Referential Relativity

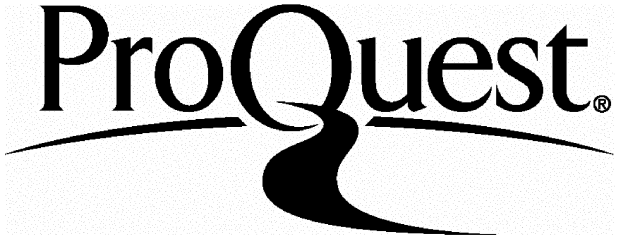
Alex Broadbent, University College London

Thesis submitted for MPhil

28, 397 words (92 pages)

"The method of 'postulating' what we want has many advantages; they are the same as the advantages of theft over honest toil."

Bertrand Russell



Abstract

This essay concerns the possibility that our words do not refer in the way that we normally take them to. I explore arguments presented by Quine, in *Ontological Relativity*, and the so-called "model-theoretic" arguments of Putnam's in the late seventies and early eighties. My aim is not to add a new position to the existing pantheon, but to add argument. In particular I defend Quine and Putnam against the most common justification for the dismissive attitude with which these arguments are sometimes met.

I start by arguing that the works of Quine and Putnam mentioned are very closely linked – in fact, they have at least one argument in common. This is of relevance because many philosophers who are scathing of Putnam are less scathing of Quine. I argue that much of what goes for one goes for the other, on this topic.

I describe in detail the shared argument: its essence is the claim that we could permute the referents of all our terms while still preserving the truth-values or truth-conditions of all our sentences. Then I set out the common challenge. The challengers say that a possible source of determinate reference has been ignored. Putnam (and Quine) only consider constraints imposed on reference by the use we make of our terms; in particular, they focus heavily on the empirical constraints that the reference relation must meet. Objectors say that there might be another kind of constraint – an *external* constraint – that acts (partially) independently of our experience and activities. I present David Lewis's version of this challenge.

I acknowledge this as a loophole in the Quine-Putnam argument. Moreover it can be developed into a consistent position. I attack not its consistency but its plausibility. I suggest that no empirical justification is available for the claim that such constraints act on reference. I suggest further that no *a priori* justification has been advanced. Thus I argue that utilising the loophole requires making a claim that is, in fact, unjustified: it is a piece of speculative metaphysics. I conclude by suggesting a way to avoid this.

Acknowledgements

I am very grateful to my supervisor, José Zalabardo, for his excellent guidance and help, and also for helping me get a grip on the technical issues presented in Chapter 1, where I rely extensively on his *Introduction to the Theory of Logic*.

For Zeynep, my love.

Contents

Introduction

This is an essay about the possibility that our words do not determinately refer in the way that we normally take them to. I present arguments to this effect from the work of Quine and Putnam (in Chapter 2). These arguments seem to me both extremely important and widely neglected. Part of my purpose, therefore, quite aside from whether the arguments work, is to convince the reader that there is more to be said about them than many philosophers seem to suppose. But I also think that the arguments do work. The focus of this essay is therefore to defend these arguments against a loophole exploited by writers like Lewis, Heller and Devitt, who present what I call the externalist challenge. (Presenting and deflecting this challenge is the task of Chapter 3.) This leaves us with a form of mild relativism, likened by Quine to relativity theory in physics, and in Chapter 4 I interpret and endorse Quine's position in *Ontological Relativity*. However that discussion is not extensive. For what I seek to add to the debate is argument, rather than another new position. The question of what position we are to adopt if we accept the Quine-Putnam arguments about reference is not a question I claim to have settled. Rather, it is a question I claim we should discuss. For, by arguing for the failure of the externalist challenge, I argue it is a question we face.

Chapter 1

The Framework

The purpose of this short chapter is to lay out clearly the logical terminology and results which will feature in the remainder of the work. At this stage I neither propose nor deny any philosophical significance for these results. Clarity is the present aim: the philosophy starts in the next chapter.

For our purposes, we have what might be called an extensional surrogate of the various features of our language. Properties are represented by sets. Relations are also represented by sets, but as sets of tuples. Thus what properties and relations do for the predicates of a language, will on our picture be done by sets.

The notion of a language to which the results to be discussed apply involves three components. Consider some language[1] L, consisting of the following three sets. The first is a set V of terms, the extralogical vocabulary of L. The second is the set of *formulas* of L. A formula is a tuple of terms and other symbols – the logical symbols. Not every tuple of elements of V and logical symbols will be a formula. The final ingredient of L is the set of *sentences,* a subset of the set of formulas. Not every formula is a sentence: this fact models the restrictions of grammar, in a natural language such as English, on what counts as a sentence. A sentence is a formula with no *free variables*. A *variable* is a part of the logical vocabulary of the language (i.e. not in V) that indicates which argument place is affected by the quantifiers. A *free* variable is one that is not bound by a quantifier.

Another notable point is that the rules of English grammar may not be precise nor fully determinate; thus there may be room for indeterminacy as to what counts as a sentence of English. One characteristic of formal

[1] Throughout, I am considering only first-order languages. It is true that some of the results I discuss do not hold for second-order languages. However, it is very uncommon for anyone to suggest that the problems of this work are solved by any appeal to second-order languages. For one thing, it is debatable whether second-order logic is really logic. But more importantly, the claims I will shortly make concerning indiscernibility of isomorphic structures still hold for second-order language (though the Lowenheim-Skolem Theorem does not). These claims are the logical underpinning of the arguments examined in subsequent chapters. I therefore restrict my discussion of language to first-order languages throughout: for "language", read "first-order language".

languages is that this indeterminacy has been removed. In other words, the three sets – of terms, formulas and sentences – that form the language have been fully specified.

I do not, in this chapter, propose to address the question of how results (such as those to be presented) obtained in the study of formal languages bear on natural languages. That question will be subsumed under the more general issue previously deferred, of discerning the philosophical significance of these results. For the philosophical issues later discussed arise in relation to natural, not formal, languages. It should become clear, however, that the difference alluded to here between natural and formal languages will not provide an instant solution to these philosophical difficulties. For the central issue to be discussed is how reference can be absolute or determinate; and it seems *prima facie* unlikely that a natural language whose sets of terms, formulas and sentences are not fully determinate will have a better chance of referential determinacy than a formal language, which is determinate in those other respects. More precisely, it seems unlikely that a natural language will have an advantage *by virtue of* its lack of determinate sets of terms, formulas and sentences. This is merely by way of provisionally licensing the discussion to come. I do not deny that an argument may be built on the difference between natural and formal languages; however, most such arguments (and certainly the ones we will examine) do not rely on the difference I have mentioned, but on other differences (e.g.: causal links between language and world; the involvement of thoughts; historical patterns of use). Thus there is no shortcut.

A *structure* $\mathscr{A}$ for a first-order language L consists of two things. The first is a set, A, called the *universe* of $\mathscr{A}$. The second is a function pairing each term in the extralogical vocabulary of L with its *interpretation* in $\mathscr{A}$. The interpretation in $\mathscr{A}$ of an L-term is an element (for singular terms) or a subset (for general terms) of A. If the term in question is a singular term a, we say that the element of A with which it is interpreted in $\mathscr{A}$ is the *denotation* in $\mathscr{A}$ of the L-term. The interpretation of an n-place L-predicate

P in $\mathscr{A}$ is an n-place relation in A. We say that a structure is *infinite* when its universe is an infinite set (that is, a set with infinitely many elements).

On this extensional picture of language, *truth* is a function from the interpretation of each L-sentence in some structure $\mathscr{A}$ to the truth-values – to the set {T,F}, where we assume only that T and F are not the same object. This approach to truth makes it structure-relative, since the function (we will call it v) is defined on a structure. A *model* of a sentence is a structure on which that sentence is true. Likewise, a model of a set of sentences is a structure on which all those sentences are true. I.e. some L-structure $\mathscr{A}$ is a model of some L-sentence Φ just in case $v_{\mathscr{A}}(\Phi) = T$; $\mathscr{B}$ is a model of some set Γ of L-sentences just in case $v_{\mathscr{B}}(\Gamma) = T$.

One theorem that we shall see appealed to is the *model-existence theorem*: Every consistent set of sentences has a model. In other words, for every set of sentences that does not syntactically entail a contradiction, there is some structure on which all those sentences come out true. At first sight this result might look surprisingly metaphysical, amounting to the claim that there will be sufficient resources in the world to find objects corresponding to each term, sets for each predicate, and so on - *whatever* we say, so long as we are consistent. You would have thought that, if the world was small, you might run out of things. However the oddness goes away when we remember abstract objects like numbers, and in particular on the reflection that the vocabulary of a language might be the universe for a structure of a set of sentences of that language – so that each term denotes itself, for instance.

We say that a set Γ of sentences *represents* a class C of L-structures if and only if all and only models of Γ are members of the class C. So if Γ is a set of sentences which has models not in C, Γ does not represent C. And again if C includes any structures which are not models of Γ, then Γ does not represent C. When the truth value of every sentence in L interpreted on two L-structures $\mathscr{A}$, $\mathscr{B}$ is the same, we say that $\mathscr{A}$ and $\mathscr{B}$ are *indiscernible*. It immediately follows that if $\mathscr{A}$ is a structure indiscernible from a structure $\mathscr{B}$ and $\mathscr{A}$ is a model of Γ, then $\mathscr{B}$ will also be a model of Γ. And bearing this in

mind, it follows from our definition of representation that if $\mathscr{A}$ is in a class C of L-structures and $\mathscr{B}$ is not, then Γ will not represent C.

We say that two structures are *isomorphic* just in case we can define a function – an *isomorphism* – between their universes, such that the function is one-to-one, and the image under the function of every term and predicate of L interpreted in one structure is the interpretation of that term or predicate in the other structure. Isomorphic structures of a given language can be shown to be indiscernible: that is, they assign the same truth value to every sentence of that language.

From these definitions of representation and isomorphism, it is clear that many classes of structures of a language will not be representable: there will be no set Γ of L-sentences that represents C. For since isomorphic structures assign the same truth values to every L-sentence, if $\mathscr{A}$ is a model of Γ and is in C, and $\mathscr{B}$ is isomorphic to $\mathscr{A}$ and is outside C, then $\mathscr{B}$ will also be a model of Γ. Thus Γ will have models outside C, so Γ will not represent C. However, for certain purposes we are happy to live with this, and to talk about whatever properties (of which isomorphism may be one) certain classes of structures have in common. This attitude is predominant in mathematics. Rarely do mathematicians try to identify what the structures they study "really" are, at least not without straying from their discipline and donning their philosophical hats.

However, the situation concerning languages such as English is more difficult. English is suited and used to make empirical claims. So is the more specialised language of empirical science. However, identifying up to isomorphism the models of the English sentence, "Yesterday I had an omelette for tea", will make that sentence represent many surprising structures. For instance, it might represent a structure whose universe consisted entirely of protons, or of numbers, or of a certain octopus in the Mediterranean. Yet if I honestly utter this sentence, then it would be surprising to claim that I am representing all of these things and more. And the same goes, of course, for the more complex and ambitious sets of sentences comprising our best empirical theory: in fact, some of them may

end up representing my omelette. On the other hand, if we abandon the notion of representation up to isomorphism, then we are faced with the prospect that our empirical sentences don't, on our current definition of *represent*, represent the classes of structures (involving omelettes or octopi, respectively) which we thought they did, prior to our current definition of *represent*. For the class of structures involving omelettes and me eating them that I had hoped to represent, will not contain all the models of the aforementioned sentence ("Yesterday I had an omelette for tea"). There will be other structures outside that class that are also models of the sentence: thus that class is not represented by that sentence. When we seek to represent some class C of L-structures with an L-sentence Φ, we speak of the model(s) of Φ in C as the *intended model(s)* of Φ.

This is a problem related to the way in which the terms of our language are defined. I will describe two sorts of definition here. *Explicit definition* introduces a new term to a language by identifying, in terms of the old language, the object or set which the new term is to denote. (E.g. a "sprog" is a particularly springy dog; or "resultant force" is a quantity equal to the product of an object's mass and its acceleration.) If terms are introduced by explicit definition, and if the classes of structures represented by every sentence in the introducing language have been fixed, then clearly isomorphism will present no problem for our ability to represent structures in the new language. For the class of structures represented by a sentence Φ containing the new term will be whatever class of structures Φ would represent if the specified term(s) of the old language were substituted for the new term.

