JOSÉ L. ZALABARDO

PUTTING REFERENCE BEYOND BELIEF[*]

1. METAPHYSICAL REALISM

In 1976 Putnam publicly abjured the realist views that he had once championed. This notorious change of mind was prompted by a perceived difficulty with the picture of reference that Putnam saw at the heart of realism.[1] In "Realism and Reason" he referred to this picture as *Metaphysical Realism*. The label has later been used for a number of related views, but I shall reserve it here for the picture of reference that Putnam presented and attacked in that paper.

Metaphysical Realism (MR), Putnam tells us, "is less an empirical theory than a model – in the 'colliding billiard balls' sense of 'model' ".[2] It is an elusive model. Its essential commitment is that "there has to *be* a determinate relation of *reference* between terms in L and pieces (or sets of pieces) of THE WORLD".[3] But its distinctive, and, for Putnam, objectionable feature is not so much its commitment to the existence of this relation as the way in which the commitment is understood. For MR, "(1) the picture is supposed to apply to all correct theories at once [...]; and (2) THE WORLD is supposed to be *independent* of any particular representation we have of it".[4]

Some critics have objected to the use of the term 'realism' for a picture of reference. Realism, they protest, is a metaphysical doctrine, and the connection between issues in metaphysics and issues in the theory of reference is, at best, not transparent.[5] In particular, it is not clear that the metaphysical doctrine of realism, in its most interesting form, would have to be abandoned if MR were found untenable.[6] In this paper, I shall not be concerned with how metaphysical issues would be affected by the demise of MR. I shall consider instead whether there is a cogent argument against MR. I shall focus, in particular, on whether the picture is undermined

by a version of the so-called 'model-theoretic considerations' that Putnam has adduced against it.[7]

The debate on Putnam's 'model-theoretic argument' against MR has been remarkably one-sided. Ever since it was first advanced, a legion of critics have contended that the argument misses its mark – that it fails to put genuine pressure on the metaphysical realist picture of reference.[8] These attacks have gone largely unchallenged. Putnam himself has failed to address them in any detail,[9] and the few philosophers who have come in his defence have advanced construals of his argument according to which Putnam's target is not the metaphysical realist model itself, but the availability of a plausible substantive account of the reference relation that MR postulates.[10]

In this paper, I shall argue that the versions of Putnam's argument that his critics have attacked are indeed unsatisfactory. I shall then contend that the basic thought underlying Putnam's model-theoretic considerations can be articulated in a more promising way. I shall present an argument against MR that appears to incorporate this basic thought and doesn't fall prey to the objections levelled by Putnam's critics.

2. IDEAL BUT FALSE THEORIES

Let me start by considering the argument against MR actually advanced in "Realism and Reason" and "Models and Reality". In these versions of the argument, a crucial role is played by the thought that truth is radically non-epistemic, i.e. that a theory that is epistemically ideal could nevertheless be false. This thought, Putnam contends, is a consequence of MR, but MR cannot make sense of the possibility of an epistemically ideal, yet objectively false theory. Since MR has a consequence that it cannot make sense of, the picture is, Putnam concludes, incoherent.[11] Thus the argument has the form of a *reductio*. It rests on the claim that MR entails two contradictory propositions: that it has to be possible for an epistemically ideal theory to be objectively false and that it is not possible for an epistemically ideal theory to be objectively false. Neither entailment has gone unchallenged.

Critics have argued, first, that it doesn't follow from MR that it has to be possible for an epistemically ideal theory to be false.[12] It

is undeniable that MR is compatible with this view.[13] It may even be that MR has been traditionally associated with this claim – that most philosophers who have thought of reference along the lines of MR have also construed truth as a radically non-epistemic notion. But it is far from clear that the latter view is entailed by the former. In fact, it would be surprising if this entailment obtained. As LePore and Loewer put it, "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' ".[14]

These considerations suggest that there is room in principle for the following position: Reference consists in a relation between the terms of all theories and pieces of the representation-independent world, but the facts that single out this relation guarantee that an epistemically ideal theory will come out true. This is the view that, according to Lewis, Putnam should have endorsed:

Contra realist orthodoxy, truth *simpliciter* is equivalent to, or simply is, epistemic truth. That is not because there is anything epistemic about truth-on-an-interpretation. Nor is it because truth *simpliciter* is anything else than truth on all intended interpretations. Rather it is because the intendedness of interpretations is an epistemic matter.[15]

Thus, on this view, the reference-determining relation between terms and pieces of the world would be singled out epistemically, i.e. as the relation that makes the epistemically ideal theory come out true. As we shall see, there is a problem with the idea that a *unique* relation can be singled out in this manner. One may attempt to put pressure on MR on these grounds, but the resulting argument would be substantially different from anything actually found in "Realism and Reason" and "Models and Reality". Hence, even if Putnam succeeded in establishing that MR cannot make sense of the possibility of a false ideal theory, the argument wouldn't seem to undermine MR itself. It wouldn't follow, at any rate, that the view is, as Putnam claims, incoherent. All that would follow is that the proponent of MR would have to abandon another article of the realist credo – the view that ideal theories can be false.

Let me now turn to the other entailment – from MR to the impossibility of a false ideal theory. Putnam's argument for this entailment proceeds in two steps. He argues first that there is a way of assigning extensions in THE WORLD to the terms of an epistemically ideal theory such that if this correspondence were the reference-determining

relation, the theory would come out true. Then he goes on to argue that the proponent of MR cannot make sense of the claim that this correspondence is not the one that determines reference. Therefore, Putnam concludes, "the supposition that even an 'ideal' theory [. . .] might *really* be false appears to collapse into *unintelligibility*".[16]

The first claim is hardly controversial. For every consistent theory T and every set of object S of the right size, there is a model of T that has S as its universe. The terms of T can be interpreted on the elements of S in such a way as to make T come out true. Hence, in particular, an epistemically ideal theory will have a model with THE WORLD as its universe. The terms of the theory can be interpreted on the objects in THE WORLD in such a way as to make the theory true, subject only to the proviso that the theory has models of the size of THE WORLD. Notice, incidentally, that, as several commentators have pointed out, the only role that the Löwenheim-Skolem Theorem plays in the argument is to weaken this proviso. For it follows from the theorem that if THE WORLD has infinitely many objects, and the ideal theory has infinite models, the theory will have a model with THE WORLD as its universe, whatever infinite size THE WORLD may have.[17]

The second claim is much more problematic. Let SAT be an assignment of extensions in THE WORLD to the terms of an epistemically ideal theory that makes it come out true. According to Putnam, the proponent of MR cannot make sense of the claim that SAT is not the reference-determining relation. He argues for this conclusion by looking at the constraints that we can impose on the relation that does the job of determining reference. We could impose, he suggests, *theoretical constraints*, i.e. we could demand that it makes (an idealisation of) our theory of the world come out true.[18] But SAT, by hypothesis, satisfies this requirement. We could also impose *operational constraints*. In "Realism and Reason", these consist in making verified observational sentences come out true.[19] But SAT will also satisfy the operational constraints. For verified observational sentences will undoubtedly be part of an epistemically ideal theory, and hence, by hypothesis, will be made true by SAT.

Thus SAT satisfies theoretical and operational constraints. Hence, unless the proponent of MR can invoke further constraints, he won't be able to make sense of the idea that SAT is not the reference-

determining relation. And for Putnam this is precisely the situation. No other constraints could be legitimately imposed on the relation that determines the satisfaction conditions of terms.

For his critics, however, this restriction is totally unjustified. It amounts to requiring that the satisfaction conditions of terms have to be determined by the truth values of the sentences in which they figure.[20] But there seems to be room in principle for accounts of how satisfaction conditions are determined in which the truth conditions of sentences play only a limited role, or no role whatsoever. This would be the situation, in particular, if the relation that determines satisfaction conditions were a *causal* relation.

Let me use the label *externalism* for the view that the facts that determine which correspondence is the reference relation are at least partly independent of the truth values that each correspondence would assign to sentences or beliefs. So long as externalism is a live option, it would seem possible for the proponent of MR to maintain that SAT is not the relation that determines satisfaction conditions, and hence that the ideal theory could be false.

Putnam has offered several lines of argument to undermine externalism. If he succeeded in this endeavour, he would establish the entailment from MR to the impossibility of an epistemically ideal theory that is objectively false. Nevertheless, as we have seen above, this result wouldn't by itself force the proponent of MR to reject his picture of reference. He could save it if he were willing to give up the claim that the ideal theory could be false. We still haven't got an argument for the claim that if externalism had to be abandoned, MR would fall with it. I shall argue in the next section that such an argument can be found elsewhere in Putnam's writings. Then I shall turn to considering Putnam's reasons for rejecting externalism.

