do not themselves distinguish right answers from wrong ones. One might think that a causal theory of reference might be of use here.

A kind of causal theory suggested Fodor (1987) and Stalnaker (1984) is this: A predicate P (in the language of thought of person X) has as its extension, say, cats iff under optimal conditions X's tokenings of P are caused by cats and only by cats. There are many difficulties with this kind of account.[6] Here we will discuss two. One is that to make this account even slightly plausible, the specification of "optimal conditions" will make ample use of intentional notions, including specifications of X's other beliefs. Obviously, this is inadmissible in a naturalistic reduction of reference. Second, a Goodmanesque problem arises: If under optimal conditions tokenings of P are caused by all and only by cats, they will also be caused by all and only by cats*, where y is a cat* iff if conditions are optimal y is a cat and if conditions are not optimal y is a dog.[7] Although neither Kripke nor Putnam has produced a proof that a reduction of semantic facts to natural facts is impossible, given the preceding it seems an extremely poor bet. If we are right in our reading of Putnam, he has noticed the same problem that Kripke has for the MR. And the prospects for the MR answering the question seem extremely dim.

Must the MR solve the Kripke problem to defend his views? He might make a familiar move that cat* is a peculiar property and really is not, other things being equal, an eligible referent for a term. This amounts to there being a special class of "natural" properties that are better candidates than the gruesome properties to be the referents of our terms.[8] In the context of physicalistic MR, this is an odd view. What makes something a gruesome or natural property? That the natural properties just happen to be the ones we refer to might be taken to mean we simply *read* the structure of our language into nature.

A comparison with moral statements is useful. A naturalistically minded MR considers moral statements as non-fact-stating unless he can show how they are made true by natural facts. If he cannot do so he will consider them as performing some other function, say, expressing preferences or attitudes (for example, Ayer 1936; Mackie 1977). Is a similar move open to the MR? The following argument suggests that it is not (Wright 1984; Boghossian, unpublished manuscript). Suppose that S is true iff p. If "S is true" expresses an attitude, not a fact, then it seems to follow

that "p" does not state a fact either. For the MR this is to throw the baby out with the bathwater. So, although an MR might be led to relegate ethics to a second-class status, he cannot do the same for semantics without undermining his own position.

We would like to consider briefly two other responses open to the MR. One is just to embrace semantical relations and facts as irreducible (or primitive). Putnam reserves his most creative name-calling for this position, accusing it of being "medieval" and "magical." It amounts, for Putnam, to saying that the mind somehow reaches out (perhaps by shooting noetic rays!) and touches objects. Beyond the name-calling, there is the irony that a position originally intended as a philosophical basis for scientific realism would end up appealing to nonnatural facts.[9] The second response is to try to mitigate the mysteriousness of nonnatural facts by claiming that they *supervene* on natural facts. But, as Simon Blackburn has argued in the case of the alleged supervenience of moral properties on natural properties, without a reduction of the former to the latter, "supervenience becomes, for the realist, an opaque, isolated logical fact, for which no explanation can be proferred" (1971). The problem is that if moral statements are realistically interpreted as stating facts, supervenience is itself a mystery unless there is a reduction of moral properties to natural ones. One can respond by abandoning moral realism while keeping supervenience as a constraint on the attribution of moral predicates. As we saw, this response is not available to the MR. So as long as he cannot produce a reduction of semantical properties to natural ones, the supervenience claim does little to dispel the mystery.

4. INTERNAL REALISM

Putnam tells us that it was Dummett who awakened him from his metaphysical slumbers, and it will be useful to describe briefly certain features of Dummett's account of meaning and truth as preparatory to discussing IR. Dummett, like Putnam, rejects the view that truth is radically nonepistemic. According to Dummett:

> A statement is true only if we have established it as true; or only if we either have done so or shall do so at some future time; or only if we have some procedure which, were we to carry it out, would establish it as true; or, at least, only if there exists something of the sort that we normally take as a basis for the assertion of a statement of that class, such that, if we knew it, we should treat it as a ground for the truth of the statement. (1981, 443)

This idea is best illustrated in Dummett's philosophy of mathematics (1978, xxiv). Associated with a mathematical statement are conditions for proving it. *Understanding* the sentence is a practical ability to manifest a particular sort of behavior, namely, that behavior which brings us into the position in which, if the condition that conclusively justifies the assertion of S obtains, we *recognize* it as so doing. In case of a mathematical statement this requires that we recognize what would count as a proof of the statement. The recognizable conditions of a sentence's conclusively

justified assertion are its verification conditions. Dummett denies that a sentence has truth conditions other than those associated with it recognitionally by the speaker's understanding. So, a sentence has only *verificationist* truth conditions. A mathematical sentence, thus, is "true" iff it has a proof (1973, 467-68; 1975, 115-23; 1976, 70-111).