However, the usefulness of explicit definition is subject to a limitation. It does not enable us to go beyond the old language. This means that if the old language is empty, i.e. if there is no old language, then we can't get started. But more relevantly in the present context, it means that we can't say anything in the new language that we couldn't say in the old: for we can't use the new language to represent any classes of structures other than those we could already represent using the old. So for example, explicit definition of new terms to create a new language can't enable us to

represent structures with dinosaurs in their universes unless our old language could also represent structures with dinosaurs in their universes.

Implicit definition involves specifying in the old language the *truth conditions,* of every sentence of a given form, of the new language featuring the new term. That is, an implicit definition supplies necessary and sufficient conditions (in terms of the old language) for the truth of every new-language sentence containing the new term. The definition of the axioms of set theory have this form. With one important qualification, implicit definition does allow us to represent new structures with new terms, since sentences containing the new terms are not restricted to representing structures that are already representable with sentences of the old languages. The qualification is that sentences containing implicitly defined terms can only represent classes of structures up to isomorphism. For whether or not the classes of structures represented by all sentences of the old language had been fixed, sentences containing the new term would not inherit the same structures. They would simply inherit truth-conditions: necessary and sufficient conditions for their truth. And this leaves open the possibility of finding isomorphic structures to the intended model(s) of the new sentence, that are also models (since isomorphic to some model) but are not intended models (since not in whatever class of structures we are considering).

These facts will become relevant when we find it claimed that isomorphic structures are very easy to come by for most interesting structures. For example, most would agree that the physical world is an interesting structure. It is not difficult to exhibit structures of the language of our best physical theory that are isomorphic to the physical world, but different in many ways that you might not at all be inclined to accept. Whether there are unintended models of a given set of sentences will of course depend on what class those sentences are intended to represent. However, there will always be at least one model provided that the sentence is consistent. The business of showing that there are unintended models for a given set of sentences thus comes down (as we will see) to showing that there are isomorphic structures to the intended model, and that these structures are not intended models.

The role of the Lowenheim-Skolem Theorem in this area is less than has sometimes been claimed. There are two halves to the Lowenheim-Skolem Theorem. The Downward Theorem claims that when a set of sentences has infinite models, it also has countable models. The Upward Theorem claims that when a set of sentences has infinite models, it has infinite models of arbitrarily high cardinalities. The effect of these combined is simply that sentences with infinite models cannot represent those models even up to isomorphism. For clearly there cannot be a one-to-one correspondence between two sets of different cardinalities, and an isomorphism is a one-to-one correspondence.

This becomes relevant to our purposes if our empirical theories are thought to have infinite models. It may be that the physical universe is infinite. But even if it turns out not to be, our scientific theory includes the number system, which must have an infinite model if it has any model. So the relevance of the Lowenheim-Skolem Theorem in the present case is to guarantee that there will be non-isomorphic unintended models of our empirical theory. But even without such a guarantee, we shall see that it is not difficult to show that there are such unintended models. The Lowenheim-Skolem Theorem is of more interest in mathematics, where, as mentioned, the aim is commonly considered to be the identification of models up to isomorphism. I have already argued that in empirical theory, even identification of models up to isomorphism is not a very satisfactory achievement. In empirical theory, *any* kind of indiscernibility will be unacceptable, regardless of whether the models in question are isomorphic. This is why the Lowenheim-Skolem Theorem is of only limited relevance.[2]

These ideas are the battlefield for the next two chapters. A suspicion is forgivable that the ideas necessary for that battle could actually be presented without even this much logical apparatus or model-theory[3]. Be that as it may, the debate has taken this form.

---

[2] For a fuller and similarly negative discussion of the role of this Theorem, see *Putnam's Paradox* (Lewis, 1984).

[3] Again, see *Putnam's Paradox* (Lewis, 1984) for a similar view more extensively discussed.

Chapter 2

Quine and Putnam

This chapter is long, but the thread should not be difficult to keep hold of. I begin by presenting Quine's views on reference as he explains them in *Ontological Relativity*. The second section deals with Putnam's rather infamous "model-theoretic" arguments against metaphysical realism. The purpose of the final section is to compare the two, partly for the intrinsic interest of the comparison and partly in order to establish that there are clear parallels and key shared assumptions between the two. This paves the way for the next chapter, where we will see these assumptions challenged.

1. Ontological Relativity

Quine's thesis of ontological relativity needs to be distinguished from the background thesis of indeterminacy of translation. Here is a concise statement of the latter:

> The infinite totality of sentences of any given speaker's language can be so permuted, or mapped onto itself, that (*a*) the totality of the speaker's dispositions to verbal behaviour remains invariant, and yet (*b*) the mapping is no mere correlation of sentences with *equivalent* sentences, in any plausible sense of equivalence however loose.
>
> (Quine, 1960: 27)

The important point for present purposes is that indeterminacy of translation deals with sentences, and correspondingly with meanings. The arguments for ontological relativity turn on an argument about terms (that is, parts of language that are smaller than sentences), and correspondingly about reference. Compare the following statement of ontological relativity:

> What makes ontological questions meaningless when taken absolutely is [...] circularity. A question of the form "What is an *F*?" can be answered only by recourse to a further term: "An *F* is a *G*." The answer makes only relative sense: relative to the uncritical acceptance of "*G*."
>
> (Quine, 1969: 53)

The present work is concerned with ontological relativity, not directly with the thesis of indeterminacy of translation, nor directly with the further thesis of indeterminacy of meaning. Nevertheless, an indeterminacy claim about translation is one ingredient in Quine's recipe for ontological relativity, so we will have occasion to discuss certain aspects of it.

However, before discussing the argument for ontological relativity, it is important to notice another slightly curious background aspect of the argument. That is the contrast between *theories* and *language*. The first third of *Ontological Relativity* mentions only language. Theories make their first appearance on page 50 (1969: 50), with no fanfare. Thereafter they are the focus of attention, and language drops out of the picture. Quine does a little to explain this, but not much. He says:

To talk thus of theories raises a problem of formulation. A theory, it will be said, is a set of fully interpreted sentences. [...] But if the sentences of a theory are fully interpreted, then in particular the range of values of their variables is settled. How then can there be no sense in saying what the objects of a theory are?

(Quine, 1969: 51)

This indicates that Quine is aware of the switch. But he treats it as an obstacle. His solution is simply "[...] that we cannot require theories to be fully interpreted, except in a relative sense, if anything is to count as a theory" (1969: 51). This still leaves unexplained why he bothered making the switch from talk of language to talk of theories. I suggest that the reason is as follows. Quine makes clear (1969: 35) that his arguments about the indeterminacy of translation have an impact on both meaning and reference. But as long as the arguments are at the level of language, a suspicion remains that they don't go very far towards establishing ontological results. This is because speaking a language need not be seen as adopting an ontology. English speakers do not, by their use of English, necessarily commit themselves to the existence of any entities. At least, a further argument would be needed by anyone who wanted to claim that they do. So establishing that English speakers might unwittingly use the same terms to refer to different objects – which follows from the indeterminacy of translation[4] – would not establish that the question, what there is, is relative or meaningless or anything of the sort. It would only establish that different English speakers might give different answers and mean the same thing, or vice versa.

Theories, on the other hand, could potentially be seen to involve ontological commitment. One obvious view to take on the nature of theory, is that a theory says something about the world. If this is so, then the claim that the terms of the theory do not really refer stands in direct opposition. It is incumbent upon Quine to say something about what a theory is, and particularly whether we should think of them as bearers of truth value, if he will deny that their terms refer.

[4] If this is not immediately clear, see Quine's discussion of intension and extension on page 34-5 of *Ontological Relativity*. He says, for instance, "The indeterminacy of translation now confronting us... cuts across extension and intension alike" (1969: 35).

Thus the switch to theories is an essential move. By talking only about language, he is open to the objection that he is pushing an open door, since the use of a language does not obviously require the adoption of an ontology. And hence there does not arise any question about the absolute or relative status of that ontology. He would simultaneously leave himself open to the objection that denying the reality of reference would falsify many theories which we might want to accept. So instead he tries to show that theories' truth-values are independent of their ontological commitment. – Of course, this and the preceding two paragraphs are by way of suggestion; Quine does not supply this argument. But I submit that he makes best sense when read this way.

With these background remarks in place, let me present and examine the argument, as I extract it. The argument for ontological relativity proceeds in two and a half stages. The first stage establishes that reference is underdetermined by all the candidates that might determine it. The second stage moves from underdetermination to indeterminacy: the result is that reference is indeterminate, and that there are no absolute facts about what our terms refer to. This is what Quine calls the inscrutability of reference. The half stage is the move from inscrutability of reference to ontological relativity. Quine presents it as if ontological relativity is just the position which someone who accepts the inscrutability of reference will take on matters of ontology. He speaks as if they are two aspects of the same position. So for the purposes of exegesis, I must follow him, and present a two-stage argument. But for the purposes of evaluation, I must diverge and treat the latter stage as two, since a thorough evaluation requires examining exactly how these remarks on reference affect metaphysics.

The first stage, that of showing that reference is underdetermined by all candidates, relies on the use of a model-theoretic argument. Given a theory, the constraints on what universe to pick for its variables to range over, and on the models satisfying the theory, stop well short of providing us with just one model in very many cases. This means that there will be multiple models of our theories. He argues that it is impossible, from within the vocabulary of a theory, to distinguish between different models. Picking

out the reference of one term using some other terms will rely – as already cited (1969: 53) – on the uncritical acceptance of the latter terms. If we want to establish what objects *all* terms in a given vocabulary refer to, then we will not uncritically accept any terms. And if we want to carry this task out using *only* terms in that same vocabulary, then we are bound to fail. For there is no way of getting started.

This seems quite clear. But it does not address the question of whether reference is underdetermined by *all* available candidates. It only tells us that the reference of a vocabulary cannot be fixed using only that vocabulary, and this is perhaps not very remarkable. Quine is sensitive to the fact that there may be other ways of fixing reference. That is, he thinks there might be one other way: pointing. However, it is clear that similar considerations will apply to ostension. There are two issues. First, there is the problem of individuation. Pointing will not distinguish between this rabbit, and undetached rabbit parts; and the reference of "rabbit" will depend upon fixing the apparatus of individuation – plurals, and the like. Second, even with this apparatus fixed, there is still the possibility that the object of the *direct* ostension – the thing actually pointed at – is itself supposed to be pointing at something else. So I might point to the petrol guage to indicate that there is no petrol (1969: 40). Or I might point at some grass in order to illustrate the reference of the term "green" when construed as an abstract singular term (i.e. referring to greenness). This is not direct ostension because "the abstract object which is the color green [...] does not contain the ostended point" (1969: 40). Quine calls these cases *deferred* ostension, and argues that it is impossible to distinguish between cases of deferred ostension to illustrate abstract singular terms, and direct ostension to illustrate concrete general terms (e.g. (my example, since Quine provides none) the term "green" when construed as referring to green things).

However, the details of Quine's discussion are not important: I do not want to get bogged down in a discussion of universals. The general point about ostension is that much of what goes for referring terms will go for ostension. This is true unless some special basic status can be claimed for ostension. For otherwise it is not even settled which direction a pointed

finger is meant to be pointing. And just as it is impossible to settle which objects the terms in some vocabulary refer to just using that vocabulary, it is impossible to settle which objects a pointing finger points at just by pointing that finger. Although ostension may have a special psychological status, this does not affect Quine's point. For this special psychological status does not affect the fact that, insofar as ostension can be seen as a device for picking out objects, there will be multiple models of ostension, just as there are multiple models of our verbal theories. The pseudo-language of ostension cannot frame a distinction between the pseudo-models of its pseudo-theories, any more than our ordinary language. Thus any psychological status that ostension may possess cannot embody such a distinction. For this reason, I do not think that it is plausible to argue for some stronger special status of ostension. Besides, those who disagree do not disagree for the sake of some special notion of ostension. No-one else thinks that ostension will solve the problem. So I will take it as granted that ostension is not the answer to these problems, and that it cannot fix reference; and I will leave the matter here.

It is in the issue of paraphrasing that the thesis of indeterminacy of translation plays a role. For:

> Ontology is indeed doubly relative. Specifying the universe of a theory makes sense only relative to some background theory, and only relative to some choice of a manual of translation of one theory into the other.
>
> (1969: 55)

But the argument for ontological relativity does not rely on translation being indeterminate. After all, to look ahead a little, Quine argues that at some point – when we consider our background theory – there is no antecedently understood vocabulary, and so paraphrasing will not fix reference. Yet he still thinks that ontology is relative for our background theory. It is his reasons for thinking *this* that I am interested in; and in this case, no translation is occurring. So the indeterminacy of translation cannot be doing vital work. That is why I feel safe ignoring the indeterminacy of translation when considering the argument for ontological relativity. We can regard the passage just cited as something of a flourish.