3. PERMUTATION

Thus, according to Putnam, the only legitimate way of singling out the relation that determines the satisfaction conditions of terms is with constraints on the truth values that would result for the sentences in which they figure. We could demand, e.g., that a successful candidate should make the ideal theory come out true. As we have seen, one problem with this proposal is that it makes the thought that

the ideal theory might be false totally unintelligible. But Putnam's restriction has another consequence that is more directly damaging for MR. The problem, highlighted in Chapter 2 of *Reason, Truth and History*, is that for any consistent theory T and any domain S of the right size, there will be, not just one, but many different assignments of extensions in S to the terms that figure in T, each of which would make T come out true. Thus, in particular, there will be many assignments of extensions in THE WORLD to the terms of the ideal theory that would make it come out true. This result is reached through the *permutation* argument. Given an assignment of extensions in THE WORLD to the terms of T that makes the theory come out true, any automorphism of the objects of THE WORLD would generate a different assignment that also satisfies this condition. If the world contains infinitely many objects, infinitely many assignments can be generated in this manner, all of which would make T come out true.[21] This means that, so long as Putnam's restriction is in place, we won't be able to single out a unique relation as the one that determines the satisfaction conditions of terms. But the existence of such a relation is the central tenet of MR. Therefore, if Putnam's restriction can be upheld, MR will have to be rejected.

We now have the argument we were looking for. If externalism had to be rejected, MR would have to go with it. This connection is now widely accepted,[22] but while Putnam expects it to undermine MR, most other writers have taken it the other way round, as lending support to externalism.[23] This is due to the fact that Putnam's arguments against externalism have carried very little conviction. The remainder of this paper will be devoted to exploring the possibility of a cogent argument against the externalist version of MR. I shall first take a brief look at three lines of argument that can be discerned in Putnam's writings. Then I shall articulate a different argument by developing a reinterpretation of the basic thought behind Putnam's rejection of externalism.

4. ARGUMENTS AGAINST EXTERNALISM

According to externalism, there is a relation, call it R, partly independent of the truth values of sentences, whose instantiation conditions determine the satisfaction conditions of each term. A term *t* is sat-

isfied by an object *a* if and only if *t* bears R to *a*. Putnam briefly considers this proposal, and dismisses it on the grounds that such an 'external' constraint would be nothing but a theoretical constraint under disguise.[24] Thus the externalist argues that the satisfaction relation is the one that pairs each term with the objects to which it is R-related. But for Putnam this is indistinguishable from a theoretical constraint: the relation that determines satisfaction conditions has to be such as to make the following sentence come out true: 'x is satisfied by y iff R(x, y)'.

Critics have objected that this reply rests on a sort of use-mention confusion.[25] It is one thing to single out the satisfaction relation as the one that pairs each term with the objects to which it is R-related, and quite another to single it out as the one that makes a sentence come out true. As Putnam has shown, many candidate satisfaction relations would make the sentence 'x is satisfied by y iff R(x, y)' come out true, but it doesn't follow from this that R isn't a perfectly determinate relation that each term bears to a unique class of objects.

Putnam has protested that, in raising this objection,

[. . .] the philosopher is ignoring his own epistemological position. He is philosophizing as if naive realism were true of him (or, equivalently, as if he and he alone were in an *absolute* relation to the world). What *he* calls causation really is causation, and *of course* there is a fixed, somehow singled-out, correspondence between the world and one definite relation in *his* case. Or so he assumes. But how this can be so was just the question at issue.[26]

As we shall see presently, a version of this thought may ultimately constitute a legitimate complaint against the metaphysical realist position. But, as it stands, Putnam's rejoinder seems to miss an important point. The externalist proposal is not trying to establish that terms have a determinate extension. Its goal is to specify *how* extensions are fixed. And the nature of the enterprise crucially affects the rules of the game. As Carsten Hansen has put it,

[i]n the context of giving an account of the means by which our terms acquire determinate extensions there can be nothing in assuming that the language in which the account is given has a determinate interpretation. No attempt to explain *how* extensions are fixed could fail by begging the question *whether* language has a determinate interpretation, and this is what Putnam's objection amounts to.[27]

A second line of argument is expressed by Putnam's protest: "It seems as if the fact that R *is* reference must be a *metaphysically unexplainable* fact, a kind of primitive, surd, metaphysical truth".[28]

The main difficulty with this thought is that if it were a fact that reference is R, it wouldn't seem to be more surd or metaphysically unexplainable than perfectly respectable facts such as the fact that water is H_2O. In fact, proponents of physicalistic theories of reference seem to conceive of their enterprise as analogous to any other scientific explication of a pre-theoretical notion.[29] Putnam sees a disanalogy. He writes:

> if there is a determinate physicalistic relation R [. . .] which just *is* reference [. . .], *this* fact [unlike the fact that water is H_2O] cannot itself be the consequence of our intentions to refer; rather, [. . .] it enters into determining what our very intentions to refer signify.[30]

But even if this is correct, and the fact that reference is R has to be construed as a brute metaphysical fact that transcends our referential intentions, this admission need not force the externalist to abandon his position. As Mark Heller puts it,

> [t]he mere bruteness of the fact should not intimidate the externalist. It seems likely to me that the externalist is going to have to accept some brute facts anyway. It seems to be a primitive, unexplainable fact that *modus ponens* is valid. Perhaps the same should be said of the laws of nature. Why should we not hold that the fact that R is reference [. . .] is also primitive?[31]

Another family of arguments advanced by Putnam target, not externalism as such, but rather the availability of a plausible account of the nature of the relation that, according to MR, links each term with the objects that satisfy it. These arguments grant that if a true sentence of the form 'x refers to y iff R(x, y)' were available, externalism could be vindicated, but attack the plausibility of the most promising strategies for instantiating this schema. They target, in particular, the attempt to reduce reference to causal-nomological notions.[32] I shall not be concerned with these arguments here. I believe that the difficulties faced by causal-nomological accounts of reference at this level are indeed formidable.[33] But these difficulties do not reveal what, if anything, is wrong with the externalist framework as such. In spite of recent attempts to construe the model-theoretic considerations as a challenge of this kind,[34] it is hard to deny that, for Putnam, the externalist version of MR rests on a fundamental mistake. My goal in what follows will be to develop an argument that would, if successful, vindicate this intuition.

5. A BETTER ARGUMENT?

In *Reason, Truth and History*, Putnam presents the intuitive thought behind his rejection of MR. As we have seen, in Chapter 2 of this book he tries to undermine MR on the grounds that "even if we fix the theory *and* fix the objects there are (if the number of objects is infinite) infinitely many *different* ways in which the same objects can be used to make a model for a given theory".[35] "This", he adds, "simply states in mathematical language the intuitive fact that to single out a correspondence between two domains one needs some independent access to both domains".[36] The problem with MR is that

[t]o pick out just *one* correspondence between words or mental signs and mind-independent things we would have already to have referential access to the mind-independent things. You can't single out a correspondence between two things by just squeezing *one* of them hard (or doing anything else to just one of them); you cannot single out a correspondence between our concepts and the supposed noumenal objects without access to the noumenal objects.[37]

Let's focus on concepts or mental representations.[38] The crucial difficulty for MR seems to arise from the fact that we cannot *single out* the correspondence that, according to MR, determines the satisfaction conditions of our concepts.

There is a way of understanding the notion of singling out a correspondence according to which Putnam's thought clearly fails to put pressure on externalism. Thus suppose that, by 'singling out a correspondence as the satisfaction relation', Putnam meant making it the case that that correspondence is the one that determines satisfaction conditions.[39] Then the externalist would agree with Putnam. We, with our theories, cannot make it the case that a certain correspondence is the one that determines satisfaction conditions. Reference facts, the externalist would urge, cannot arise from internal constraints alone.

If this is Putnam's reason for rejecting MR, the externalist seems to be right in thinking that he is over-reacting. That internal constraints don't fix reference does not entail that nothing does. Perhaps external constraints do. Nothing in this reading of Putnam's thought seems to preclude this possibility. But there is another way of understanding the notion of singling out a correspondence that results in a more promising line of argument. Suppose that by 'singling out

a correspondence as the satisfaction relation', Putnam didn't mean making it the case that that correspondence determines satisfaction conditions, but rather *grasping this fact*. On this construal of the notion, the passage quoted above would suggest a different difficulty for MR. Putnam would not be questioning directly the metaphysical realist thesis that reference facts are determined by a correspondence between all our concepts and things in the mind-independent world. On this reading, the problem would be that if this were the way in which reference facts are determined, it wouldn't be possible for us to grasp them – we wouldn't be able to grasp facts concerning which objects of the mind-independent world satisfy each of our concepts. In what follows, I want to explore the possibility of reading Putnam's arguments as aiming to establish this conclusion. I shall contend that his basic thoughts can be naturally construed as undermining the most plausible accounts of how we could grasp the referential connections postulated by MR. Let me emphasise that I am not making any claim to exegetical accuracy. My goal is not to be faithful to Putnam's own intentions, but to develop an argument against MR that seems to be suggested by some of the things he says.

It will save us some confusion to conduct the discussion in the third person, by considering the conditions under which a subject, call him Peter, would grasp the satisfaction conditions of his concepts. Peter's system of mental representation will include the resources of first-order logic plus a repertoire of general and individual concepts. I shall avoid committing myself to any specific account of the nature of these mental symbols, and stipulate that they are letters of the Latin alphabet – lower-case letters in the case of singular concepts, upper-case letters in the case of general concepts.[40] I shall form names for Peter's concepts and for the beliefs that can be formed with them by enclosing them in single quotation marks.[41] It will also help clarify matters to differentiate between Peter's representational resources and our own. For this purpose, I, unlike Peter, shall use letters of the Greek alphabet to refer to objects, properties and relations in the mind-independent world – lower-case letters for objects, upper-case letters for properties and relations.