All understanding for Dummett is verificationist. Although Dummett has not developed a well-worked-out account for nonmathematical sentences, he thinks the situation for these sentences is similar. To understand, for example, "There is a chair in the room," is to know what would count as a (conclusive) justification of it: perhaps seeing a chair in the room under good light conditions, not being intoxicated, and so on.[10] On this view, there may be sentences such that neither they nor their negations are true. (To justify the negation of S is to [conclusively] refute S.) We may know what counts as a (conclusive) justification (and refuation) of a sentence S, but none might exist. Finally, on this account, reality is mind-dependent (at least) in this sense: Whether the sentence "There is a chair in the room" is true depends on its justifiability conditions. Truth and rationality (justifiability) are conceptually bound—no truth without conclusive justifiability. For example, our finite abilities limit what we can justify.

Dummett's view contains numerous subtleties. Here we want to ignore most of them and pursue a line of thought that will be useful for our comparison with Putnam.[11] It is notorious that Dummett wavers between considering "justification" to mean "conclusively justified" and "sufficiently well justified" (1973, 146, 467, 514, 586; 1978, xxxviii). The problem is that on first reading, it may be that only mathematical and sense-perception statements, if even they, are ever justified. If truth is identified with conclusive justification, it would follow that only mathematical and sense-perception statements are capable of having truth value. Also, since for Dummett the justification conditions of a sentence are its meaning and since understanding a sentence is just knowing its meaning, it follows that anyone who understands a sentence knows its justification conditions, that is, the canonical conditions such that if they were satisfied the sentence would be true. An obvious difficulty with this view is that it makes it extremely unlikely that two different people, especially different people who live at different times or in different societies, speak languages that are translatable. For example, it is extremely unlikely that any sentence spoken by Tycho Brahe has the same justification conditions as our sentence "The sun revolves around the earth." The result is incommensurability. We would like to say that Tycho Brahe was wrong to believe that the sun revolves around the earth, but Dummett's theory together with the fact that our procedures for verifying or confirming sentences change seems to preclude us from doing so.

With this capsule summary of Dummett under our belts, we can describe IR by constrast. Putnam's own proposal, IR, is very sketchily presented. The heart of his view seems to be a rejection of the metaphysicalist realist picture. It is replaced with a view that truth is justification. He writes:

> My own view . . . is that truth is to be identified with justification in the sense of *idealized* justification, as opposed to justification on the present evidence. (1983, xvii)

> Consider the sentence "There is a chair in my office right now." Under sufficiently good epistemic conditions any normal person could verify this, where sufficiently good epistemic conditions might, for example, consist in one's having good vision, being in my office now with the light on, not having taken a hallucinogenic agent, etc. . . . There is no single general rule or universal method [contra Dummett] for knowing what conditions are better or worse for justifying an arbitrary empirical judgement. (1983, xvi)

Putnam is not trying to give a formal definition of truth but an informal elucidation of the notion. The two central ideas of his idealization theory of truth are:

(a) Truth, though independent of justification here and now, is not independent of all justification. To claim a statement is true is to claim that it could be justified under ideal conditions.

(b) Truth is expected to be stable or convergent; if both a statement and its negation could be "justified," even if conditions were as ideal as one could hope to make them, there would be no sense in thinking of the statement as having a truth value. (1981, 56).

His view seems to be this: Our practices of forming judgments, testing them, arguing about them, and so on, are sufficient to associate with statements what would count as reasons for believing the statements. For some statements we have a pretty clear idea of what it would take to justify them and the conditions under which they would be justified, for example, "There is a chair in my office right now." Putnam offers no account of knowledge, but presumably he has in mind that there is some account of better and worse reasons for my believing that there is a chair in my room right now. Somehow, community practices result in associating justification conditions with statements. These are conditions such that if we believed they obtained we would be justified in stating that X believes that p. Since truth is idealized justification, it is not the case that "The sun revolves around the earth" changed truth value and/or meaning between the time of Tycho Brahe and now. It is rather that views have changed about whether that statement is (ideally) justified. This is an improvement over Dummett's views, but it is committed to the position that it is impossible (by definition) for our ideally justified beliefs to be false. We saw before that despite the MR's protests, Putnam argues that the MR cannot maintain his intuition about the logical possibility of the ideal theory being false. IR does not rule out the possibility that we might be very well (ideally) justified in believing another to be a brain in a vat. But IR seems to rule out the possibility that for any individual he can think of himself that he is a brain in a vat and also be (ideally) justified in this thought.