So Quine suggests that the only way in which we can, and do, fix the reference of terms, is by "paraphrase in some antecedently understood vocabulary" (1969: 54). It is here that doubts arise. For although he has shown that two candidates for fixing reference – definition using the vocabulary of the theory in question, and ostension – are inadequate, he has not ruled out the possibility that there are other candidates. At least, if he has ruled it out then he has done so very subtly. This is where attacks are directed, as we will see in the next chapter. For now let us flag it as a hole in the argument.

The second part of the argument towards ontological relativity is the move from the claim about reference being underdetermined by all the candidates, to the claim that it is indeterminate. The model-theoretic argument discussed, ignoring the flagged hole for the sake of argument, shows that the only way to pick a universe for a theory's model is to use the vocabulary of a further theory. No theory can distinguish between its own models. But it is important to realise that Quine is not merely claiming that, using only the vocabulary of a theory, we can't *tell* the difference between models. He is saying that, from within the vocabulary of the theory, there *is* no difference: the vocabulary cannot frame any difference.

The argument is this. We are granting the first stage, that nothing is available to determine what the terms in the vocabulary of a theory refer to, apart from paraphrase in other vocabularies. But if we set paraphrase aside for a moment, then "to question the reference of all the terms [...] becomes meaningless, simply for want of further terms relative to which to ask or answer the question." Confining ourselves to the vocabulary of a theory, we will be unable to distinguish between different models. We will be unable to frame any such distinction, in the terms available. So, from the confines of the vocabulary in question, it becomes impossible to maintain that there is any such distinction. For it is impossible to make the distinction in the first place. If it is impossible to maintain that there is a distinction between models, or to say that there is, then it follows that there is no fact *which can be stated in the terms of the theory* about that distinction.

Of course, if we have recourse to a background theory,[5] then we will be able to frame this distinction. And in the background theory, there may be a fact about which model of the object theory is the intended one. But the closing move of this second stage of the argument is to point out that there will come a point at which no background theory will be available. There are two other possibilities, admittedly; but neither is attractive. Regress is the first. There might be an infinity of background theories, each more general and inclusive than the last. However, since it is plausible that we are finite beings, this appears to be false of our actual situation; and if it were true it would hardly be a comfort. Circularity is the second. Various theories might be available for paraphrasing, with none as the most basic. Quine does not consider this option. However, I do not see that it will help. For if the circle of interdefined theories were complete, then it should be possible to show a vicious circularity, by fixing the reference of the terms of each theory using the vocabulary of another until we reached the theory we started with. This would amount to settling the reference of a theory using its own vocabulary, and this we have already seen gets us nowhere. And if it were not possible to go full circle, as it were, and reach the theory we started with, then the situation would not be a case of circularity but of regress, which we have also seen gets us nowhere.

So when we consider our background theory, as when we consider any other theory, there is no fact of the matter that we can state concerning what our terms really refer to. This means that there is no fact of the matter *at all*. We can see, by analogy with our other non-background theories, that our background vocabulary is unable to frame the distinction between the models of the background theory. But for this reason, we can also see that any attempt to go ahead and talk about the reference of the terms of the background theory directly is bound to fail. It is thus senseless to hold that there is a fact of the matter about what our terms really refer to. For

[5] Quine is not totally explicit as to what makes a background theory suitable for fixing the reference of the object theory. He generally writes as if the universe of the subordinate theory must be "some portion of the background universe" (1969: 50-1). But sometimes he leaves this qualification out: "...it makes no sense to say what the objects of a theory are, beyond saying how to interpret or reinterpret that theory into another" (1969: 50). I do not see that anything turns on this ambiguity.

whatever it was that we held (e.g. that there is such a fact of the matter), it would not have the sense we wanted it to have. For it would, necessarily, be framed in the vocabulary of our background theory.

In this context Quine's unacknowledged move from reference to ontology is reasonable. (This is the two-and-a-halfth stage of the argument.) Whatever we say about the ontology of a theory will only make sense relative to a background theory: for "it makes no sense to say what the objects of a theory are, beyond saying how to interpret or reinterpret that theory in another" (1969: 50). And questions of ontology are questions about what the objects of a theory are. So questions of ontology will only make sense relative to a background theory. There are no questions of ontology which make sense in any other way. And this is ontological relativity.

So I am inclined to agree with Quine, that the results about ontology go through with the first two stages of his argument. In addition, I have flagged what I perceive as the most significant gap in his argument – namely, the failure to rule out the possibility that some means of fixing the reference of terms has been overlooked. This gap is the subject of Chapter 3. But first I want to consider some parts of Putnam's work, and then to relate the two.

2. Internal realism

My aim in this work is to see whether a certain family of considerations, especially model-theoretic considerations, undermine a substantial notion of reference. Putnam's aim, in the work we will look at, is to refute metaphysical realism. Yet his arguments are relevant to my project, since they turn on the sorts of considerations about reference that concern me.

Putnam presents something of an array of arguments, along with considerations, thoughts, historical allusions, and plenty of italics. He is not consistent between versions of his arguments (and sometimes not even within the same version). By *not consistent*, I do not simply mean that his arguments contradict one another (or contain contradictions); that is an unremarkable occupational hazard. I mean that he is not consistent about *which* argument is supposed to be the crucial one for dispensing with metaphysical realism. In particular, two non-equivalent arguments stand out. One is the version which attacks the (alleged) consequence of metaphysical realism, that an epistemically ideal theory might be false. This version is prominent in *Realism and Reason* (in Putnam, 1978: 123-140) and *Models and Reality* (in Putnam, 1983: 1-25). The other version suggests that we can permute the referents of every term in our language such that no sense can be made of the (again, alleged) metaphysical realist claim that we can determinately refer. This version is described in detail in *Reason, Truth and History* (Putnam, 1981). Somehow, the epistemically ideal theory also features; but Putnam does not make the link very clear. This I hope to rectify. I propose to focus on these two arguments, and to relate them through the following suggested reading of Putnam.

Metaphysical realism, as Putnam sees it, is the following picture:

> *Metaphysical* realism [...] is less an empirical theory than a model – in the "colliding billiard balls" sense of "model". It is, or purports to be, a model of the relation of *any* correct theory to all or part of THE WORLD [...] In its primitive form, there is a relation between each term in the language and a piece of the world (or *kind* of piece, if the term is a general term).
>
> (Putnam, 1978: 123-4)

[...] THE WORLD is supposed to be *independent* of any particular representation we have of it [...]

(Putnam, 1978: 125)

The relation, between terms in a theory and pieces of THE WORLD, is the relation of reference; and THE WORLD is supposed to be independent of what we think or say about it. Putnam makes some further claims about the place of the reference relation in the metaphysical realist's picture:[6]

Minimally, however, there has to *be* a determinate relation of *reference* between terms in L [a language] and pieces (or sets of pieces) of THE WORLD, on the metaphysical realist model[7] [...]

(Putnam, 1978: 125)

My suggestion is that we see this claim as bipartite. Putnam's target metaphysical realist holds two essential views about reference. The first is that there must be such a relation: our language and the world must be related by a reference relation; that is, our terms must refer.[8] The second is that this relation must be determinate. An obvious way to read this would be as saying that, in each case, the reference relation succeeds in relating the term in question to only one piece of the world. That is, the relation must be one-to-one. (Let us waive mereological issues for the sake of argument.)

So I suggest the following formulation of Putnam's target view.

(a) The world is independent of any representation we may have of it.

(b) A special relation, called *reference*, sometimes holds between language and the world, by virtue of which linguistic constructions succeed in representing the world.

(c) When a theory is true, there is at least one reference relation between language and the world. (This is the *existence clause*.)

---

[6] In the text, this passage comes between the previous two.

[7] It should be obvious that the sense of "model" that Putnam is employing here is not the technical one I use. I will only use the word in it's technical sense; for the sense used by Putnam in these quotations I shall use "view", "position", "picture", or some such.

[8] Remember that this is a picture applying to *true* theories; obviously "unicorn" might not refer, but it is not part of the vocabulary of a true theory.

(d) When a theory is true, there is at most one reference relation between language and the world. (This is the *uniqueness clause*.)

Obviously this leaves a lot unsaid. In particular, it should by no means be read as an elucidation of the notions of truth or reference. They are assumed. Putnam sometimes says that he is attacking a correspondence theory of truth; however, I will not be primarily concerned with those claims.

One terminological note: I will use "metaphysical realism" to refer to the view demarcated by these four claims. I realise that many self-professed metaphysical realists might characterise their position with more, fewer or different claims. But I will simply stipulate that, as I use it, the term will be defined by these four claims. I stipulate this because I think the terminology is not wholly unreasonable. If the reader disagrees, I recommend mentally substituting some novel or more appropriate term.

The two lines of argument I have indicated may now be seen as focussing on (c) and (d) respectively. The argument concerning the existence of a model for the epistemically ideal theory shows that (c) is too easily achieved, even allowing that reference relations must meet idealised empirical constraints. The near-triviality of (c) Putnam uses in an attempt to undermine a substantive realist notion of truth. And the argument concerning permutations of our language shows that claim (d) is false. (I emphasise my awareness that this is not how Putnam presents his arguments; I am undertaking an organisational manoeuvre.) Before examining the arguments in more detail, I will briefly outline them.

We start with the argument concerning the existence of a model for the epistemically ideal theory, which I suggest renders (c) nearly trivial and which consequently undermines a realist notion of truth. The structure of the argument is this. First, Putnam claims that it follows from metaphysical realism that our best possible theory, the product of an ideal investigative process, might be false. Such a theory is the *epistemically ideal theory*, and the claim that it might be false is the claim that "truth is *radically non-epistemic*" (1978: 125). He implicitly assumes that there is only one ideal theory. This is the theory that satisfies operational constraints (those

imposed by experience) and theoretical constraints (formal constraints on the theory, e.g. simplicity). It satisfies them exceptionally well – as well as we could ask, or better. (Obviously this is an imaginary theory.) The argument is then simply that, by the model-existence theorem, the theory will have a model (assuming, as seems plausible, that the ideal theory is consistent). This means that, with the addition of a further clause along the lines that the ideal theory might be false, the negation of (c) follows from that extra clause plus (a) and (b). I.e., there will be no reference relation such that the world is independent of our representations and the epistemically ideal theory might be false. In fact, Putnam does not see any need for an extra clause; I discuss this shortly. This argument relies on a standard Tarskian analysis of truth (which I presented in Chapter 1), which makes truth relative to a structure. In essence, Putnam exploits this feature of the analysis to question the sense the metaphysical realist's further claims about the possible falsity of an ideal theory.

The other argument – the permutation argument – shows that the negation of (d), the uniqueness clause, follows from (a), (b) and (c), although in the end Putnam wants to preserve (d) and reject one of the other clauses. The point Putnam proves is that for a language like ours, if there is an interpretation, then there will be another, yielding indiscernible models for every possible sentence; and it will not be possible to "intend" one model rather than another from the vocabulary of the language. (See Putnam, 1978: 126.) These considerations have a broad application, for they can be used to show that we are not able to distinguish between interpretations of our whole language. (See Putnam, 1981: Ch 2.)

Let me lay each argument out in more detail, starting with the argument for the existence of a model for the epistemically ideal theory.

Putnam's first premise is that it follows from metaphysical realism that the epistemically ideal theory might be false. There is a common objection to this premise, which is that this does not follow.[9] After all, suppose for a moment that MR is metaphysical realism, and IT stands for

[9] For a concise discussion of such sentiments, see Zalabardo 1998: 222-223.

"the ideal theory is true". Then at first glance the following inference does not appear to be valid:[10]

$\neg(MR \rightarrow \Box IT)$

$(MR \rightarrow \blacklozenge\neg IT)$

However, I maintain that it *does* follow, from Putnam's formulation and mine, that metaphysical realism implies the possible falsity of the epistemically ideal theory. In the first place, it follows from the plausible assumption that metaphysical realism implies that the ideal theory is not necessarily true. In other words:

$(MR \rightarrow \neg\Box IT)$

$(MR \rightarrow \blacklozenge\neg IT)$

The premise of this inference is plausible because we are considering empirical theories, whose truth are among the most plausible candidates for contingent truths.