I propose to assume that MR is the right account of the reference of Peter's concepts. Thus, I shall assume that there is a correspondence, Σ, that pairs each of Peter's concepts with the object or set of objects

by which it is satisfied. The question that we need to consider is how Peter would go about in trying to grasp the fact that each of his concepts is satisfied by the objects to which it is Σ-related.

6. DIRECT GRASP OF REFERENCE FACTS

What could grasp of reference facts consist in? One natural thought is that the way to grasp a fact is to form a belief that represents it.[42] On this account, all I need to do in order to grasp the fact that John is taller than Mary is to form a belief that represents the fact that John bears this relation to Mary. Similarly, all I would need to do in order to grasp the fact that 'John' refers to John is to form a belief that represents the fact that 'John' bears this relation to John.

We need to say a little bit at this point about the relation of representation between beliefs and facts. It is natural to suppose that this relation is closely connected with the truth conditions of beliefs. In order for a belief B to represent a fact F, the obtaining or otherwise of F will have to determine the truth value of B, i.e. B will have to be true if F obtains and false if it fails to obtain. It is widely admitted that this is a necessary condition for a belief to represent a fact. Treating it also as a sufficient condition raises familiar problems. But I propose to waive these and work with a truth-conditional account of which fact is represented by each belief, since my arguments will not trade on the controversial features of this account.[43]

Thus the proposal is that the way to grasp a reference fact is to form a belief that represents it. Let's consider how Peter could pursue this strategy. Let 'a' be one of Peter's singular concepts, and let α be the object to which it refers. According to this proposal, in order to grasp the fact that 'a' refers to α, Peter would have to form a belief that represents it, i.e., on my working construal of the notion, a belief that is true if 'a' refers to α and false if it doesn't. Similarly, let 'P' be one of Peter's monadic general concepts that is satisfied by the instances of a property, Π. In order to grasp the fact that 'P' is satisfied by the instances of Π, Peter would have to form a belief that is true if 'P' is satisfied precisely by these objects and false if it isn't.[44]

In order for Peter to be able to form these beliefs, we need to endow him with a minimal repertoire of semantic concepts, consist-

ing of two items. On the one hand, he will need a way of referring to his own concepts – i.e. a device of 'mental quotation' that enables him to form, for each of his concepts, a singular concept that refers to it. I have been using (single) quotation marks to form singular terms that refer to Peter's concepts. But just as I am differentiating between Peter's 'first order' referential devices (letters from the Latin alphabet) and my own (letters from the Greek alphabet), I propose to differenciate between his meta-linguistic referential devices and mine. For this purpose, instead of using quotation marks, Peter will refer to one of his concepts by enclosing it in square brackets. Thus '[c]' will be Peter's singular concept that refers to his own concept 'c'.[45] On the other hand, he will need a binary concept that is satisfied by a concept and an object or tuple just in case the object or tuple satisfies the concept. The same considerations that I have adduced to motivate the use of square brackets recommend the introduction of a conventional symbol to represent this binary concept. I shall use 'SAT' for this purpose.

With these resources at his disposal, Peter can form beliefs that represent the satisfaction conditions of his concepts. Thus let 'b' be a singular concept of Peter's which refers to α. Then the belief 'SAT([a], b)' will represent the fact that 'a' refers to α. Similarly the fact that 'P' is satisfied by the instances of Π will be represented by the belief '$(\forall x)(\text{SAT}([P], x) \leftrightarrow Q(x))$', where 'Q' is a monadic general concept that is satisfied by the instances of Π. I would like to consider next whether forming beliefs of this form would enable Peter to succeed in grasping the satisfaction conditions of his concepts.

7. THE VEIL OF CONCEPTS

In the passages quoted in Section 5 above, Putnam provides an explanation of why we can't single out a correspondence between concepts and objects. The reason is, in his view, that we have no *independent access* to objects. In "Realism and Reason" he briefly explains why objects are not independently accessible:

If concepts are particulars ('signs'), then any concept we may have of the *relation* between a sign and its object is *another sign*. But it is unintelligible, from my point of view, how the sort of relation the metaphysical realist envisages as holding

between a sign and its object can be singled out either by holding up the sign itself [. . .] *or* by holding up yet another sign [. . .].[46]

Putnam's thought can be construed as an attack on the proposal that we are considering – that a subject can grasp the fact that one of his concepts, C, is satisfied by a class of objects by forming a belief that represents this fact. On this construal, Putnam's objection to this proposal is that such a belief would use concepts in order to represent this fact about concept C. In particular, it would use concepts to single out the objects that it pairs with C as its satisfaction conditions.

Consider Peter's attempt to grasp the fact that 'a' refers to α by forming the belief 'SAT([a], b)'. This belief represents the fact that Peter is trying to grasp – the fact that one of his concepts bears a certain relation to an object. But it does so by singling out each of the relata, and indeed the relation itself, by means of concepts. In particular, the object that it pairs with 'a' as its referent is singled out by virtue of the fact that it is the referent of another one of Peter's concepts, namely 'b'. On my construal of Putnam's thought, the problem faced by this strategy for grasping reference facts arises from the involvement of concepts in the belief that represents the satisfaction conditions of a concept.

To see why this is a problem, we need to consider what makes us think that forming, say, the belief 'SAT([a], b)' would enable Peter to grasp the fact that 'a' refers to α. A natural explanation is that this belief would enable him to derive a grasp of the satisfaction conditions of 'a' from his grasp of the satisfaction conditions of 'b'. We normally take for granted that subjects have a grasp of the satisfaction conditions of the concepts that figure in their beliefs. Hence, if Peter can form the belief 'SAT([a], b)', we assume that he has a grasp of the satisfaction conditions of the concepts that figure in it, including 'b'. By forming this belief, Peter will generate a grasp of the fact that 'a' is satisfied by α out of his grasp of the fact that 'b' is satisfied by α. On this account, if Peter didn't antecedently grasp the fact that 'b' is satisfied by α, forming the belief 'SAT([a], b)' wouldn't enable him to grasp the fact that 'a' is satisfied by α. Hence this proposal as to how Peter can grasp the satisfaction conditions of 'a' presupposes that he has a grasp of the satisfaction conditions of 'b'.

This presupposition will have to be discharged sooner or later. We are looking for an account of how Peter can grasp the fact that each of his concepts is satisfied by the objects to which it is Σ-related. An account of how he can attain this goal cannot presuppose that he has grasp of the satisfaction conditions of some of his concepts. It has to explain how he can achieve grasp of the satisfaction conditions of all of them.

It is this need for generality that poses a difficulty for the proposal under consideration. As we have seen, by forming the belief 'SAT([a], b)', Peter can grasp the fact that 'a' refers to α, provided that he grasps the fact that 'b' refers to α. He could attain grasp of the latter fact by the same strategy. If 'c' is another one of Peter's concepts that refers to α, and he grasps the fact that it does, then he could grasp the fact that 'b' refers to α by forming the belief 'SAT([b], c)'.

But it seems obvious that this cannot be the general account of how Peter grasps the satisfaction conditions of his concepts. This method can only be used for deriving grasp of the satisfaction conditions of some concepts from grasp of the satisfaction conditions of others. In the end, a basic stock of concepts has to be reached whose satisfaction conditions Peter grasps by other means. Thus on the standard construal of his argument, Putnam's contention is that concepts can't in general obtain their satisfaction conditions from beliefs, as a belief that attempts to fix the satisfaction conditions of a concept will involve other concepts, whose satisfaction conditions will also have to be fixed. On the construal that I am recommending, the claim is that the satisfaction conditions of concepts cannot in general be grasped by means of beliefs, as a belief with which we try to grasp the satisfaction conditions of a concept will involve other concepts, whose satisfaction conditions we also need to grasp.

8. JUST MORE THEORY? I: TARSKIAN THEORIES OF REFERENCE

It may seem, however, that Peter could use this strategy to better effect. One could grant that it will take him nowhere to try to derive grasp of the satisfaction conditions of 'a' from grasp of the satisfaction conditions of other concepts, unless the chain ends with some concepts whose satisfaction conditions Peter grasps non-

derivatively. But instead of chasing grasp of the satisfaction conditions of concepts around his conceptual repertoire in this manner, couldn't he try to form a set of beliefs that enables him to grasp the satisfaction conditions of all his concepts at once?

One way of carrying out this strategy would be to form what I propose to call a *Tarskian theory of reference*. This theory contains, for every singular concept 'c', the belief 'SAT([c], c)', and for every monadic general concept 'C', the belief '$(\forall x)(SAT([C], x) \leftrightarrow C(x))$', as well as similar beliefs for each of his relational and functional concepts. I want to consider whether by forming this theory Peter would come to grasp the fact that each of his concepts is satisfied by the objects to which it is Σ-related. This would require that the theory represents this fact. On my construal of the notion of fact-representation, this would require, in turn, that the theory is true if the fact obtains, and false if it fails to obtain.

Peter's Tarskian theory of reference would seem to satisfy this constraint. Consider, e.g., the belief 'SAT([a], a)'. We are assuming that Σ is the satisfaction relation. One could argue that, on this assumption, if α is the object to which 'a' is Σ-related, Peter's belief 'SAT([a], a)' would be true if α satisfies 'a', and false otherwise. Similarly, if 'P' is Σ-related to the instances of property Π, the belief '$(\forall x)(SAT([P], x) \leftrightarrow P(x))$' would be true if 'P' is satisfied by the members of Π, and false otherwise. Thus, taken as a whole, Peter's Tarskian theory of reference would be true if each of his concepts is satisfied by the objects to which it is Σ-related, and false if these aren't the satisfaction conditions of Peter's concepts.