It is important to clear up certain misconceptions, perhaps owing to Putnam's own rhetoric, about IR. In denying the MR's (1), in the first section of this paper,

Putnam is denying that the world consists of a fixed totality of mind-independent objects. This smells of "idealism," and Putnam's favorite metaphor that the mind and the world together make up the mind and the world encourages the charge that IR has idealist tendencies. One way of expressing these tendencies would be to say that under IR, ordinary commonsense counterfactuals fail to be true. But it is no consequence of IR that counterfactuals like "If we had not constructed the theory of electrons, then there would be no electrons" are true. In fact, on Putnam's, but perhaps not on Dummett's, account the counterfactual "Even if we had not constructed the theory of electrons there would be electrons" is justified. So we have reason to think it is true. In general, whenever S is justified, so is "Even if I had not thought of S it would be justified." (Of course, this counterfactual is not justified for every S.) This may not satisfy an MR since he wants to say not merely that one would be justified in asserting the counterfactual, but that it is a *fact* that even if I had never thought of electrons they would have existed. But the IR *is* justified in saying this as well. So whatever the alleged mind-dependence of the world consists in for the IR, it cannot be expressed by him counterfactually since he and the MR agree on (most) counterfactuals they assert.[12] The difference, rather, is in what they think the truth of these counterfactuals consists in. For the MR it is correspondence to facts. For the IR it is in being ideally justified.

As early as 1978 (50–51), Putnam made some rather strong statements about the relationship between realism and equivalent descriptions. He stated that we cannot ignore the existence of pairs of equivalent descriptions. Scientific realism is not committed, he contends, to there being one true theory (and only one). He says that "assuming there is a 'fact of the matter' as to 'which is true' if either whenever we have two intuitively 'different' theories is naieve." Putnam's intuition is well accommodated by his IR. Since idealized justification is theory relative, then, for example, relative to field theory "Fields are real" may be ideally justified, though relative to particle theory it may not be ideally justified. Therefore, it is consistent with IR that there be more than one "true" theory or description of the world (contra the MR's (3)). It is not surprising, therefore, that Putnam should be charged with relativism (Pears 1982). He is a relativist inasmuch as whether a statement is true is relative to a theory. But it is a very limited relativism (1978, 38–41; 1981, 117–19). Putnam claims that ordinary statements about ordinary objects, for example, "There is a chair in the room," are, if true, true relative to all acceptable theories (Putnam ms$_2$). Relativism applies only to statements on the periphery, for example, the one about fields above, statements in set theory. Whether Putnam's "relativism" is this confined or not, he argues that IR does not possess what he takes to be the two most pernicious features of relativism: (1) relativism cannot distinguish between "P is true" and "I (we) think that p is true" and (2) relativism leads to rampant incommensurability. We cannot go into why Putnam's views escape these charges here except to remark that (a) there is a distinction on IR between ideally justifying "I think that p" and ideally justifying "p" (1981, 124) and (b) as we have mentioned a number of times, Putnam thinks that translation-claims between theories can themselves be ideally

justified and we frequently have good reason to believe these claims are satisfied (1981, 116–17).

The obvious question that one is itching to ask Putnam is this: Doesn't a problem analogous to the MR's problem of explaining how the reference relation is fixed arise for IR? What makes it the case that a particular sentence has the justification conditions it has as opposed to other justification conditions? Similar questions could be asked concerning the reference and meanings of terms. The MR may smugly suppose that the IR will have as much difficulty specifying the naturalistic facts that make these semantical relations hold as he had. This way of asking the question reveals an MR bias. The MR is asking what *fact* makes it the case that "There is a chair in the room" has whatever ideal justification conditions it has? But the IR has rejected the account by which facts make assertions true. This being so he need not satisfy the MR's demand in order to defend his theory. (In contrast, the MR did need to satisfy the IR's request since the MR held that there is a fact about what terms refer to.) What he needs to do instead is to show that the sentence "It is part of the ideal justification conditions of 'There is a chair in the room' that . . . " is itself justified—similarly, for sentences like " 'Cat' refers to cats" and "Bohr's term 'electron' in 1912 has the same reference as his term in 1934." Someone may ask what fact makes it the case that these semantical sentences have the justification conditions they have. The IR answer is that there is no such fact. Rather, certain assertions about their justification conditions are justified. As Putnam says (quoting Dummett), "Facts are soft all the way down" (1978, 128).