But now notice that the premise of the second, valid, argument follows directly from the premise of the first argument. It is a straightforward case of $\neg(A \rightarrow B)$ entailing $(A \rightarrow \neg B)$. Thus by inserting this intermediate step, the first inference is valid after all:[11]

$\neg(MR \rightarrow \Box IT)$

$(MR \rightarrow \neg\Box IT)$

$(MR \rightarrow \blacklozenge\neg IT)$

Perhaps this inference might be more open to question if the relation involved was strict entailment rather than material implication, although in

[10] "¬" is negation; "→" is the material conditional; "□" is the necessity symbol; "◆" is the possibility symbol.
[11] I am indebted to José Zalabardo for alerting me to this point.

my opinion it would still be valid. But my point here is that those who scorn the claim that metaphysical realism implies the possible falsity of the epistemically ideal theory, seem oblivious to the fact that a *prima facie* plausible argument for that claim is quite easy to construct.

Nevertheless I think it is terrible strategy to use this claim as a premise in the larger argument. First, an argument concerning the soundness of some modal logic (Putnam's or mine) is not to the present point; so I will not do any more to defend my claim. Second, the claim that however good our theories get, we might nonetheless be wrong, is an unfair characterisation of metaphysical realism. In fact, it is more akin to a characterisation of scepticism. Realists are often epistemological optimists. Hence the metaphysical realist is almost certain to object that she has been inadequately characterised. The debate on how best to characterise metaphysical realism will be even less fruitful than the debate about the logic of modal inference. Hence it should be avoided. And finally third, a modification of the claim exists which – though it doesn't follow from (a)-(d) – I believe that the metaphysical realist will accept, yet which supports Putnam's conclusion. Let me now exhibit this modification:

Suppose, *contra* Putnam, that the metaphysical realist believes that the epistemically ideal theory must be true, for some reason or other. For example, she may think that, as it happens, the facts singling out the epistemically ideal theory are also such as to guarantee its truth. Whatever the reason is, I suggest that she will not want it to be Putnam's reason. That is, she will not want it to follow from the claims (a)-(c) that the epistemically ideal theory will be true. If the truth of the ideal theory turns out to be a mere logical consequence of (a)-(c) (plus the notion of *epistemically ideal theory*), then the metaphysical realist will not be happy. For she will, I think, not be inclined to view that theory's truth as a logical consequence of the minimal position we have attributed to her.

If I am right, then we need to supplement claims (a)-(d) above with a further claim on behalf of the metaphysical realist:

(e) If the epistemically ideal theory must be true, then this is not a logical consequence of (a)-(c). (Note that (d) is not required for this argument.)

Thus, if she accepts that the ideal theory might be false, the argument can proceed, and if she denies it, then all she needs to accept for the sake of the argument is that the truth of the ideal theory does not follow from (a)-(c).

The second premise of Putnam's argument is the bipartite claim that the epistemically ideal theory must satisfy idealised operational and theoretical constraints, and that there is nothing else for the theory to satisfy. A discussion of the latter part – the assertion that there is nothing besides operational and theoretical constraints available for the ideal theory to satisfy – I will set aside. For here lies the externalist challenge, to which I will devote the remaining two chapters of this work. As for the claim that the ideal theory *must* satisfy operational and theoretical constraints, the metaphysical realist will presumably want more constraints, not fewer; and hence there are only two reasons she would reject the ones Putnam offers. The first is that she might want to propose some better constraint, and this line, as mentioned, I defer to a later discussion. The second is that she might suspect him of surreptitiously smuggling something in with his constraints, which allows his argument to go through. We should therefore briefly examine them to check that this is not the case.

There are no obviously relevant problems with the notion of a theoretical constraint. Those who attack the notion of "superempirical virtues" or "theoretical virtues" such as simplicity, elegance, explanatory power, scope, etc., typically do so by way of attacking the epistemological optimism which, I suggested, frequently goes hand-in-hand with metaphysical realism. So as an anthropological matter, I doubt that many realists will object to Putnam's appeal here. And more importantly, if they were to object, then Putnam could simply drop the constraints without harming his argument.

The notion of operational constraints do the real work in Putnam's notion of an epistemically ideal theory. This comes out when Putnam

contrasts the intelligibility of our imperfect current theories being false, with the alleged senselessness of the ideal theory being false. If T is a formalisation of present-day total science, then:

"[...] T is, we may suppose, well confirmed at the present time, and hence rationally acceptable on the evidence we *now* have; but there is a clear sense in which it may be false. Indeed, it may well lead to false predictions, and thus conflict with OP [operational constraints]."

(Putnam, 1983: 12-3).

What he leaves unsaid is that it is much less likely that our current theories will suddenly be deemed insufficiently elegant or some such; if we think they are our best current theories, then we presumably have already cast them as simply and inclusively as we can. Only further empirical research will provide impetus to change our theories. And it is this possibility which gives our current theories "a clear sense in which [... they] may be false" (1983: 13). In this mood, Putnam starts toying with verificationism, in some modified sense; so perhaps his opponents should pay attention to his formulation of operational constraints.

The most precise and complete formulation occurs in *Models and Reality*. He formulates operational constraints in terms of three things. The first is "a sufficiently large 'observational vocabulary' [...] - call it the set of O-terms" (1983: 11). The second is "a set of S which can be taken to be the set of macroscopically observable things and events" (1983: 12). The third is "a valuation (call it [...] 'OP') which assigns the correct truth value to each n-place O-term (for n=1, 2, 3, ...) on each n-tuple of elements of S on which it is defined" (1983: 12). And "it is the valuation of OP that captures our 'operational constraints'" (1983: 12).

The argument is then, in this version, pushed a stage further by the assumption that there are something like sense-data. So, by "taking the operational constraint this time to be that we wish the ideal theory to correctly predict all sense data" (1983: 15), he hopes to repeat his argument and show that "even terms referring to ordinary material objects – terms such as 'cat' and 'dog'" (1983: 16) turn out to be "formal constructs variously interpreted in various models" (1983: 16).

The formulation in terms of sense-data presents obvious problems. For it is an open philosophical question whether there are any such things, and it is not immediately obvious that a metaphysical realist is any more likely to take one side than the other in this debate. Once again, I accuse Putnam of a strategic error: he commits his opponents to more than his purposes require. So I will focus on the first formulation.

However, even this is hardly unproblematic. The notion of a "macroscopically observable thing" is notoriously hard to pinpoint. According to Putnam, it means "observable with the human sensorium" (1983: 12). But what, we might ask, is so special about the human sensorium? Do we count what we observe through microscopes and telescopes? And if not, do we ban scientists from wearing spectacles in the lab? Moreover, as a species we change over time (according to one of our best current theories), and our senses presumably evolve too. Observability, it has been argued, is both contingent (not based on a principled distinction) and a matter of degree. Once again, I do not mean to enter the debate about observability; I simply indicate that there *is* a debate there. And it is open to Putnam's opponents to take either side in that debate. Someone who thought observability was a matter of degree might not accept that we have an "observational vocabulary", or that there is "a set of *S* which can be taken to be the set of macroscopically observable things and events" (1983: 12).

Another problem arises when we ask how "a valuation (call it [...] 'OP') which assigns the correct truth value to each *n*-place O-term (for n=1, 2, 3, ...) on each *n*-tuple of elements of *S* on which it is defined" (1983: 12) is supposed to happen. What determines the correct truth-value[12] for an observational term? Stipulation obviously plays some role: for we can choose what symbols (spoken or written) to apply to which element of S. But once we have thus stipulated, we are not free to assign truth-values as we like: for that would cut our language entirely free of experience. That is why Putnam speaks of assigning the *correct* truth-value to O-terms on *n*-

[12] When Putnam speaks of the truth value of an O-term on an *n*-tuple, he is not (nonsensically) assigning truth-values to terms. Rather, he is describing a function for each *n*-place O-term from each *n*-tuple of elements of *S* to a truth-value. This is a variant but equivalent procedure to simply assigning elements of *S* to each O-term.

tuples (i.e. the correct referent to each O-term). But what, then, determines the correct truth-value? It looks like some work is being done by a further notion of constraint; and if this is so, then we could be forgiven for suspecting that this further constraint is the important part of the notion of an operational constraint. And Putnam says nothing about this important further part. If, on the other hand, Putnam protests that there is no further constraint at work when he talks about assigning correct truth-values to O-terms, then he must explain what makes such assignments correct or incorrect. This, again, he does not do.

No help is available in any of his other formulations of operational constraint. The intuitive idea is that experience is supposed to constrain our theory. But the problem is making clear the link between experience and the vocabulary of the theory. This is a substantial philosophical problem in its own right. In his defence, this is not just a problem for Putnam; it is a problem for everyone. Nonetheless I suspect an ambiguity in his attitude to experience. Compare:

> So "I seem to myself to push the button", when understood in the "bracketed sense" (as meaning that I have a certain subjective experience of voluntarily pushing a button) has not just the same truth conditions but the same *interpretation* under *J* [a deviant interpretation] and under the normal interpretation *I*.
>
> (Putnam, 1981: 40)

...with the following:

> Even our description of our own sensations, so dear as a starting point for knowledge to generations of epistemologists, is heavily affected (as are sensations themselves, for that matter) by a host of conceptual choices.
>
> (Putnam, 1981: 54)

It is beyond the scope of this work to decide whether there is a genuine indiscrepancy here. I suspect that there might be. (His mention of "sensations themselves" must surely be regarded as a slip.)

But to conclude what has become something of a digression, consider that none of these remarks are of direct use to the realist. The purpose of the discussion was to see whether anything objectionable had

been smuggled in along with the notions of operational and theoretical constraints. And nothing obviously suspicious has been found. For it is no surprise that operational constraints are hard to formulate, and that the link between language and experience is hard to pinpoint. All that Putnam needs his target to accept is that there *are* constraints imposed by experience on what an ideal theory could say. Unless she comes up with some account of these constraints which spoils Putnam's argument, then his failure to formulate them well can be regarded as unfortunate but not fatal. The only obvious account of operational constraints which might spoil Putnam's argument would be one which claimed some kind of extra-linguistic, external link between O-terms and their referents. And this sort of challenge comes under the umbrella of externalism, which I have already said we will discuss in the remaining two chapters, but ignore for the remainder of this.

The argument, in its clearest form, now proceeds as follows. The epistemically ideal theory, which I follow *Models and Reality* in calling T_I, will be consistent. So, by the model-existence theorem, T_I will have models. As in the formulation given previously, S is the set of all observable things and events, and each of its members is denoted by a term we call an O-term. Consider a model of T_I such that the model "is standard with respect to P [...] restricted to S [...] for each O-term P." Putnam says:

> Now, such a model satisfies all operational constraints, since it agrees with OP. And it satisfies those theoretical constraints we would impose in the ideal limit of inquiry. So, once again, it looks as if any such model is 'intended' – for what else could single out a model as 'intended' than this? But if this is what it *is* to be an 'intended model', T_I must be *true*: true in all intended models! The metaphysical realist's claim that even the ideal theory T_I might be false 'in reality' seems to collapse into unintelligibility.
>
> (Putnam, 1983: 13)

Notice a clear statement of the point externalists challenge: the rhetorical question, "what else could single out a model as 'intended' than this?" The important point, however, is that on a model meeting the conditions he describes, T_I will be true. So we have it as a mere logical consequence of the metaphysical realist's view that the epistemically ideal theory will be true. I

agree with Putnam that, with the provisos mentioned, this follows; and moreover, by (e), my modified version of his first premise, I agree with Putnam that the metaphysical realist (defined as someone who would accept (a)-(e)) cannot accept this conclusion. A consequence of (a)-(c) is in conflict with (e): therefore metaphysical realism, defined as (a)-(e), is incoherent.

This argument has the form of a *reductio ad absurdum*. Claims (a)-(c) state that there must be a reference relation between an independent world and what we say about it, when what we say is true. Claim (e) states that the truth of the epistemically ideal theory must not follow from (a)-(c). Putnam shows that it does.

Since the argument is a *reductio*, at least one of its premises must be rejected. I claimed earlier that the argument put pressure on (c), but obviously a *reductio* puts equal pressure on all its premises. If the argument is construed to put pressure on (c), the existence clause, the pressure is that no such relation exists such that (a), (b) and (e) are satisfied. Conversely, insofar as (c) is plausible or desirable, we should reject assumptions that place pressure on it. There is a subsidiary difficulty that, as formulated, (c) depends on (a) and (b) (for it mentions truth, the world, and so on). Putnam's solution is to reject (a), (b) and (e), and to hold on to something like (c) – but we will examine his position shortly.

Let us first turn to the other strand of Putnam's argument, the permutation strand. This may be seen as inessentially supplementing what we have just discussed by showing that there is no way of ruling out the model of the epistemically ideal theory on which it comes out true. It also constitutes a standalone attack on the conjunction of (a), (b) and (d) – the combined claim that the world is independent of our language and that the reference relation is unique.