Let's suppose for the moment that this is correct – that Peter's Tarskian theory of reference is true if Σ is the satisfaction relation, and false if it isn't. I want to suggest that conceding this point doesn't force us to accept that Peter has now grasped the fact that Σ is the satisfaction relation. True, he now has the belief 'SAT ([a], a)', and we are assuming that this belief represents the fact that 'a' is satisfied by α. But we saw before that the natural explanation of why a belief of the form 'SAT([c_1], c_2)' would enable Peter to grasp the satisfaction conditions of 'c_1' is that such a belief would enable him to derive this grasp from his grasp of the satisfaction conditions of 'c_2'. Hence 'SAT([c_1], c_2)' would enable Peter to grasp the satisfaction conditions of 'c_1' only if he already grasped the

satisfaction conditions of 'c$_2$'. On this account of why beliefs of this form can bring about grasp of the satisfaction conditions of concepts, a 'disquotational' belief such as 'SAT([a], a)' would fail to discharge this task. For forming this belief would enable Peter to grasp the satisfaction conditions of 'a' only if he already grasped them. Thus Peter's Tarskian theory of reference may represent the fact that Σ is the satisfaction relation, but in his attempt to grasp this fact, the theory will be either unnecessary or ineffectual. If, on the one hand, Peter has no grasp of the satisfaction conditions of his concepts, forming this theory won't enable him to grasp them. If, on the other hand, he has grasped them already, the theory is no longer needed for the job. Either way, Peter's Tarskian theory of reference doesn't constitute grasp of the fact that Σ is the satisfaction relation.

The proponent of MR may protest that this conclusion has only been reached by sneaking through the back door extra conditions on when a subject can be said to have grasped a fact. I accepted some time ago that a subject can grasp a fact by forming a belief that represents it. Now I have concluded that Peter's Tarskian theory of reference wouldn't enable him to grasp the fact that Σ is the satisfaction relation, even after granting that the theory represents this fact. This outcome shows that somewhere along the way I have abandoned the account of grasp of facts with which I started, by adding an extra condition on when a subject can be said to grasp a fact.

The controversial addition arises from the requirement that if Peter is to grasp the satisfaction conditions of a concept C by forming a belief, he has to grasp already the satisfaction conditions of the concept in terms of which the belief singles out the objects that it pairs with C. Once this additional constraint is in place, forming a belief that represents the satisfaction conditions of C may not enable Peter to grasp them.

An externalist may want to resist this constraint. He could argue that it amounts to requiring that in order to grasp the fact that 'a' refers to α, Peter would have to form a belief that represents this fact *and* grasp the further fact that α is the object that the belief ascribes to 'a' as its referent. In general, the externalist would contend, I am requiring that in order to grasp a fact F, a subject has to satisfy two conditions:

(a)   Form a belief B that represents F.

(b)   Grasp the fact that B represents F.

It is the second condition that generates the problem. For it entails that grasping a fact requires grasping another fact. Hence any attempt to grasp a fact, semantic or otherwise, will be rendered impossible by the ensuing regress.

The externalist could try to avoid the difficulty by rejecting condition (b), and maintaining that all one needs to do to grasp a fact is to form a belief that represents it. In particular, Peter could grasp the fact that Σ is the satisfaction relation by forming a set of beliefs that represents this fact. Hence, if Peter's Tarskian theory of reference represents the fact that Σ is the satisfaction relation, forming this theory is all he needs to do to grasp this fact. Just as an externalist account of reference saved MR from the standard construal of Putnam's attack, an externalist account of grasp of facts may save it now from the version of the argument that I am recommending.

This proposal can be attacked on two fronts. On the one hand, the account of grasp of facts that it puts forward seems extremely implausible. It is true that we normally consider forming a belief B that represents a fact F sufficient for grasping F. However, I want to suggest that this is because we assume that subjects have a grasp of what facts are represented by their beliefs, not because we think that no such grasp is necessary. But the proponent of MR cannot appeal to this assumption to save grasp of semantic facts from my challenge. For it is hard to see how a subject could grasp what fact is represented by one of his beliefs unless he had a grasp of the satisfaction conditions of the concepts that figure in it. But I am arguing precisely that MR would render grasp of satisfaction conditions impossible. In this situation, the proponent of MR would be begging the question if he invoked the assumption that we normally grasp what state of affairs is represented by each of our beliefs.

The externalist proposal amounts to the claim that the existence of a relation between a mental representation and a fact suffices for the subject that forms this mental representation to grasp the fact, even if the subject has no grasp of which fact the mental representation is related to, let alone of the nature of the relation. This would be a very striking feature for mental representations to have. Certainly non-mental representations don't exhibit it. The gauge pointing in

a certain direction may represent the fact that the tank is full. But seeing the direction in which the gauge points would only enable you to grasp the fact that the tank is full if you grasped another fact – that the position of the gauge represents the tank being full. The relevance of the illustration is underscored by the fact that, in the externalist picture, the link between the representation and the fact that it represents is as dependent on factors external to the representation in the case of a belief as in the case of a petrol gauge. No doubt there are important differences between seeing a petrol gauge and forming a mental representation. But if the externalist account could only be saved by an unsupported appeal to the *sui generis* character of mental representation, it would be fair to describe the view, adapting Putnam's epithet, as a magical theory of grasp of facts.[47]

Notice that the proponent of MR cannot support the externalist account of grasp of facts on the grounds that we seem perfectly capable of grasping facts, and only the externalist proposal can accommodate this possibility. I am not denying that we can grasp facts, semantic or otherwise. My claim is rather that no such grasp would be possible if MR were the correct account of reference. The pressure that this result would put on MR cannot be deflected by insisting that grasp of facts has to be possible. If accepting an implausible account of what this grasp amounts to were the only way to save it, perhaps we would be compelled to bite the bullet. But insofar as grasp of facts can also be saved by giving up MR, the proponent of this picture of reference is in no position to defend the externalist account of grasp of facts as the only way of rendering this grasp possible. In the last section I shall address briefly the question whether giving up MR might enable us to provide a more satisfactory account of grasp of reference facts.

9. BLIND SPOTS

Let me now move on to consider a second difficulty with this proposal. Even if we accept the externalist account of grasp of facts, there are reasons for doubting that Peter's Tarskian theory of reference would enable him to grasp the fact that Σ is the satisfaction relation. For it is far from clear that the theory succeeds in representing this fact. The claim that it does rests on the idea that the

disquotational theorem corresponding to each concept represents the fact that it is satisfied by the objects to which it is Σ-related. Thus, if 'P' is Σ-related to the instances of Π, since Σ is the satisfaction relation, the belief '$(\forall x)(\text{SAT}([P], x) \leftrightarrow P(x))$' would represent the fact that 'P' is satisfied by the instances of Π.

I want to offer a line of reasoning that seems to cast doubt on this conception of what facts are represented by beliefs of this kind. I am assuming that beliefs come to represent facts by virtue of their truth conditions. A belief, B, represents a fact, F, just in case B is true if F obtains and false if it fails to obtain. But beliefs acquire their truth conditions by virtue of the satisfaction conditions of the concepts that figure in them. Hence what fact is represented by a belief is ultimately determined by the satisfaction conditions of the concepts that figure in it. Thus given that predicate 'P' is satisfied by the instances of property Π, and singular term 'a' refers to object α, the belief 'Pa' will be true if α is an instance of Π, and false if it isn't. By having those truth conditions, the belief comes to represent the fact that $\Pi(\alpha)$.

The same account can explain how one of Peter's beliefs can represent the satisfaction conditions of a concept. Thus given that (a) his concept 'Q' is satisfied by the instances of Π, (b) 'SAT' is satisfied by all pairs $\langle x, y \rangle$ such that x satisfies y, and (c) '[P]' is satisfied by 'P', the belief '$(\forall x)(\text{SAT}([P], x) \leftrightarrow Q(x))$' will be true if 'P' is satisfied by the instances of Π, and false if these aren't the satisfaction conditions of the concept. Hence the belief represents the fact that 'P' is satisfied by the instances of Π.

Let's consider now the situation for a disquotational theorem, such as '$(\forall x)(\text{SAT}([P], x) \leftrightarrow P(x))$'. In light of the preceding discussion, this belief would represent the fact that 'P' is satisfied by the instances of Π just in case the obtaining of this fact would make the belief true, and its failure to obtain would make it false. However, this is not a correct description of the truth conditions of the belief. If the fact obtains, the belief is true, but its failure to obtain wouldn't make the belief false. The reason is not that the belief would still be true in these circumstances. The reason is rather that it doesn't make sense to ask what truth value that situation would yield for the belief, given the satisfaction conditions of the concepts that figure in it. For the situation in which 'P' is not satisfied by the instances of Π

is one in which the concepts that figure in the belief wouldn't have the satisfaction conditions that they actually have. Notice that the proponent of MR cannot avoid this difficulty by evaluating the belief by reference to the satisfaction conditions that its concepts would have in that situation. For so long as 'SAT' and '[P]' keep their satisfaction conditions, the belief will continue to be true, whatever satisfaction conditions 'P' happens to have.