The IR can reply to the permutation argument as follows: He says that " 'Cat' refers to cats, not to cats*" is justified. He can support this by pointing out that "cat" is appropriately causally related to cats but not to cats*. Neither objection made against the MR applies here since the IR is *not searching for a naturalistic fact* with which to identify reference.

5. CONCLUSION

The IR's response to the Kripkenstein problem may appear to be a sleight of hand. Simply say that sentences have justification conditions, not realist truth conditions, and—Voilà!—the problem of intentionality vanishes. But does it? Isn't there, for example, still a problem of accounting for the causal efficacy of intentional (and other mental) states, if these are not naturalistically reducible? These are questions we address elsewhere.[13] Here we rest, content to have uncovered what we think to be a provocative "interpretation" of Putnam's text.[14]

Notes

1. For a detailed assessment of SR and Putnam's arguments for it see Devitt 1984.

2. Putnam has himself on several occasions remarked that the notion of an (epistemically) ideal theory is far from clear (1983; 12, 161).

3. MR does not mention "ideal" theory. So it is hard to see how it could imply "It is possible that

an ideal theory is false." But since MR is supposed to provide a nonepistemic characterization of truth, it does seem that the MR would embrace (6). See Glymour (1982, 176).

4. Carsten Hansen writes:

That Putnam in "Models and Reality" does not discuss his previous work in "The Meaning of 'Meaning' " is quite puzzling. For in the latter, after having argued that 'meanings aren't in the head', he goes on to give a realistically acceptable account of the extensions in a nonsolipsistic, social setting. . . . It needs to be explained why Putnam no longer thinks this explanation is available to a realist" (1987, 91).

There are two reaons why the causal theory of reference, which Putnam *continues* to hold, does not provide a sufficient response to the challenge. The first is that the causal theory as developed by Putnam provides at most constraints on what counts as an acceptable interpretation and it itself uses intentional notions and so presupposes that the problem of specifying an interpretation has already been solved. For my use of "water" to refer to water on Putnam's theory requires that I have certain perceptions and intentions, for example, that I intend to refer to the same substance that whoever I acquired the term used it to refer to. The second response concerns a particular feature of reference. On some versions of the causal theory of reference certain terms in my language—names and natural kind terms—can refer only if there is a causal chain connecting my use of the term with its reference. This causal chain may go via other speakers. [It is not clear that Kripke holds the view that there must be such a causal chain.] This requirement seems much too weak to single out a unique interpretation or even to remove from consideration the interpretation SAT which makes the ideal theory true.

5. When I ask why is reference fixed by causation and not by causation*, the only answer a physicalist can give me is "because that is the nature of reference." To say that Nature itself singles out objects and put them into correspondence with words is a claim that has no meaning that I can make out at all (Putnam ms₁).

6. See Loewer 1987.

7. A slightly different theory is found in Dretske 1981: If the original tokens (those that occurred during what Dretske calls the "learning period") of the predicate P carried the information cats are present, then subsequent tokens (those that occur subsequent to the learning period) refer to cats even if they are not caused by cats. The trouble with this is that the appropriate notions of information and the learning period will likely include intentional concepts. See Loewer 1987.

8. David Lewis has recently (1986) taken up the suggestion that there are certain classes of things "out there" that are intrinsically distinguished and he suggests it is a "natural constraint" on reference, that is, a constraint that is built into nature, that as many of our terms as possible should refer to these classes. Lewis's natural constraint is not brought into existence by our interests. It has to be thought of as something that operates together with those interests to fix reference (Putnam ms₁).

9. In some places Putnam makes the stronger charge that this view of taking reference as primitive is unintelligible. In "Models and Reality," Putnam says that this move is unintelligible only if you resist making certain kinds of moves that some do not resist (1983, 14). For example, Chisholm, following Brentano, contends that the mind has a faculty of referring to external objects which he calls Intentionality. The MR Putnam is criticizing, though, is naturalistically minded and would find the postulation of an unexplained mental faculty unhelpful epistemologically and almost certainly bad science as well (1983, 5).