The attack goes like this. First, operational and theoretical constraints are introduced as the way we fix interpretations of our language.

The most common view of how interpretations of our language are fixed by us, collectively if not individually, is associated with the notions of an operational constraint and a theoretical constraint.

(Putnam, 1981: 29)

He does not argue for the claim that this is the most common view. Again, this is a point where the externalist will protest, as we will see later; we will let it pass for now. We have already examined operational and theoretical constraints. I will only remark that they are constraints not on language itself but on its *use*; the underlying thought is that it is only by using language to make claims that we can hope to assess which terms refer to what. At this point, truth-conditions come onto the scene, and some surface ambiguity sets in. For on the one hand:

> Since the constraints we use to test the theory *also* fix the extensions of its terms, the thinkers' estimate of the theory "working" is at the same time an estimate of its truth.

(Putnam, 1981: 32)

Here, the picture he apparently wants to attack has operational and theoretical constraints fixing reference of the terms of a theory directly, and thereby determining the truth-value of the theory. But on the next page, it is the other way round:

> The received view [...] tries to fix the intensions and extensions of individual terms by fixing the truth conditions for whole sentences.

(Putnam, 1981: 33)

In fact, this is the picture he goes on to attack. The ambiguity may be resolved as follows. Putnam takes it for granted that "there is nothing in the notion of an operational or theoretical constraint to do this [fix reference] directly" (1981: 33). What operational and theoretical constraints *can* do, he thinks, is fix truth-conditions of whole sentences. And he then goes on to argue that truth-conditions can't fix reference either. So the view he wants to attack does have operational and theoretical constraints fixing reference, but only indirectly through fixing truth conditions for sentences.

Before we consider his argument, it is worth examining the picture he dismisses out of hand. That is the picture on which reference is fixed directly by operational and theoretical constraints, and sentences' truth conditions are determined either in parallel or subsequently. If this were so,

it would fatally undermine his subsequent argument. Moreover, the idea has a certain initial plausibility: we normally think of our sentences being true or false by virtue of the terms used, and what they refer to; we do not think of our terms referring to whatever they need to refer to in order to preserve a sentence's agreed or desired truth condition.

We are seeking a justification for the claim that "there is nothing in the notion of an operational or theoretical constraint to do this [i.e. determine what our terms refer to] directly" (1981: 33). We might consider two possibilities. The first might be that the sentence is the basic semantic unit. However, this would be a bad premise, because attacking it would provide a way out for the metaphysical realist which completely avoided the issues at which Putnam's argument is aimed. Second we might consider the idea that these constraints are only effective when a term is used. Plausible as the suggestion might be, it appears too general to be of relevance in a linguistic context. If I shout "Putnam!" to scare an over-interested cow, then I am indeed using the term, and I am operationally constrained (by the approach or departure of the cow). Yet my success when the cow departs is not relevant to the reference of "Putnam". In short, the unqualified notion of "use" is rather vague. And I do not propose to attempt to tighten it up. Besides, this suggestion also runs the previous risk of providing an irrelevant focus for debate.

To achieve the justification we must therefore take the matter head on, and suggest reasons why operational and theoretical constraints can't fix the reference of terms directly. And it isn't hard to see that these constraints are ill-suited to fixing the reference of terms directly. Theoretical constraints constrain theory form – they concern formal features of theories, such as complexity. Terms do not exhibit any such features. It might be suggested that the extension or intension of a term should be subjected to theoretical constraints. But this is a mistake. The error is to construe the assignment of objects to a term as itself a theory. Of course, we might have such a theory – perhaps concerning the terms of an unfamiliar language we are learning. However, if such a theory is seen as *constituting* the assignment in question, then we have arrived at a severe regress by a shorter route than Putnam has

in mind. For we are admitting the unabating need for a further theory in order to assign objects to terms of a theory. (This is, of course, extremely reminiscent of Quine.) If, on the other hand, theoretical constraints such as simplicity are supposed to work on reference directly *without* necessitating any further theory in which the assignment of objects to terms takes place, then I start to suspect that the notion in play is not one of a theoretical constraint at all. Nor can I see any related notion that will do the job, this side of metaphysics. It might be that there is some direct constraint of, e.g., simplicity, imposed on extensions in general. (David Lewis suggests something of the sort.) But this would have to come under the externalist heading I am postponing till the next chapter. For it would not be something we could appeal to, employ, or operate with, in the way that theoretical constraints are normally appealed to, employed or operated with.

Operational constraints are patently unsuited to directly constraining reference. There is no kind of experience that can directly constrain the reference of any term. For in order that experience should play a role, something must happen involving that term and the speaker. The speaker must have an experience that is somehow relevant to the reference of the term. It is hard to see how this might happen in any way except by using the term to "say something", whatever that turns out to mean. Merely brandishing a term is unlikely to yield any informative experiences regarding that term's reference. If I say "cat" to you, then you might think that I was asserting that there was a cat around, or calling you a cat, or asking you where the cat was; but in each case, "cat" would count as a sentence. After all, in each case you are understanding more to my utterance than cat: proximity, attribution and querying respectively. (If this isn't obvious, notice that the three interpretations are incompatible. This means that they can't all say the same thing: and therefore they can't all be *just* cat.) If, on the other hand, I simply say "cat" and do not mean anything more by it than cat – just try to use the term on its own, as it were – then I haven't said anything that you, or anybody else, might object or assent to. I may have used the term, but not to say anything. (Perhaps I just like the sound of the word.) Nor could my utterance be somehow shown to be

incorrect by any experience I might have. (There don't even need to be cats: I might say "unicorn".) There is nothing for an operational constraint to get a grip on in my use of "cat" alone.

This concludes my supplied argument to support Putnam's claim that "there is nothing in the notion of an operational or theoretical constraint to do this [determine what our terms refer to] directly" (1981: 33). Instead, Putnam suggests that operational and theoretical constraints constrain *truth conditions of sentences* – at least in the "ideal limit of inquiry" (1981: 33). Presumably the idea is that in our present imperfect state of knowledge, there might be a certain degree of underdetermination – that is, operational and theoretical constraints might not at present determine truth conditions of all our sentences. Since that is irrelevant to Putnam's point, he allows that we might have a determinate answer at some stage to every question we could ask (even if we think this an unlikely scenario). His argument is now that even then, when the truth condition of every sentence in our language is fixed by operational and theoretical constraints, the reference relation cannot be one-to-one on the metaphysical realist picture.

His argument is a proof of the following theorem:

> Let L be a language with predicates F_1, F_2, ..., F_k (not necessarily monadic). Let I be an interpretation, in the sense of an assignment of an intension[13] to every predicate of L. Then if I is non-trivial in the sense that at least one predicate has an extension which is neither empty nor universal in at least one possible world, there exists a second interpretation J which disagrees with I, but which makes the same sentences true in every possible world as I does.
>
> (Putnam, 1981: 217)

The proof goes through, as far as I can see, and there is no philosophical interest in examining the logic (which basically turns on the completeness theorem). The idea is that there will exist more than one interpretation of the whole language, such that all and only possible sentences having models (in any given possible world) under one interpretation will have different but

[13] By intension, he means extension at each possible world. If you do not want to admit possible world apparatus, then the argument can be obviously adjusted to run for truth values and extensions, rather than truth conditions and intensions.

indiscernible models under another interpretation. The interpretations will differ in the intensions they assign to predicates. The intuitive idea is easily grasped: it is possible that you and I might interpret our common language such that we both hold exactly the same sentences to be true/false under exactly the same conditions, yet the terms in our sentences might be referring to different things.

The argument obviously turns on the claim that the relation between each model and *L* is a reference relation. This might initially appear question-begging. But notice that if we accept that only operational and theoretical constraints are available to fix truth-conditions to, in turn, fix reference, then we must accept that every model *L* bears the reference relation to all its models satisfying those constraints. For to deny that would be to deny that only operational and theoretical constraints are available (via truth-conditions) to fix reference. Hence the structure of the externalist challenge, which denies the latter claim and therefore rejects the claim that *L* bears the reference relation to all the specified models. But this must wait until the next chapter.

Hence the reference relation is not unique. The permutation argument does not attribute any particular theory of reference to the metaphysical realist. Rather, it shows that whatever properties the reference relation has, it cannot possibly satisfy (a), (b) and (d).

It remains to summarise. *Internal realism* is a term Putnam applies rather generously; however, in the context of the arguments we have examined here, I suggest that the view best attributed to that term is essentially the negation of metaphysical realism as described. Putnam rejects the picture of a world independent of our representations of it. Ignoring some of his wilder claims, the validity of this core move deserves emphasis. Granting certain provisos, his argument goes through; he uses the place of reference in metaphysical realism to demonstrate two problems with the metaphysical realist view. The first is a problem to do with truth, and the second is a problem for reference itself. It is important to note that Putnam does not attack any particular theory of reference; he attacks the features that the reference relation must have, according to (his construal of)

metaphysical realism. He focuses on the work the relation must do, not on how that work is done. This work, he claims, is impossible. So he rejects the view which relies upon that work, impossibly, being done. Whatever else we might think about his arguments, there is an admirable clarity in this response to the perceived problems.

Notice, importantly, that Putnam chooses to reject (a), (b) and (e), for the sake of keeping something like (c) and (d) – jointly, the claim that there exists exactly one reference relation. Obviously as I have formulated them, (c) and (d) are stated in terms not amenable to someone who rejects (a) and (b), but Putnam wants to preserve the spirit of them. It is very important to see that this need not be the expression of a bare preference; it can be supported by an argument. In particular, it would not, logically, be open to reject (c), (d), or both, and hold on to (a), (b) and (e). Notice, first, that (c) and (d) can't be rejected together: it is not consistent to maintain that the number of reference relations is both less than and greater than 1. Since the two arguments I have ascribed to Putnam are somewhat independent, a problem would remain for any view which rejected only (c) and kept (d), or vice versa. Thus it is not permissible to maintain (a), (b) and (e) along with any combination of (c) and (d), and it is not permissible to deny both (c) and (d). Therefore (c) or (d) must be accepted, and the conjunction of (a), (b) and (e) must be rejected. This is an important point, but it is not something that Putnam stresses: he merely says that the denial of determinate reference is absurd. I hope that I have provided an argument for this claim. It renders plausible, though not unavoidable, the suggestion that we accept both (c) and (d) and reject (a), (b) and (e).

It might be objected that nothing forces the metaphysical realist to hang on to (a), (b) and (e) together, and that (c) or (d) might be compatible with some lesser combination of those three. In particular, the existence clause, (c), is surely compatible with (a) and (b) on their own, minus clause (e) concerning the ideal theory. After all, we needed to insert that clause in order to run the first of the two arguments we looked at. This objection is correct. However, the rejection of (e), or of (a) or (b) for that matter, will hardly be seen as a victory for Putnam's opponents. There may indeed be

other logically permissible alternatives to the rejection of all of (a), (c) and (e) and the acceptance of (c) and (d). But this takes us further away from a discussion of anything that looks remotely like metaphysical realism. There may be other ways of salvaging the wreck, but the wreck won't bear much resemblance to its former self. So while Putnam's own solution does not follow from his arguments, it is not clear than any other solution will offer any more comfort than Putnam's chosen one to the metaphysical realist.

Putnam's position he calls *internal realism.* The "internal" comes from his response to the same sort of questions that he presses on the metaphysical realist. Putnam maintains that he is a realist, when realism is construed as an "empirical theory" (1978: 130). He explains what he means by this in his second John Locke Lecture (in Putnam, 1978: 18-33). Thus he believes that the best explanation for the success of our investigations into the world is that they converge on the truth and that their terms typically refer. This sounds eminently realistic. Where he will not be drawn is on questions about reference, and indeed truth, of theories when those questions are asked from outside of any theory. Internally, from within a theory, we can say what, e.g., "cat" refers to: cats. From within another theory, we might question whether there are any such things. But he emphasises that we cannot legitimately ask any questions from outside *all* theories whatever, and say whether there are *really*, e.g., cats, independent of any theory. The reason we can't legitimately ask such questions is that they presuppose a picture – of an independent world, of truth and of reference – which he believes to have shown is incoherent.

3. Relating the arguments of Quine and Putnam

The binding of this work is a bundle of claims about reference, given explicit logical form by model theory. The connecting thread between Quine and Putnam is spun from the same bundle. Once again, I propose to be highly selective in my approach. I am seeking to make explicit a link between certain thoughts of two complex and prolific thinkers. In fact there is a clear historical link, as Putnam frequently mentions. But I am not aiming to establish that, in general, there are similarities between the two; I aim to establish that there is a specific vein of thought common to both. I will have occasion to disagree with Putnam's own analysis of the connection.