In order for a belief to represent the fact that 'P' has certain satisfaction conditions, the satisfaction conditions of the concepts that figure in the belief have to bestow on it a truth value for each possible state of affairs in regard of the satisfaction conditions of 'P' – it has to be true if 'P' has the satisfaction conditions that the belief represents it as having, and false otherwise. But a disquotational belief cannot fulfil this requirement. The satisfaction conditions of its concepts are incapable of bestowing a truth value on the belief for all the satisfaction conditions that 'P' could have. They cannot determine the truth value that the belief would have if the satisfaction conditions of 'P' were other than what they are, as 'P' is one of the concepts whose satisfaction conditions are supposed to effect this determination.

We can think of the satisfaction conditions of the concepts that figure in a belief as a 'scanning device' that surveys all possible states of affairs and bestows on the belief a truth value for each of them. It is thanks to this device that a belief can represent a fact. If a belief B is made true by all states of affairs in which F obtains and false by all those in which F fails to obtain, we say that B represents F. The problem with a disquotational belief is that the scanning device constituted by the satisfaction conditions of its concepts has 'blind spots'. It fails to assign a truth value to the belief for all states of affairs in which the concept that it mentions has satisfaction conditions other than the actual ones. This shortcoming renders the belief incapable of representing the fact that a concept has certain satisfaction conditions, even if we grant that the concepts that figure in it have, as a matter of fact, determinate satisfaction conditions.

This means that even if we accept the externalist account of reference, we can resist the claim that Peter's Tarskian theory represents the satisfaction conditions of his concepts. For even if we grant that Σ is the satisfaction relation, the theory won't represent the fact that

this is so. Therefore, even if we accept the externalist account of grasp of facts, Peter's Tarskian theory of reference won't enable him to grasp the satisfaction conditions of his concepts.[48]

10. JUST MORE THEORY? II: FIELDIAN THEORIES OF REFERENCE

Let me consider now another version of this proposal. A disquotational theory is not the only kind of theory of reference that Peter could adopt. Let's suppose that Peter's conceptual repertoire includes a binary concept 'R' that is satisfied by a concept and an object or tuple just in case the former bears relation Σ to the latter. Then Peter could form a theory of reference along the lines of the proposal that Putnam attributes to Field,[49] by forming the belief

(F)   '$(\forall x)(\forall y)(\text{SAT}(x, y) \leftrightarrow R(x, y))$'

Let's consider whether forming this belief would enable Peter to grasp the fact that Σ is the satisfaction relation.

As before, even if we grant that (F) represents the fact that Σ is the satisfaction relation, forming this belief won't enable Peter to grasp this fact – according to the conception of grasp of facts that I have advocated. The theory would enable him to grasp the satisfaction conditions of his concepts by deriving this grasp from his grasp of the satisfaction conditions of 'R' – the concept with which the theory singles out the pairs connected by the satisfaction relation. But we are looking for a general account of how Peter could grasp the satisfaction conditions of *all* his concepts. An explanation of how he can derive from grasp of the satisfaction conditions of some of his concepts grasp of the satisfaction conditions of all the rest does not fit the bill. The general account that we are looking for cannot presuppose grasp of the satisfaction conditions of any concept, but unless we presuppose that Peter grasps the satisfaction conditions of 'R', his Fieldian theory of reference won't enable him to grasp the satisfaction conditions of the rest of his concepts.

In light of this outcome, it may be tempting once more to resort to an externalist account of grasp of facts, according to which all one needs to do to grasp a fact is to form a belief that represents it. But independently of the plausibility of this account, there remains an issue as to whether (F) succeeds in representing the fact that Σ is

the satisfaction relation. I want to argue that the considerations that block this conclusion in the case of a Tarskian theory of reference would have the same effect for its Fieldian counterpart.

I argued in the previous section that in order to represent the fact that Σ is the satisfaction relation, the satisfaction conditions of the concepts that figure in (F) would have to confer on this belief the following truth conditions: (F) would have to be true if the fact obtains and false if it fails to obtain. If Σ is the satisfaction relation, (F) would indeed be true. And unlike in the previous case, some alternative hypotheses as to the identity of the satisfaction relation would make (F) false. Thus suppose that 'R' were satisfied by the pairs that instantiate Σ, but some other of Peter's concepts failed to be satisfied by the objects to which they bear Σ. In those circumstances, the satisfaction conditions of the concepts that figure in (F) would make the belief false. But there still remain other hypotheses for which the satisfaction conditions of the concepts that figure in (F) would fail to confer a truth value on it – namely if the satisfaction relation failed to pair 'R' with the instances of Σ. These hypotheses constitute blind spots for the satisfaction conditions of the concepts that figure in (F).

Hence, as in the case of a Tarskian theory, a Fieldian theory of reference fails to be related to the fact that Σ is the satisfaction relation in the way that would enable it to represent this fact. For even though the obtaining of the fact would make the theory true, its failure to obtain would not suffice in general for making the theory false. The inability to represent the fact that Σ is the satisfaction relation doesn't afflict disquotational theories alone. The problem arises for every attempt to specify the satisfaction conditions of a class of concepts C with beliefs that include some of the concepts in C. And since any specification that Peter can provide of the satisfaction conditions of all his concepts is bound to exhibit this feature, the difficulty will thwart all his attempts to form a belief that represents the fact that his concepts are satisfied by the objects to which they are Σ-related.[50] Hence, we have to conclude that the identity of the reference relation postulated by MR cannot be grasped by forming beliefs that represent it.

11. IMPLICIT THEORIES OF REFERENCE

I turn now to a different proposal as to how Peter could grasp the fact that Σ is the satisfaction relation. There is another way in which beliefs can generate grasp of reference facts. Even if we presuppose no grasp of the satisfaction conditions of any of his concepts, there are some reference facts that Peter can be said to grasp thanks to his beliefs. Suppose that his theory of the world contains the following (true) belief: '$(\forall x)(P_1(x) \to P_2(x))$', where '$P_1$' is satisfied by the instances of property Π_1, and 'P_2' by the instances of property Π_2. If he grasps the satisfaction conditions of one of the concepts that figure in it, this belief would enable Peter to grasp reference facts concerning the other concept. Provided that he grasps the satisfaction conditions of 'P_1', forming this belief would enable him to grasp a fact about the satisfaction conditions of 'P_2' – namely that it is satisfied (at least) by all the instances of Π_1. Similarly, if he grasps the satisfaction conditions of 'P_2', the belief will enable him to grasp the fact that 'P_1' is satisfied only by instances of Π_2. Thus, he grasp of reference facts that the belief would effect in this manner presupposes a pre-existent grasp of other reference facts.

There is, however, another reference fact which Peter could be said to grasp by forming this belief even if we presupposed no grasp of the satisfaction conditions of 'P_1' or 'P_2': namely the fact that all the objects that satisfy 'P_1' (whatever they are) also satisfy 'P_2'. In general, by forming a belief, Peter can be said to grasp the fact that the satisfaction conditions of the concepts that figure in it are such as to make the belief true. This suggests a way in which Peter can use his beliefs to grasp the identity of the satisfaction relation. According to this proposal, Peter could grasp the fact that Σ is the satisfaction relation by forming a theory of the world that would be true only if the satisfaction conditions of his concepts were the ones that Σ ascribes to them.

Obviously, not any theory would satisfy this constraint. On the one hand, Peter's theory of the world would fail to satisfy the constraint if it contained false beliefs. For, since the satisfaction conditions of his concepts *are* the ones that Σ ascribes to them, a false theory is made false by these satisfaction conditions. On the other hand, Peter's theory of the world would fail to satisfy the constraint if it were incomplete – if there were beliefs that could be formed with his

concepts such that neither they nor their negations were contained in the theory. For in this case there would be more than one assignment of satisfaction conditions to his concepts that would render his theory true. Thus, suppose that Peter's theory of the world contains neither the belief '$(\forall x)(P_1(x) \to P_2(x))$' nor its negation. Then if his theory is made true by a correspondence that pairs 'P_1' with a subset of the set of objects with which it pairs 'P_2', it will also be made true by another correspondence that differs from this one only in this respect. Hence a true but incomplete theory of the world would not enable Peter to grasp the fact that Σ is the satisfaction relation, but only the fact that this role is played by one of the correspondences in a range that includes Σ, but also other candidates.

It may seem, however, that a true complete theory of the world would enable Peter to grasp the fact that Σ is the satisfaction relation. If this suggestion were correct, it would be possible after all to grasp the identity of the satisfaction relation by forming beliefs. But the proposal falls prey to Putnam's permutation argument. If Peter's theory of the world is true, the satisfaction conditions that Σ ascribes to his concepts will make it true. But even though it is a complete theory, Σ is far from being the only correspondence that assigns satisfaction conditions to Peter's concepts in such a way as to make his theory come out true. Any automorphism on the objects of THE WORLD would enable us to obtain from Σ another correspondence that would also make the theory come out true. Hence even a complete theory will exhibit the shortcoming that we found in incomplete theories. It wouldn't enable Peter to grasp the fact that Σ is the satisfaction relation, but only the fact that this role is played by one of the correspondence within a range that includes, along with Σ (if the theory is true), many other correspondences.