10. Dummett is well known for his subscription to the dictum "a theory of meaning *is* a theory of understanding" (1975, 99; 1976, 69–70).

11. See Dummett 1978, xxxviii, where he articulates various worries about the plausibility of his antirealism.

12. Peter van Inwagen (1988) for reasons unspecified seems to think otherwise.

13. See LePore and Loewer 1987.

14. We would like to thank for comments on earlier drafts of this paper Bruce Aune, Michael Hand, Peter van Inwagen, Tim Maudlin, Eva Piccardi, Michael Root, and especially Hilary Putnam.

References

Ayer, A. J. 1936. *Language, Truth and Logic*. London.

Blackburn, Simon. 1971. "Moral Realism." In *Morality and Moral Reasoning*, edited by J. Casey. London.

Blackburn, Simon. 1984. *Spreading the Word*. Oxford.

Boghossian, Paul. "A Survey of Recent Work on Rule Following." Unpublished manuscript.

Brueckner, A. 1984. "Putnam's Model-Theoretic Argument against Metaphysical Realism." *Analysis* 44: 134–40.

Brueckner, A. 1986. "Brains in a Vat." *Journal of Philosophy* 83: 148–67.

Conee, Earl. 1987. "Hilary Putnam, *Reason, Truth and History*." *Nous* 21: 81–95.

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

Devitt, M. 1984. *Realism and Truth*. Princeton.

Dretske, Fred. 1981. *Knowledge and the Flow of Information*. Cambridge, Mass.

Dummett, M. 1973. *Frege: Philosophy of Language*. London.

Dummett, M. 1975. "What is a Theory of Meaning?" In *Mind and Language*, edited by S. Guttenplan. Oxford.

Dummett, M. 1976. "What is a Theory of Meaning? (II)." In *Truth and Meaning*, edited by G. Evans and J. McDowell. Oxford.

Dummett, M. 1978. *Truth and Other Enigmas*. Cambridge, Mass.

Dummett, M. 1981. *The Interpretation of Frege's Philosophy*. Cambridge.

Dummett, M. 1983. "Language and Truth." In *Approaches to Language*, edited by Harris and Roy. Oxford.

Field, Hartry. 1972. "Tarski's Theory of Truth." *Journal of Philosophy* 69: 347–75.

Field, Hartry. 1982. "Realism and Relativism." *Journal of Philosophy* 79: 553–67.

Fodor, J. 1987. *Psychosemantics*. Cambridge, Mass.

Glymour, C. 1982. "Conceptual Scheming or Confessions of a Metaphysicalist Realist." *Synthese* 51: 169–80.

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

Harman, G. 1982. "Metaphysical Realism and Moral Relativism." *Journal of Philosophy* 79: 568–74.

Kripke, Saul. 1981. *Wittgenstein on Rules and Private Language*. Oxford.

Leeds, Stephen. 1978. "Theories of Reference and Truth." *Erkenntnis* 13: 111–29.

LePore, E., and B. Loewer. 1987. "Mind Matters." *Journal of Philosophy* 84: 630–42.

Lewis, David. 1986. "Putnam's Paradox," unpublished manuscript.

Loewer, B. 1987. "From Information to Intentionality." *Synthese* 70: 287–317.

Mackie, J. L. 1977. *Ethica, Inventing Right and Wrong*. New York.

Pears, David. 1982. "Review of Putnam's *Reason, Truth and History*." *London Review of Books*.

Putnam, H. 1975. "The Meaning of 'Meaning'." In *Mind, Language and Reality*, Philosophical Papers. Vol. 2. Cambridge.

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

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

Putnam, H. 1983. *Realism and Reason*, Philosophical Papers. Vol. 3. Cambridge.

Putnam, H. "A Defense of Internal Realism." Unpublished manuscript[1].

Putnam, H. "A Note on Internal Realism," Unpublished manuscript[2].

Stalnaker, R. 1984. *Inquiry*. Cambridge, Mass.

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

van Inwagen, P. 1988. "On Always Being Wrong." This volume, pp. 95–111.

Wright, C. 1984. "Kripke's Account of the Argument against Private Language." *Journal of Philosophy* 81: 759–77.