First I will argue for a strong similarity between Putnam's permutation argument and Quine's argument for the inscrutability of reference. On the other hand, I will deny that Putnam's other argument, the model-existence argument, has any parallel in Quine. Second, I will try to make explicit the shared assumptions on which the shared arguments run.

Some striking surface similarities should be apparent at once. However they can be misleading. The permutation argument of Putnam might appear very similar to Quine's indeterminacy of translation thesis, as Putnam himself suggests (1981: chapter 2). I begin this section by arguing (*contra* Putnam) that, despite certain links, they are not so similar, and that the permutation argument is actually much more closely related to Quine's argument for the inscrutability of reference. I will move on to suggest that on the other hand, there are certain real differences that deserve emphasis. In particular, there seems to be no parallel in Quine of Putnam's argument concerning the possibility of the falsity of the ideal theory. I will finish by exhibiting the core assumption common to both thinkers, rendering them both vulnerable to the externalist challenge.

The appearance of similarity between Putnam's permutation argument and Quine's indeterminacy of translation thesis comes from the fact that both involve a permutation of the referents of terms. This much similarity may be conceded. Moreover, both arguments seek to establish

that something can be preserved across several permutations. The difference I wish to highlight lies in what is preserved. Indeterminacy of translation preserves "the totality of the speaker's dispositions to verbal behaviour" (Quine, 1969: 27). As I argued in the first part of this chapter, at this level it is possible to see the argument as epistemological: it may be seen as a form of scepticism, a claim about how little we can know about what other people really mean. This scepticism might even play a role in an argument against any more substantive notion of meaning beyond a fairly basic behaviourism. But without development, this sort of argument need not touch metaphysical questions. Accepting a radical indeterminacy of translation casts doubt only on our *collective* ability to represent the world, not directly on the possibility of representing the world in general.

I hereby directly contradict Putnam's own presentation in *Reason, Truth and History*, where he claims that the only difference, between Quine's indeterminacy of translation thesis and Putnam's own argument, is that you might get the impression from Quine that non-equivalent interpretations had to be closely linked (e.g. "rabbit" v. "undetached rabbit part"). There is, indeed, some undeniable similarity. Both arguments claim that even if truth conditions can be fixed (by verbal dispositions in Quine, and by operational and theoretical constraints, in Putnam), this will not suffice to establish reference or meaning in any substantive sense. However, this similarity is misleading. The arguments are by no means equivalent, nor do they have similar conclusions. The best way to demonstrate this is to exhibit a case which tells them apart, which I shall now do.

One way to bring the difference out is by considering a case where translation is not going on. Such a case might be one when, alone and in a reflective mood, we ask whether there are really such things as cats. Indeterminacy of translation has no obvious relevance here: for the speaker is presumably not to be seen as translating her own words, but as using them[14]; and denying that there is anything more to her musings than various

[14] It doesn't much matter exactly how she is using them, or what she is using them for; if we asked her, she would most likely say she was using words to help her work something out. Whether they are essential or not is irrelevant. The point is that in this use, whose

verbal dispositions would simply fail to answer her question, and therefore be only indirectly relevant. (For on a dispositionalist account as much as on any other, questions are to be answered.) But Putnam's permutation argument would be of direct relevance. In the light of that argument, she might reply to herself that the question as to whether there are really any cats only makes sense as a (rather silly) empirical question. There is no deeper transcendental or metaphysical question as to whether there really are any such things, because the attempt to ask such questions amounts to the attempt to distinguish between different interpretations of the language from within the language. And the permutation argument concludes that this is impossible.

Once models of a theory have been identified up to isomorphism – or even if they have not, provided the models are indiscernible – anything the theory can say about those models goes for them all. This is the logical thread linking the permutation argument with the argument for the inscrutability of reference, and consequently with ontological relativity.

Quine's argument for ontological relativity moves from the claim that it is impossible to distinguish the models of a theory using only the vocabulary of that theory, to the claim that this goes equally for our most basic, background theory. The relativity of theoretical ontology is supposed to consist in the impossibility of specifying a theory's ontology – that is, picking a universe – except relative to some other theory. Putnam's permutation argument does not appeal to an analogy with the situation of other theories, which I think makes it harder for him to cogently formulate the conclusion of his argument. But the analogy is not essential to either argument. The arguments thus share an identical load-bearing structure. In both cases, it is claimed that no use of a vocabulary can fix the reference of that vocabulary. Deciding truth values of a set of sentences (Quine) or fixing truth-conditions of all sentences (Putnam) won't help (more on this difference shortly). And nothing else is available to fix reference. If this is not obvious, compare the following two quotations.

---

existence in undeniable even if inessential, it is hard to make any sense of the idea that she is translating her own words.

Suppose next that in the statements which comprise the theory, that is, are true according to the theory, we abstract from the meanings of the nonlogical vocabulary and from the range of the variables. We are left with the logical form of the theory, or, as I shall say, the *theory form.* Now we may interpret this theory form anew by picking a new universe of quantification for its variables of quantification to range over, and assigning objects from this universe to the names, and choosing subsets of this universe as extensions of the one-place predicates, and so on. Each such interpretation of the theory form is called a model of it, if it makes the theory come out true. Which of these models is meant in a given actual theory cannot, of course, be guessed from the theory form.

(Quine, 1969: 53-4)[15]

In the following, which should be familiar from its appearance in the previous section, Putnam is stating part of the theorem he subsequently proves.

Let *I* be an interpretation, in the sense of an assignment of an intension to every predicate of *L*. Then if *I* is non-trivial in the sense that at least one predicate has an extension which is neither empty nor universal in at least one possible world, there exists a second interpretation *J* which disagrees with *I*, but which makes the same sentences true in every possible world as *I* does.

(Putnam, 1981: 217)

Disregarding for the moment the difference in formulation (models of theories vs. interpretations of language), it should be clear that these two snippets are driving at the same point. That point, to reiterate, is that there will be more than one assignment of objects to terms, such that the same sentences come out true.

The shared conclusion is the rejection of the idea that we can make any sense of absolute questions about the reference of our terms, beyond what we can say about them in another theory. The two writers therefore end up saying remarkably similar things, in places. For instance, compare:

[15] Here we see an explicit statement of something Lewis (following Farrell) objects to Putnam doing, namely taking the logical part of the vocabulary for granted. I regard the objection as minor; it is therefore passed over here for fluency, and treated in the next chapter.

[...] "the way the theory is understood" can't be discussed *within* the theory; and [...] the question whether the theory has a *unique* intended interpretation has *no* absolute sense.

(Putnam, 1978: 236)

with the following:

It is thus meaningless within the theory to say which of the various possible models of our theory form is our real or intended model.

(Quine, 1969: 54)

It is evident that the two writers are making the same point: that it is impossible to fix the reference of a vocabulary using only that vocabulary, even preserving the truth values or truth conditions of claims made using that vocabulary.

There remains an apparent difference in formulation, as indicated: Quine uses truth values, theories (sets of decidable sentences), and models; whereas Putnam uses truth conditions (truth at each possible world), language (mostly), and interpretations. However this difference is not deep. I suspect that it merely reflects what Quine and Putnam are willing to commit to, or perhaps what they feel that their audience will willingly accept. (These remarks are purely speculative; there is little in the texts to explain the difference.) For Putnam's argument to work, we need some apparatus of modal logic. Following David Lewis's suggestion that talk of possible worlds should be taken pretty much literally, it has become gradually less controversial to accept – if only for the sake of argument – some such apparatus.[16] But when Quine was writing, framing the discussion in terms of quantification over possible worlds might have gone down less smoothly. At any rate, the crucial point is that the difference doesn't matter. For if somebody rejected the apparatus required by Putnam's argument in terms of truth-conditions, an analogue would remain in terms of truth-values: and I have suggested that this is Quine's argument. To be fair, Quine

[16] Putnam himself explicitly rejects Lewis's modal realism: "To me this smacks more of science fiction than philosophy" (Putnam, 1983: 218). He maintains that talk of possible worlds should be construed as talk of "possible *states* of the world... not possible worlds *à la* Lewis" (Putnam, 1983: 220). But as I go on to claim, it does not seem likely that any

himself is somewhat inscrutable on the question. He does not explicitly rule out that the possible situations in which theories would be true could be considered when fixing reference. I suspect (stressing again that these remarks are merely speculative) that Quine would have said, correctly, that this wouldn't help. For it should be obvious that somebody suggesting such an amendment would, by a reverse move to the previous, be referred back to Putnam's version, which is explicitly framed in terms of truth-conditions. The only way to raise a problem against both these formulations, would be to suggest that truth-conditions needed to be considered, and simultaneously to reject the possible world apparatus. But then the objector would need to explain what a truth-condition was. And as far as I can see, there will be no deeper difficulty in devising a form of Putnam's argument that will work for any proposed apparatus of modal logic.

I have suggested, therefore, that apart from minor differences of formulation, the permutation argument suggested by Putnam is essentially the same argument as Quine's argument for ontological relativity. To recap: the permutation argument centres on the claim that it is impossible from within our language to frame any distinction between interpretations of the language that preserve the same truth-conditions for every sentence. The argument for ontological relativity says that it is only possible to fix the referents of terms of one theory's vocabulary by using the vocabulary of a further theory – and that to fix the referents of a theory's vocabulary is to give that theory's ontology. The common element is the claimed impossibility of singling out a unique assignment of objects to terms using only those terms, even when the truth values or conditions of all sentences made up of those terms are fixed.

By contrast, I suggest that Putnam's argument concerning the epistemically ideal theory has no parallel in Quine. Since there is no parallel, I can't cite any relevant passages from Quine to support my claim. Instead I will digress slightly for a discussion of the role of truth in this aspect of Quine's work. The role of truth in the essay *Ontological Relativity*

---

proposed modal apparatus will present anything more than a technical obstacle for Putnam's argument.

is mostly as that which can be preserved across different assignments of objects to terms. The exception comes in the last paragraph (the one prior to the *Note added in proof*), which I quote in full since it resists easy abstraction:

> Regress in ontology is reminiscent of the now familiar regress in the semantics of truth and kindred notions – satisfaction, naming. We know from Tarski's work how the semantics, in this sense, of a theory regularly demands an in some way more inclusive theory. This similarity should perhaps not surprise us, since both ontology and satisfaction are matters of reference. In their elusiveness, at any rate – in their emptiness now and again except relative to a broader background – both truth and ontology may in a suddenly rather clear and even tolerant sense be said to belong to transcendental metaphysics.
>
> (Quine, 1969: 67-8)

This passage shows that Quine is aware that there are strong links between issues of truth and reference. But it can hardly be construed as an argument concerning the nature of truth. It is a closing remark, and therefore licensed to a vague suggestiveness which would not be permissible in the body of the work. The suggestion seems to be that something along the lines of what he has argued will go for truth as well, because in both cases a further theory is required in order to ask and answer questions concerning the object theory.

Presumably he is thinking of examples like

> "Snow is white" is true if and only if snow is white

whereby the truth predicate can be defined, but only in a metalanguage and only for the object language (so the predicate is properly "true$_L$" or some such). Quine's much earlier work in *Truth by Convention* undermines, among other things, the viability of any attempt to give some special status to the notion of truth beyond that of a mark next to some of the sentences on an infinite list of them all. Consider:

> Each such convention [as the ones Quine has been considering] assigns truth to an infinite sheaf of the entries in our fictive list, and in this function the conventions cannot conflict; by overlapping in their effects they reinforce one another, by not overlapping they remain indifferent to one another. [...] any inconsistency among the general conventions will be of the sort previously considered, viz. the arbitrary

adoption of both "---" and "~---" as true; and the adoption of these was seen merely to impose some meaning other than denial upon the sign "~".

(Quine, 1966: 90)

And moreover, there is an interesting statement of something very like the claim concerning constraints on reference to which externalists object – although here, the context is not reference but "meaning":

[...] in point of *meaning*, however, as distinct from connotation, a word may be said to be determined to whatever extent the truth or falsehood of its contexts is determined.

(Quine, 1966: 82)

The idea expressed here is that, ignoring psychological or poetic associations, the meaning of a word is identical with[17] (i.e. "determined" exactly as far as its) truth-conditions (i.e. "the truth or falsehood of its contexts"). I argued in the first section of this chapter that this very assumption plays a crucial role in the argument for ontological relativity. Moreover it is this assumption which we will see challenged. Its presence in Quine's work on both truth and reference suggests a link between the issues themselves, not just Quine's treatment of them: for it suggests that the same assumption leads you to interesting conclusions in both fields, which in turn suggests that the fields are related.