This result wouldn't be altered if Peter's true, complete non-semantic theory of the world were supplemented with a Tarskian theory of reference for his concepts. Supplementing his theory of the world in this way won't improve Peter's situation. For the semantically enriched theory will be made true by the same assignments of satisfaction conditions as the original theory.[51] Hence, as before, we would be forced to conclude that the new theory only enables Peter to grasp the fact that the satisfaction relation is one of the

correspondences within a range that includes many other candidates besides Σ.

What if Peter supplemented his theory of the world with a Fieldian theory of reference? On the face of it, this enlargement of Peter's theory would not be as ineffectual as the previous one. A non-semantic theory plus a Tarskian theory of reference would be made true by all the assignments of satisfaction conditions that make the non-semantic theory true. Adding a Fieldian theory of reference to a non-semantic theory would not have this disappointing result. Some of the assignments of satisfaction conditions that make the non-semantic theory true would make the non-semantic theory *and* the Fieldian theory come out false. However, the addition of a true Fieldian theory of reference to a true, complete non-semantic theory will still fail to be made true by Σ alone. There will remain several other assignments of satisfaction conditions that also make the expanded theory come out true. They will include every correspondence X that makes the non-semantic theory come out true and is such that X(x, y) iff X('R', <x, y>). Σ satisfies this constraint, but once again any automorphism on the objects of THE WORLD will yield another correspondence that also satisfies it.

Each version of this proposal is undermined by the fact that "even if we fix the theory *and* fix the objects there are (if the number of objects is infinite) infinitely many *different* ways in which the same objects can be used to make a model for a given theory". The satisfaction relation cannot be singled out by a theory as the correspondence that would make the theory come out true. For even if it is true and complete, and even if it incorporates a theory of reference for its own concepts, there will always be several correspondences that make it come out true.

12. IMMEDIACY

One could try to side-step these difficulties by denying that our grasp of reference facts always has to be mediated by concepts. According to this proposal, a belief that ascribes to a concept C its satisfaction conditions could single out the objects that it pairs with C without the intervention of concepts. By what other means could a belief single out a class of objects? The answer provided by the empiricist

tradition appeals to the direct, unmediated perceptual acquaintance with objects that results from acts of ostension.

The most direct difficulty for this proposal arises from its limited range of applicability. In some version of the idea, we only enjoy this kind of immediate access to our sense impressions. In others, mid-size physical objects in our surroundings are also accessible in this manner. But wherever one draws the line, the privileged range of objects is bound to be a relatively insignificant fraction of the class of objects with which our concepts can be referentially connected. And grasp of whether our concepts are satisfied by objects outside the privileged range will have to be derived from grasp of their satisfaction conditions within the range.

The standard account of how this derivation proceeds appeals to the links established by our theory of the world between the satisfaction of concepts by observable and unobservable objects. By virtue of these links, our theory is supposed to generate grasp of the referential connections that our concepts bear to unobservable objects from our directly acquired grasp of their referential connections to observable ones. It does so by ruling out the correspondences between our concepts and unobservable objects that would render the theory false, given the facts concerning which observable objects satisfy each of our concepts. Hence, on this account, we would be able to grasp the identity of the satisfaction relation by forming a theory that would only be made true by the correspondence that plays as a matter of fact this role.

But this strategy is invalidated by the difficulties that I considered in the previous section. For even if the referential connections between our concepts and observable objects are fixed, there will always remain many assignments of unobservable objects to our concepts that would make a given theory true. In particular, if a theory is made true by Σ, it will also be made true by any other correspondence that results from Σ by an automorphism on the unobservable objects of THE WORLD.[52]

It is not clear, in any case, that referential links to ostensible objects should be treated as unproblematic. For it is hard to see how we could grasp which object is singled out by an ostensive gesture without producing a mental representation of it. And the task of grasping the satisfaction conditions of this representation would

seem to face the same difficulties that I have considered for concepts in general.[53] If grasp of the objects singled out by ostensive gestures is also mediated by mental representations, then these become more grist for the permutation mill. Forming a belief such as 'This is a pen' would only enable me to grasp the following reference fact – that the object to which the demonstrative concept refers is one of the objects that satisfies the concept 'pen'.

13. NECESSARY TRANSCENDENCE

Let's suppose, in any case, that the conclusion cannot be resisted – that the proponent of MR is forced to accept that if his picture of reference is correct, we cannot grasp reference facts. Would the acceptance of this outcome force him to abandon MR? This is a complicated question. For it may turn out that, in spite of this corollary, MR is still the best account of reference to which we can aspire. I shall not try to reach a verdict on this matter here, but I would like to end by considering two ways in which the proponent of MR might try to make this outcome more palatable than it may seem.

First, he could contend that this corollary is not as disastrous as one may think. There is nothing wrong, he would argue, with reference relations of which we have no grasp. In fact, the standard anti-individualist thought experiments seem to suggest that the phenomenon is widespread.[54] Someone who doesn't have the concepts of modern chemistry may not be able to represent the respects in which H_2O differs from XYZ. But someone with this conceptual deficit can perfectly well refer to H_2O, and not to XYZ by his concept 'water'. The proponent of MR could argue that a subject in this situation has no grasp of the fact that his concept 'water' refers to H_2O, and not to XYZ. But this is the situation in which we find ourselves with respect to many of our concepts. Hence, quite independently of MR, we would seem forced to conclude that our concepts often bear referential relations to the world of which we have no grasp.

This observation, the proponent of MR would point out, makes it easier to accept the consequence that I have drawn from his position. Even if we reject MR, we would have to accept that subjects often have no grasp of the referential connections between their concepts

and the world. In light of this fact, that MR forces us to accept this situation across the board doesn't seem a formidable liability.

Let's grant that the proponent of MR is right in his interpretation of anti-individualist thought experiments. I want to suggest that the consequence that I have drawn from MR is much harder to accept than any limitation on our grasp of reference facts that one could derive from anti-individualist considerations. What makes MR unappealing is not just the number of concepts for which this situation would have to be accepted. There is a crucial disanalogy between the situation faced by a chemically illiterate subject with respect to his concept 'water' and the situation that Peter would face with respect to each of his concepts if MR were correct. The disanalogy lies in the fact that the former situation is in principle remediable. The chemically illiterate subject could come to grasp the fact that his concept refers to H_2O, and not to XYZ, if he acquired the concepts of modern chemistry. Armed with these concepts, he would be capable of forming a belief that would rule out XYZ as a referent of his concept 'water'.

In this respect, Peter's situation is totally different. As we have seen, the permutation argument enables us to obtain a correspondence, call it Σ^*, that assigns the same truth values as Σ to every belief that can be formed with Peter's concepts. If Σ and Σ^* assigned different truth values to beliefs formed with concepts that Peter doesn't yet have, then acquiring these concepts would put him in a position to rule out Σ^*. But this is not the end of the story, as, for *any* set of concepts that Peter could acquire, and any theory that he could form with them, the permutation argument would enable us to find a correspondence that would assign the same truth value as Σ to every belief in the theory.

The transcendence of reference facts that follows from MR is not a contingent shortcoming due to the specific cognitive repertoire available to a subject at a certain time. MR makes reference facts necessarily transcendent. No improvement in the representational tools at Peter's disposal will enable him to form a theory that is made true by Σ alone. There will always remain other assignments of satisfaction conditions that would also make his theory come out true.

14. REJECTING MR

Let me now turn to a second way in which the proponent of MR may attempt to save his picture of reference in the face of the argument that I have developed. He could agree that the transcendence of reference facts is an intolerable result, but question the claim that his picture of reference is to be blamed. According to this line of thought, the problem that I have raised would be inherent to the notion of reference. Nothing short of abandoning the notion would make the problem go away.

Naturally, the problem can only be shrugged off in this manner if there is no account of reference for which it doesn't arise. And this is by no means a foregone conclusion. I am not going to try to present a satisfactory alternative here, but I would like to end by pointing in the direction of an approach to reference that may succeed in avoiding this difficulty. I am going to illustrate this approach by looking at the semantics of arithmetical terms. In his classical paper, "What Numbers Could Not Be", Paul Benacerraf attacked the idea that a reduction of arithmetic to set theory would reveal the real nature of numbers. According to the view that Benacerraf attacked, numbers are sets – the elements of a certain set-theoretic sequence. This proposal can also be presented as an account of the reference of arithmetical terms. Numerals, on this account, would refer to the sets of a certain sequence, and the satisfaction conditions of arithmetical predicates, relation symbols and function symbols would also be defined on this sequence of sets.

Consider now Jane, a subject who has arithmetical concepts, but no set-theoretic concepts at all. In fact, we can suppose that the concepts of arithmetic are all the extra-logical concepts Jane has. I want to suggest that, if the set-theoretic account of the reference of her concepts is correct, Jane's attempts to grasp their satisfaction conditions will face similar difficulties to those that Peter encountered with his concepts. Beliefs such as ' "1" refers to the successor of 0' won't do the trick. This belief would pair her concept '1' with the right set, but it wouldn't enable her to grasp the reference of '1' unless she grasped already which set is referred to by the complex concept 'the successor of 0'. As in Peter's case, beliefs of this kind would only enable her to grasp the satisfaction conditions of her

concepts if she already grasped the satisfaction conditions of some of them.