I do not propose to do anything more explicit by way of linking the two concepts here, either in Quine's thought or in general, for that would take me beyond the scope of this work. Suffice it to conclude that although there might be interesting links between Quine's work on reference and his work on truth, we have seen nothing to suggest any parallel of Putnam's argument concerning the trivial truth of the epistemically ideal theory.

Nevertheless, even considering this argument of Putnam's and finding no parallel in Quine, a shared core is now becoming apparent. I turn to the task of exhibiting this shared core.

---

[17] Perhaps "determined to exactly the same extent as" and "identical with" are distinct. But in this case, I don't see what more there might be to meaning beyond its being determined. However, if the reader will quibble, I will concede the point: Quine may then be read as

There are various components to this core, varying in their levels of generality, and the links between them are complex. One component we have just glimpsed: a commitment to the notion that truth-conditions exhaust meaning. I suggest that the motivation for such a belief is a philosophical naturalism, or even empiricism – a conviction that the world of the senses is the only world, and an abhorrence of postulating further special kinds of entities, such as "meanings". These two strands are somewhat distinguishable. The empiricist strand wants to avoid suggesting there is more to meaning than what is empirically accessible. This is Quine's reason for taking whole sentences as the basic semantic unit, and the relevant property of whole sentences is their truth-conditions, which are plausibly empirically accessible through patterns of assent and dissent, and indeed through direct interaction with the empirical world. At any rate they are more plausibly empirically accessible than a direct relation between word and object. The naturalistic strand resists the postulation of further entities, particularly "mysterious" or "non-natural" ones, where – to tie the threads neatly together – "non-natural" is frequently cashed out in terms of empirical accessibility.

This should be pretty clear, as a comment on the general outlook of these thinkers. In fact a lot of contemporary philosophers have largely naturalistic sympathies. My task is to show how this not unusual outlook takes some responsibility for these rather less popular arguments proposed by our two thinkers.

Quine begins *Ontological Relativity* by explicitly endorsing, as well as summarising, the vague position we are indicating by "naturalism":

> With Dewey I hold that knowledge, mind, and meaning are part of the same world that they have to do with, and that they are to be studied in the same empirical spirit that animates natural science. There is no place for a prior philosophy.
>
> (Quine, 1969: 26)

It would be surprising if the subsequent discussion of reference were completely unrelated to this view. In related tone, but more relevantly to the

---

saying just that meaning and truth-conditions are determined to exactly the same extent. I

present topic, Putnam dismisses both the notion that our minds can somehow reach out and fix reference, and the notion that the world might do so for us.

> Pure mental states of intending – e.g. intending that the term "water" refer to water *in one's notional world* – don't fix real world reference at all. Impure mental states of intending – e.g. intending that the term "water" refer to actual water – *presuppose* the ability to refer to water.
>
> (Putnam, 1981: 43)

And again:

> Given that there are many "correspondences" between our words and things, even many that satisfy our constraints, what *singles out* one particular correspondence *R*? [...] It seems as if the fact that *R* is the reference relation must be a kind of *metaphysically unexplainable* fact, a kind of primitive, surd, metaphysical necessity.
>
> (Putnam, 1981: 46)

In the light of these passages and others like them, it is not particularly contentious to claim that Quine and Putnam share a broadly naturalistic outlook. But exactly what features of this outlook are important for present purposes? What are the key assumptions that underlie their respective arguments about reference? I want to argue that, for Quine and Putnam, the *use* we make of language is the fundamental determinant of *truth-conditions*, and in turn that *truth-conditions* are all there is to say about *meaning* – and about reference.

The key to the thought of both thinkers in this area is the doctrine that whatever constraints there are on reference, they do their constraining indirectly. Both thinkers reject the notion that a word can be linked to objects directly, prior to the contexts in which it features, its linguistic uses, and so on. They share a fundamental assumption that the only way terms can have reference is by virtue of their roles in linguistic contexts. Thus:

> Uncritical semantics is the myth of a museum in which the exhibits are meanings and the words are labels. To switch languages is to change the labels.

---

cannot see that conceding this makes much difference.

(Quine, 1969: 27)

And:

[...] what can out "understanding" come to, at least for a naturalistically minded philosopher, which is more than *the way we use our language?*

(Putnam, 1983: 4)

Regarding *use* as central underpins their arguments in this area (including the one where Putnam is not paralleled in Quine). For we have seen that all their arguments turn on the possibility of finding ways to assign objects to terms such that some feature of the use of those terms is preserved. In Quine's ontological relativity, that feature is the truth-value of the theory whose vocabulary is under discussion; very closely, in Putnam's permutation argument, it is the truth conditions of all sentences in the language; and in Putnam's argument concerning the epistemically ideal theory, the preserved feature of use is the satisfaction of idealised operational and theoretical constraints plus the truth of that theory. None of these arguments will run if it is possible for terms to be linked to the world, or to admit of standards of correctness and so forth, in other ways apart from the deployment of further terms.

The reason for this focus on *use* of language is, I suggest, its empirical accessibility. Putnam explicitly links his naturalism with the use-doctrine (see the last quotation). *Truth-conditions* come into play because they, too, seem suitably non-mysterious. At the end of the day truth, for the naturalistically minded philosopher, might (with a bit of luck) be reduced to a more basic notion of agreement with others or with experience. This certainly seems amenable to Quine's view. I want to side-step issues about truth here, because they are beyond the scope of the present work. Scepticism about the extent to which use determines substantial truth-conditions is very justifiable. Suffice it to suggest that there is at least *some* plausibility to the idea that using language will tell you quite a bit about the truth-conditions of its sentences. I hope this brief remark will justify my leaving this can of worms shut.

What is less contentious is that both Quine and Putnam share a belief that truth-conditions are the only direct determinants of meaning and reference. They are the intermediaries between *use of language*, which is empirically accessible and thus naturalistically acceptable, and semantic notions of meaning and reference. Remember Quine's contention, previously cited, that "a word may be said to be determined to whatever extent the truth or falsehood of its contexts is determined" (Quine, 1966: 82). That Putnam relies upon a related assumption should be apparent from the following:

> So what *further* constraints on reference are there that could single out some other interpretation as (uniquely) "intended" [...]?
>
> (Putnam, 1978: 126)

This question is rhetorical. So is the next.

> Now, such a model satisfies all operational constraints [...]. And it satisfies those theoretical constraints we would impose in the ideal limit of inquiry. So, once again, it looks as if any such model is "intended" – for what else could single out a model as "intended" than this?
>
> (Putnam, 1983: 13)[18]

There is an obvious difference: Quine makes no explicit reference to operational and theoretical constraints. But towards dissolving this difference, it will be remembered (as I argued in the last section) that Putnam thinks operational and theoretical constraints can only hope to fix reference indirectly, by fixing truth-conditions. Truth-conditions, thus fixed, are then supposed to exhaust the possible ways of fixing reference.

These are complex issues, and I do not pretend to have ironed them out. This doesn't matter, however, so long as the rough ideas I have discussed are conceded. Let me tie up and summarise these ideas. Quine and Putnam share a broadly naturalistic outlook. This disposes them to view

[18] In both these quotations, and elsewhere, Putnam runs together the two threads of argument which I have previously suggested should be distinguished. I do not regard this as a problem for my distinction, which, as I suggested, is organisational rather than exegetical. For the argument concerning the *existence* of a model for the ideal theory need make no mention of the *multiplicity* of such models; and conversely, the permutation argument need not make any mention of the existence of a model for the ideal theory.

meaning as exhausted by truth-conditions (ignoring poetry), rather than as a special kind of entity. For truth-conditions are things that can, plausibly, be empirically accessed, constrained, determined – through the use we make of our language. Such use is entirely accessible to the senses, and thus healthily non-mysterious. And in similar spirit, the determining constraints on reference are sought in the truth-conditions of sentences. To postulate reference as a further sort of thing, constituted in part or whole by things not reducible to truth-conditions and thus ultimately to use, would be distasteful to both, harking back to the kind of medieval dualisms that both hope to avoid.

As Lewis points out (Lewis, 1984: 229), it is no argument to accuse your opponent of being medieval. Many modern philosophers share at least a part of the empirical spirit of Quine and Putnam. Yet they do not regard this as to committing them to the view that the reference of a group of terms can only be fixed by using those terms, on however loose a definition of "use". Intentionality as a mental power is not the popular solution to the problem; nor is it the solution I intend to consider. Instead, philosophers have tended to adopt something like the view Putnam dismisses in the latter quotation – minus the negative rhetoric, of course. They regard Putnam as failing to take on board the possibility that reference is a relation in the world, external to our language and independent of our claims about it – a relation very much in line with the metaphysical realist's picture of everything else in the world. (They may preserve the empiricist spirit in a mission-statement to the effect that finding out about this relation is a job for natural science, or something very like it.) It is this relation, either alongside or independent of our use of terms, that fixes the reference of those terms.

Such a view rejects the Quine-Putnam assumption that terms cannot have their reference fixed in a direct way, independently of other terms or outside linguistic contexts. Yet this line has been recommended frequently and often vigorously against Putnam, much more than against Quine. I hope I have said enough in this section to make it clear that what goes for Putnam, in this regard, goes for Quine: for we have seen that their

arguments turn on the same assumption. It is time to make clear the externalist challenge.

Chapter 3

The Externalist Challenge

Hitherto, I have spoken regularly and confidently of a challenge, which I have dubbed the Externalist Challenge, and which must be met by anyone proposing the sort of arguments about reference that we have seen Quine and Putnam propose. It is time for me to confess what has probably been guessed: that this challenge does not take quite the same form in any two challengers. However, I maintain that there is a common thread running through some of the major responses to Putnam, and once again I propose to pick this thread out while disregarding the other aspects and idiosyncracies of individual presentations.

To this end I will focus on David Lewis's paper *Putnam's Paradox* (Lewis, 1984), which gives stark form to the externalist challenge I hope to meet. In the first section of this chapter, I will try to swiftly isolate and bring out the main point as strongly as possible, and show how Quine is also in Lewis's target zone. In the second section I will air some doubts about the way that Lewis presents the core issue. The third section will formulate the externalist view more clearly, in order to facilitate my attack, which comes in the fourth and final section.

1. Lewis's Challenge

For Lewis, the central problem with Putnam's arguments is that they rely on an incredible view Lewis calls *global descriptivism*. *Descriptivism* about some term or group of terms is the view that their reference can be fixed by description using other terms:

> [...] we associate clusters of old-language descriptions with our new terms; and thereby, if the world co-operates, we bestow reference on our new terms.
>
> (Lewis, 1984: 222)

This is a familiar view, we are told, and one with familiar objections. Lewis thinks that these objections are not fatal, and that a local descriptivism is a useful theory of "how to get more reference if we have some already" (1984: 223). But what Putnam relies on is the *global* version:

> The intended interpretation will be the one, if such there be, that makes the term-introducing theory come true. [...] But this time, the term-introducing theory is total theory! Call this account of reference: *global* descriptivism.
>
> (Lewis, 1984: 224)

This is the view that Lewis claims Putnam relies on, and it is one that he rejects. He rejects it precisely because "it leads straight to Putnam's incredible thesis" (1984: 224).

"Putnam's incredible thesis", for Lewis, is primarily that presented in *Reason, Truth and History* (Putnam, 1981). It is what I referred to in the previous chapter as the permutation argument, and it is an attack on the determinacy of reference – by being an attack on the uniqueness of the reference relation. Lewis does not really distinguish the other main version of Putnam's argument (model existence) identified in the previous chapter. However, he does discuss *Models and Reality* (Putnam, 1980), and his favourite summary of Putnam's view seems to refer more to the model existence version of the argument than to the permutation argument. His summary is that "almost any world can satisfy almost any theory" (Lewis, 1984: 227). This failure to enforce the distinction I made last chapter does not matter from Lewis's point of view, because his objection would be a problem for both versions. For the model-existence argument and the

permutation argument both rely on the same assumption – whether or not Lewis has correctly characterised it – that terms get their reference fixed through having their contexts of use fixed. This licenses the notion of a model to which Putnam appeals: assignments of objects to terms need be answerable to nothing else, than making all the contexts where that term is used come out true or false as desired. It should also be clear, if the argument of the previous chapter holds, why Lewis must also address himself to Quine: for I argued there that Putnam's permutation argument is extremely close to Quine's argument for ontological relativity, and relies on the same crucial assumption.