Trying to form arithmetical beliefs that are only made true by the right interpretation won't get her very far either, even if we waive the difficulties involved in having a body of beliefs that is not recursively axiomatisable and suppose that her beliefs constitute a complete arithmetical theory. For there are several different ways of interpreting arithmetical concepts in set-theoretic terms all of which would make her theory come out true. A complete arithmetical theory would only enable her to grasp the fact that the satisfaction conditions of her concepts are determined by one of the correspondences within this range. If her concepts refer to sets, Jane, with her arithmetical concepts and beliefs, cannot grasp their satisfaction conditions.

I want to suggest that a different account of the reference of her concepts would make the difficulty disappear. Notice that the reason why she faces a problem is that the identity of the objects in the range of the satisfaction relation is not completely determined by which of her concepts apply to them. Her beliefs specify that her singular concepts refer to the elements of a progression, but they leave it open which specific progression is the relevant one. Obviously, this situation generates a problem only if we can meaningfully ask whether Jane's concepts refer to this or that progression. The problem would disappear if we maintained, instead, that Jane's concepts refer, not to these or those objects exemplifying the abstract structure of a progression, but rather to locations in the abstract structure itself.[55] On this account, in order to identify the referent of '3', Jane doesn't need to find out whether it is this or that set. She only has to decide where in the abstract structure it stands with respect to the referents of the rest of her concepts. She has to determine, e.g., that it bears the relation denoted by '>' to the referents of '0', '1' and '2', that it is the image under the function denoted by '+' of the referents of '1' and '2', etc. But this is precisely what is achieved by her beliefs '0 < 3', '1 < 3', '2 < 3', '1 + 2 = 3', etc. On this account, acquiring true arithmetical beliefs is all one needs to do to identify the referents of arithmetical concepts. If Jane could have all true arithmetical beliefs, there would be nothing she didn't know about the referents of her concepts.

A similar approach to the reference of Peter's concepts may enable him to avoid his difficulties. According to this approach, the identity of the referents of Peter's concepts is exhaustively determined by which beliefs involving these concepts are true. On this account, the range of the satisfaction relation consists of locations in the abstract structure defined by the true theory of the world involving these concepts. Hence, in order to grasp which objects satisfy one of his concepts, Peter only needs to grasp where in the abstract structure these objects stand with respect to those that satisfy the rest of his concepts. But this is precisely what he would achieve by acquiring true beliefs.[56]

This approach to reference faces very serious problems. To mention just one, it would make it hard to explain how a subject can change his body of first order beliefs and take his new beliefs to concern the same subject matter as the old ones, or how two subjects could take their concepts to have the same satisfaction conditions even though they disagree about the truth values of beliefs involving them. I am not endorsing this approach. My goal in sketching it has been to argue that the difficulty that I have raised in this paper is not intrinsic to the notion of reference. Giving up metaphysical realism might enable us to avoid the difficulty. Perhaps, if the view I have sketched is the only alternative, learning to live with transcendence may be seen as the lesser evil. But we shouldn't presuppose that our options are so severely limited. As Putnam himself puts it, echoing Strawson,

we are constantly being asked to choose between *metaphysical* positions on the one hand and *reductionist* positions on the other, and what is terribly difficult (but what makes the game of philosophy worth the candle) is to show that the metaphysical mystery is not the only alternative to the simplistic position of the reductionist.[57]

NOTES

[1] Cf. *Realism and Reason*, p. viii and "Sense, Nonsense, and the Senses", pp. 459–460.

[2] "Realism and Reason", p. 123.

[3] Ibid., p. 125.

[4] Ibid.

[5] See, e.g., M. Devitt, "Realism and the Renegade Putnam", pp. 291–293.

[6] Cf. C. Hansen, "Putnam's Indeterminacy Argument", pp. 78–79, C. McGinn, "Ideal Justifications", D. Lewis, "Putnam's Paradox", pp. 231–232, and M. Heller, "Putnam, Reference, and Realism", p. 118.

[7] Cf. H. Putnam, "Realism and Reason", "Models and Reality" and *Reason, Truth and History*, Ch. 2.

[8] Cf. e.g., G. H. Merrill, "The Model-Theoretic Argument Against Metaphysical Realism", M. Devitt, "Realism and the Renegade Putnam", A. L. Brueckner, "Putnam's Model-Theoretic Argument Against Metaphysical Realism", D. Lewis, "Putnam's Paradox", M. Heller, "Putnam, Reference, and Realism", and J. van Cleve, "Semantic Supervenience and Referential Indeterminacy".

[9] He briefly takes issue with his 'externalist' critics in the Introduction to *Realism and Reason*. Cf. pp. x–xii. See also "Beyond Historicism", pp. 295–296.

[10] Cf. e.g., D. L. Anderson, "What Is the Model Theoretic Argument?" and B. Hale and C. Wright, "Putnam's Model-Theoretic Argument Against Metaphysical Realism: A Short Tour".

[11] Cf. "Realism and Reason", pp. 125–126, and "Models and Reality", pp. 471–476.

[12] See J. Koethe, "Putnam's Argument against Realism", for an interesting discussion of this issue.

[13] E. LePore and B. Loewer, "A Putnam's Progress", p. 462.

[14] E. LePore and B. Loewer, op. cit., pp. 471–472.

[15] D. Lewis, "Putnam's Paradox", p. 232.

[16] H. Putnam, "Realism and Reason", p. 126.

[17] Cf., e.g., D. Lewis, "Putnam's Paradox", p. 229.

[18] In *Reason, Truth and History* Putnam refers by 'theoretical constraints', not to the scientific theory itself, but rather to general methodological principles. Cf. p. 31.

[19] In "Models and Reality", operational constraints receive a slightly different construal. An extension assignment satisfies operational constraints if it assigns to each observational term in the language the observable objects that satisfy it (pp. 472–473). But it will always be possible to find an extension assignment that fulfils this requirement and makes the ideal theory true – since an epistemically ideal theory will only contain true observational sentences. This version of the notion had appeared already in J. A. Winnie, "The Implicit Definition of Theoretical Terms". Cf. §12 below.

[20] Cf. van Cleve, "Semantic Supervenience", pp. 354–355.

[21] Putnam makes, in fact, the stronger claim that infinitely many assignments of extensions in S to the terms that figure in T would yield the same truth values for the sentences of T *in every possible world*. Cf. *Reason, Truth and History*, pp. 32–38, 217–218.

[22] Cf., e.g., D. Lewis, "Putnam's Paradox", p. 226.

[23] Cf. e.g., A. L. Brueckner, "Putnam's Model-Theoretic Argument Against Metaphysical Realism", p. 139: "the argument seems to force us to adopt some

or other account of reference according to which extra-linguistic facts help fix reference".

24 Cf. "Realism and Reason", p. 127, "Models and Reality", p. 477, and *Reason, Truth and History*, pp. 45–46.

25 Cf., e.g., C. Glymour, "Conceptual Scheming", p. 177.

26 *Realism and Reason*, p. xi.

27 C. Hansen, "Putnam's Indeterminacy Argument", p. 88. Cf. also B. Hale & C. Wright, "Putnam's Model Theoretic Argument Against Metaphysical Realism: A Short Tour", p. 14.

28 *Reason, Truth and History*, p. 47. Cf. also *Realism and Reason*, p. xii and "Beyond Historicism", p. 296.

29 Hartry Field's "Tarski's Theory of Truth" is a prime example of this attitude.

30 *Reason, Truth and History*, p. 47.

31 M. Heller, "Putnam, Reference, and Realism", pp. 125–126. See also A. L. Brueckner, "Putnam's Model-Theoretic Argument Against Metaphysical Realism", p. 138.

32 Cf. Putnam, "Model Theory and the Factuality of Semantics", and *Renewing Philosophy*, Ch. 3.

33 In addition to Putnam's texts cited in the previous footnote, cf. P. Boghossian, "Naturalizing Content" and J. Zalabardo, "A Problem for Information Theoretic Semantics".

34 Cf. D. Anderson, "What Is the Model-Theoretic Argument?"

35 P. 74.

36 Ibid.

37 P. 73.

38 It is clear that Putnam is thinking of concepts as mental representations or signs, not as abstract entities that the mind somehow grasps and whose satisfaction conditions are intrinsic to them.

39 Cf. Brueckner, "Putnam's Model-Theoretic Argument Against Metaphysical Realism", p. 137, for an explicit endorsement of this reading of Putnam's argument.

40 The argument does presuppose, however, that mental representation requires mental symbols. This assumption is widespread, and Putnam himself, in *Reason, Truth and History*, considers it unavoidable: "[. . .] if we assume that we have no 'sixth sense' which enables us to *directly* perceive extra-mental entities, or to do something analogous to perceiving them ('intuiting' them, perhaps), then 'grasping' an intension, or any extra-mental entity, must be mediated by representations in some way" (p. 27). This is not, however, his current view. In his Dewey Lectures of 1994, he urges us to understand that "giving up the idea of representations as interfaces requiring a 'semantics' is not the same thing as giving up on the whole idea of representation" ("Sense, Nonsense, and the Senses", p. 505).

41 Quotation marks are standardly used to form names of linguistic expressions, but it seems natural to extend their use in the way that I suggest, once we think of concepts and beliefs as mental signs, akin in role and structure to terms and (assertoric) sentences of a language. The use of quotation marks has been extended in this way, among others, by Jerry Fodor. Cf., e.g., his *Psychosemantics*, Ch. 4.

42 It may be more natural to construe the notion of grasp of a fact as incorporating also a component of justification. Be this as it may, for the purposes of this paper

grasp of facts will not require justification. As will emerge, the difficulties that I shall raise have nothing to do with Peter's ability to justify claims about the reference of his concepts. Neither do they concern specifically his ability to grasp the referential states of affairs *which actually obtain*.

[43] Cf. J. Fodor, *Psychosemantics*, pp. 72 ff. for a defence of this model as the core of the notion of mental content.

[44] The notion of grasp of the satisfaction conditions of a general concept needs some clarification. There is an ambiguity in "the fact that 'P' is satisfied by the instances of Π". It can denote either the fact that 'P' bears a certain relation to a class of objects (the instances of Π) or the fact that it bears a certain relation to a property (Π). The ambiguity is important for our purposes, as Peter could grasp one fact and not the other. Thus he could grasp the fact that the instances of Π are the objects that satisfy 'P', but fail to grasp that they are instances of Π. Similarly, he could grasp the fact that Π is the property whose instances satisfy 'P' but have no grasp of which objects instantiate Π. I propose to work with a very liberal notion of grasp of satisfaction conditions. Peter will count as grasping the satisfaction conditions of 'P' if he grasps the connection between the concept and the objects that satisfy it or if he grasps the connection between the concept and a property (Π or any other) whose instances are, as a matter of fact, the objects that satisfy 'P'. Grasp of the fact that 'P' is satisfied by the instances of Π will be meant as incorporating all these possibilities. Similar remarks will apply to grasp of the fact that Peter's concepts are satisfied by the objects to which they are Σ-related.

[45] The introduction of a second kind of quotation device is a potential source of confusion. In this connection, it may help to reflect that the same situation arises if we speak (in English) about German expressions involving quotation, as German uses angle brackets the way English uses quotation marks. Thus 'Schnee' refers to snow, but '<Schnee>' refers to 'Schnee'. Similarly, speaking about Peter's concepts, we can say that 'a' refers to α, but '[a]' refers to 'a'.

Peter's square brackets differ from our quotation marks in that the former generate mental representations, whereas the latter generate English singular terms. Nevertheless, I don't intend to exploit this contrast in my argument, and my motivation for using square brackets as Peter's quotation symbols is not to highlight it, but to keep Peter's referential devices separate from mine. I am indebted on these points to an anonymous referee.

[46] Pp. 126–127.

[47] Putnam speaks of magical theories of reference in *Reason, Truth and History*, pp. 3 ff. The magic is particularly striking in the case of the disquotational theorems of a Tarskian theory of reference. On this account, someone with no previous grasp of the satisfaction conditions of a concept 'P' would come to grasp them by forming a belief, '$(\forall x)(SAT([P], x) \leftrightarrow P(x))$', that pairs 'P' with a class of objects singled out as those that satisfy 'P', the very concept whose satisfaction conditions he is trying to grasp.

[48] This is not to say that a Tarskian theory of reference would be completely useless. It would play the important role of enabling Peter to grasp the satisfaction conditions of the satisfaction concept 'SAT', provided that he grasps the satisfaction conditions of his non-semantic concepts. What it cannot do is enable him to grasp the satisfaction conditions of these.

[49] Cf. Putnam, *Reason, Truth and History*, p. 45. Field put forward his proposal in "Tarski's Theory of Truth".

[50] This line of thought underlies Frege's puzzling contention that any attempt to define truth is bound to be circular. Cf. "Thoughts", p. 4. It also grounds Quine's claim that it is meaningless to question the reference of all the terms of our all-inclusive theory. Cf. "Ontological Relativity", pp. 51 ff.

[51] More accurately, by extensions of the original assignments that take account of the expansion of the language.

[52] Cf. Putnam, "Models and Reality", pp. 472–474 and J. A. Winnie, "The Implicit Definition of Theoretical Terms', pp. 225–227.

[53] Cf. Putnam's remark: "the reference of gestures is just as problematic as the reference of terms, if not more so" ("Models and Reality", p. 476). The point I am making is that grasping the reference of gestures may be just as problematic as grasping the reference of concepts.

[54] Cf. H. Putnam, "The Meaning of 'Meaning'" and T. Burge, "Individualism and the Mental".

[55] I take this proposal to be congenial to the account of numbers advocated in "What Numbers Could Not Be". "For arithmetical purposes", Benacerraf writes, "the properties of numbers which do not stem from the relations they bear to one another in virtue of being arranged in a progression are of no consequence whatsoever" (pp. 69–70). And later: "To *be* number 3 is no more and no less than to be preceded by 2, 1, and possibly 0, and to be followed by 4, 5, and so forth" (p. 70).

[56] Cf. Curtis Brown's suggestion: "Perhaps Putnam means to simply make it true by definition that reference is determinate. Perhaps he simply intends to use the word 'object' in such a way that knowing the truth conditions of sentences *does* suffice to determine what objects those sentences are about" ("Internal Realism: Transcendental Idealism?", p. 153).

[57] Putnam, *Realism and Reason*, p. xvi.

REFERENCES

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

Benacerraf, P. (1965): 'What Numbers Could Not Be', *Philosophical Review* 74, 47–73.

Boghossian, P. A. (1991): 'Naturalizing Content', in G. Rey and B. Loewer (eds.), *Meaning in Mind: Fodor and His Critics*, pp. 65–86, Cambridge, MA: Blackwell.

Brown, C. (1988): 'Internal Realism: Transcendental Idealism?', in P. A. French, T. E. Uehling and H. K. Wettstein (eds.), *Midwest Studies in Philosophy, Volume XII: Realism and Antirealism*, pp. 145–155, Minneapolis: University of Minnesota Press.

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

Burge, T. (1979): 'Individualism and the Mental', in P. A. French, T. E. Uehling and H. K. Wettstein (eds.), *Midwest Studies in Philosophy, Volume IV: Studies in Metaphysics*, pp. 73–121, Minneapolis: University of Minnesota Press.

Devitt, M. (1983): 'Realism and the Renegade Putnam: A Critical Study of *Meaning and the Moral Sciences*', *Noûs* 17, 291–301.

Field, H. (1972): 'Tarski's Theory of Truth', *Journal of Philosophy* 69, 347–375.

Fodor, J. A. (1987): *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, Cambridge, MA: The MIT Press.

Frege, G. (1977): 'Thoughts', in P. T. Geach and R. H. Stoothoff (trans.), *Logical Investigations*, pp. 1–30, New Haven: Yale University Press.

Glymour, C. (1982): 'Conceptual Scheming or Confessions of a Metaphysical Realist', *Synthese* 50, 169–180.

Hale, B. and C. Wright (1994): 'Putnam's Model-Theoretic Argument Against Metaphysical Realism: A Short Tour', unpublished.

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

Heller, M. (1988): 'Putnam, Reference and Realism', in P. A. French, T. E. Uehling and H. K. Wettstein (eds.), *Midwest Studies in Philosophy, Volume XII: Realism and Antirealism*, pp. 113–127, Minneapolis: University of Minnesota Press.

Koethe, J. (1979): 'Putnam's Argument Against Realism', *Philosophical Review* 88, 92–99.

LePore, E. and Loewer, B. (1988): 'A Putnam's Progress', in P. A. French, T. E. Uehling and H. K. Wettstein (eds.), *Midwest Studies in Philosophy, Volume XII: Realism and Antirealism*, pp. 459–473, Minneapolis: University of Minnesota Press.

Lewis, D. (1984): 'Putnam's Paradox', *Australasian Journal of Philosophy* 62, 221–236.

McGinn, C. (25 November 1983): 'Ideal Justifications', *Times Literary Supplement*, p. 1307.

Merrill, G. H. (1980): 'The Model-Theoretic Argument Against Realism', *Philosophy of Science* 47, 69–81.

Putnam, H. (1975): 'The Meaning of "Meaning"', in K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science, Vol. VII: Language, Mind, and Knowledge*, pp. 131–193, Minneapolis: University of Minnesota Press.

Putnam, H. (1978): 'Realism and Reason', reprinted in *Meaning and the Moral Sciences*, London: Routledge & Kegan Paul, pp. 123–140.

Putnam, H. (1980): 'Models and Reality', *Journal of Symbolic Logic* 45, 464–482.

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

Putnam, H. (1983): *Reason and Reason: Philosophical Papers, Volume 3*. Cambridge: Cambridge University Press.

Putnam, H. (1983): 'Beyond Historicism', in *Realism and Reason: Philosophical Papers, Volume 3*, pp. 287–303.

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

Putnam, H. (1992): *Renewing Philosophy*, Cambridge, Mass.: Harvard University Press.

Quine, W. V. O. (1969): 'Ontological Relativity', in *Ontological Relativity, and Other Essays*, New York: Columbia University Press, pp. 26–68.

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

Winnie, J. A. (1967): 'The Implicit Definition of Theoretical Terms', *British Journal for the Philosophy of Science* 18, 223–229.

Zalabardo, J. L. (1995): 'A Problem for Information Theoretic Semantics', *Synthese* 105, 1–29.

Philosophy Department
University of Birmingham
Edgbaston Birmingham B15 2TT
England