Putnam defends his assumption in various ways. His discussion in *Reason, Truth and History* appeals to several distinguishable thoughts, whose compatibility with one another is sometimes open to question. However his most famous line of defence, upon which I want to focus, is better expressed by the following oft-cited sentence from *Realism and Reason*:

> Notice that a "causal" theory of reference is not (would not be) of any help here: for how "causes" can uniquely refer is as much of a puzzle as how "cat" can.
>
> (Putnam, 1978: 126)

This summarises, with specific but inessential mention of causal theories of reference, what is known as the "just more theory" argument. In *Reason, Truth and History*, he presents it more substantially, using Field[19] as a foil. Putnam argues:

> Suppose there is a possible naturalistic or physicalistic definition of reference, as Field contends. Suppose
>
> (1) *x refers to y* if and only if *x bears R to y*
>
> is true, where R is a relation definable in natural science vocabulary without using any semantical notions [...]. (1) is a sentence which would be part of our 'reflective equilibrium' or 'ideal limit' theory of the world.
>
> (Putnam, 1981: 53)

[19] Field argues for a physicalist theory of reference in *Tarski's Theory of Truth* (Field, 1972).

He then refers us to his permutation argument, which shows that even the ideal theory does not have determinate reference. Ideal theory includes a theory about R, the reference relation. But "R" does not have determinate reference, any more than the rest of ideal theory. So making a theory of R part of the ideal theory won't help us fix reference: because we are just appealing to more theory, whose objects can also be permuted. This is the "just more theory" response, and it is to this that Lewis objects.

Lewis suggests that there might be other constraints on reference, besides the indirect action, via truth-conditions of sentences (or other contexts), of operational and theoretical constraints.[20] Lewis's counter is straightforward:

> [According to Putnam,] Constraints that work within it [global descriptivism] are the only possible constraints on reference. His [Putnam's] reason is that global descriptivism is imperialistic: it will annex any satisfactory account of constraints on reference. [...] To which I reply: *C* [the reference-saving constraint, whatever it may be] is *not* to be imposed just by accepting *C*-theory. That is a misunderstanding of what *C* is. The constraint is *not* that an intended interpretation must somehow make our account of *C* come true. The constraint is that an intended interpretation must conform to *C* itself.
>
> (Lewis, 1984: 225)

So Lewis accuses Putnam of a fairly simple mistake: he is confusing the claim that a certain something – *C* – constrains reference, with the claim that the theory about that constraint – the theory about *C* – constrains reference. Lewis's proposal is not *just* more theory: if he is right (or possibly even if he is wrong), then it's not his theory, but the facts that his theory hopes to express, that constrain and constitute the relation of reference.

This central thought comes out very clearly from Lewis's paper. In the next section I question an aspect of Lewis's presentation, and in the final

[20] Lewis, like me, struggles to clarify the exact status of operational and theoretical constraints in Putnam: "It is hard to tell from his words whether these are supposed to constrain reference or theory. Probably he thinks they do both: they constrain ideal theory, ideal theory is the term-introducing theory to which global descriptivism applies, so in this indirect way they constrain reference also." (Lewis, 1984: 224.) This is pretty much how I see it.

section I will attempt a rebuttal. But I will close this section by emphasising that Lewis's objection does not rely on any particular theory of reference. In particular, Lewis is diffident about causal theories of reference:

> The causal theory often works, but not as invariably as philosophers nowadays tend to think. Sometimes old-fashioned descriptivism works better; sometimes there are puzzling intermediate cases[...].
>
> (Lewis, 1984: 227)

In fact, Lewis's own proposal is not terribly mainstream. He suggests that certain classes of things are intrinsically more eligible to act as referents for words:

> Among all the countless things and classes that there are, most are miscellaneous, gerrymandered, ill-demarcated. Only an elite minority are carved at the joints, so that their boundaries are established by objective sameness and difference in nature. Only these elite things and classes are eligible to serve as referents. When we limit ourselves to the eligible interpretations, the ones that respect the objective joints in nature, there is no longer any guarantee that (almost) any world can satisfy (almost) any theory.
>
> (Lewis, 1984: 227)

I am not inclined to accept this as a theory of reference, even a schematic one, and I suspect that it would be a fairly substantial pill to swallow for many philosophers. For one thing, it is very metaphysical; many modern philosophers aspire to a higher degree of what might loosely be called naturalism, and are suspicious of metaphysics' ability to tell us about the world. But let us not be side-tracked: the strength of Lewis's attack on Putnam is that it does not rely on any particular theory of reference. It may not be compatible with every theory of reference, but it does not confine your choice to one.

This is just as well for Lewis. For Putnam explicitly anticipates Lewis's proposal, saying: "[...] 'of the same kind' makes no sense apart from a categorical system which says what properties do and what properties do not count as similar" (Putnam, 1981: 53). Maybe not. But that has no direct bearing on Lewis's criticism of Putnam. If that criticism is to be deflected, it must be tackled directly.

2. A Question About Lewis's Diagnosis

I suspect that Lewis mislocates the crucial assumption of the permutation argument. I have given my own account of that crucial assumption at the end of Chapter 2, but on the face of it perhaps Lewis's analysis need not disagree with mine. The problems I want to raise for Lewis's diagnosis come from a different direction – namely, the problems mentioned in Chapter 1 concerning implicit definition.

As discussed in the previous section, Lewis characterises Putnam's mistaken premise as global descriptivism. This is the view that the reference of our extralogical vocabulary is fixed by "total theory" – the view, in effect, that reference of our vocabulary is fixed by sentences constructed out of that vocabulary. This is a recipe for indeterminism of reference, for the kinds of reasons we have seen Quine discuss: multiple models of the sentences will be available, no matter how many sentences we add to our theory. Lewis contrasts the perfectly wholesome *local* variety of descriptivism, whereby we can get more reference if we have some already. Local descriptivism is the view that the reference of some set of terms is fixed by sentences involving terms, whose reference is already fixed, which are not elements of the set being defined.

What I suggest in this section is that there are good reasons for doubting that local descriptivism escapes the permutation net. I outlined one such reason in Chapter 1, where I discussed the contrast between *implicit* and *explicit* definition. As they were presented there, these both fall under the title of local descriptivist techniques. I argued that *explicit* definition could preserve determinacy of reference, by specifying with the old, reference-fixed terms exactly what structures were represented by the newly introduced terms. However, we saw that this strategy will never enable us to represent structures with our new terms that are not already representable with the old terms. *Implicit* definition of the newly-introduced terms, on the other hand, does not restrict us to the structures of the old terms. However, I suggested that it does not allow us to distinguish between models of sentences of the language containing the new terms, even if – somehow – the models of the old-language sentences had been identified beyond

isomorphism and indiscernibility (i.e. the reference of the old terms is fixed.) The reason is that implicit definition involves specifying necessary and sufficient conditions for the truth of *some* sentences containing the new term, but not *all* possible sentences. Usually the *some* are identified by being those of a given form; classically, we might define "direction" in terms of being parallel without saying anything about the truth-conditions of other forms of sentence involving "direction". Intuitively, we would not have said what a direction is, only that there are certain things that it is not.

If I am right about this, then there is a question as to how Lewis thinks local descriptivism is supposed to work "to get more reference if we have some already" (1984: 223). Explicit definition won't get more reference, in the sense that it won't enable us to use our new terms to refer to anything we couldn't have referred to with our old terms. Implicit definition won't get more reference in the sense that it won't enable us to identify models of (at least some of) the new-language sentences beyond isomorphism, which is pretty much a succinct way of putting Putnam's permutation result. It appears that global descriptivism is no more of a culprit than local descriptivism.

Of course Lewis is very well known for his work in this area and I can hardly accuse him of ignorance of these issues. These considerations are very general, and to make them more convincing I need a more detailed argument. Luckily one has been provided by Winnie, in his short paper *The Implicit Definition of Theoretical Terms* (1967: 223-229).

Winnie's explicit concern is with the problems of fixing the reference of terms whose referents are not empirically accessible – *theoretical* terms – using terms whose referents are empirically accessible – *observational* terms. It is regularly pointed out that a sharp theory/observation distinction doesn't stand up well to scrutiny, either as a distinction between types of entities or between types of terms. In *How to Define Theoretical Terms* (Lewis, 1970: 427-446) Lewis suggests, plausibly, that the best distinction to recognise is not a general one, but is specific to the theory in question. So for some theory T, there will be T-terms – which are just new terms peculiar to the theory – and O-terms,

which are all other terms, and with which we are already assumed to be familiar. But the official motivation for Winnie's argument is a little misleading, because his arguments do not depend on a principled theory/observation distinction, among either entities or terms. Thus denying that distinction does not go all the way to refuting his arguments.

So what are his arguments? He offers two theorems and two proofs. The first hopes to show that "[t]here would always be, given one true interpretation, another true interpretation" (Winnie, 1967: 226). The second hopes to show that "under certain trivial assumptions, a true numerical interpretation will always be forthcoming" (1967: 225) – in other words, the entities designated by the theoretical terms might not even be physical objects at all, but numbers.

Without going into details, the proofs both work in a similar way to Putnam's permutation argument. The proof that there is another physical model of the theory proceeds by defining a one-to-one function from the universe of the vocabulary of the theory onto itself such that if some object is part of the universe of the theoretical vocabulary, then in at least one case its image under the function will be a different object, whereas the function will map objects in the universe of the observational vocabulary onto themselves. The result will be an isomorphic but non-equivalent model of the entire vocabulary of the theory. The proof regarding the existence of a numerical model proceeds similarly by defining a one-to-one function between the universe of the theoretical vocabulary and a set of numbers of the same cardinality.

It should be plausible from my formulation, even if not from Winnie's, that that these arguments should work using Lewis's weaker definition of a theoretical term. For the proof doesn't rely on any special properties for theoretical terms other than that the ones under consideration should have their reference fixed solely by the other terms in the language. This is a property that Lewis appears to accept in *How to Define Theoretical Terms*.

Lewis has a response to Winnie, but before turning to it, notice that Lewis *can't* make the externalist move at this stage. For instance, it would be damaging to him if he proposed (along the lines of his proposal in *Putnam's Paradox*) that the objects referred to by theoretical terms should not be gerrymandered, ill-demarcated or miscellaneous. This move would damage Lewis because it would belie his claim to have located the source of Putnam's paradox in global descriptivism. If appeal to world-imposed constraints on reference is necessary even in the case of local descriptivism to prevent permutation-type results, then global descriptivism can't be the problematic assumption underlying those results.

Lewis's discussion of Winnie is brief. He accepts Winnie's arguments, but he says that he doesn't agree with Winnie's way of treating *O*-terms:

> I am concerned only with realizations under a fixed interpretation of the *O*-vocabulary; whereas Winnie permits variation in the interpretation of certain *O*-terms from one realization to another, provided that the variation is confined to theoretical entities. For instance, he would permit variation in the extension of the *O*-predicate '__is bigger than__' so long as the extension among observational entities remained fixed. Winnie's proof does not show that a theory is multiply realized in my sense unless the postulate of the theory is free of "mixed" *O*-terms[...]. I claim that mixed *O*-terms are omnipresent[...].
>
> (Lewis, 1970: 434)

Here Lewis appears to rely heavily on the fact that his notion of an *O*-term and a *T*-term differs from Winnie's. His point is that Winnie's proof works by allowing that when a familiar term – an *O*-term – is applied to a theoretical entity (one to which we do not have direct empirical access, whatever that might be), its extension is up for grabs. He is attributing to Winnie the view that "__is bigger than__" might mean something different when applied to, say, electrons and protons, than it does when applied to, say, bricks.

This is a strange way to dismiss Winnie. I have already suggested that one might agree with Lewis in rejecting Winnie's explicit project and formulation in terms of observational and theoretical entities, while

nevertheless remaining interested in a version of Winnie's argument. *Prima facie* it is plausible that Winnie's argument might be amended to patch the hole Lewis spots. Winnie could simply accept that *O*-terms such as "__is bigger than__" mean the same regardless of the entities to which they are applied. He would still be able to carry out a permutation in many cases. For example, the claim that a proton is bigger than an electron closes off the possibility of swapping protons and electrons, since protons are bigger than electrons but electrons aren't bigger than protons. But that can scarcely be hoped to reduce the number of eligible proton candidates to one. Nor, indeed, can we rely on the addition of further similar sentences to do so. Structures will be eliminated; but the possibility that there are multiple models of the sentences in question can't be ruled out merely by those sentences, even if the reference of some of the terms in those sentences is fixed.

I don't claim to have proved that there will always be multiple models. I merely claim that Lewis's point doesn't show that there won't be. And as such it is a puzzling reply to Winnie's argument that there are.

A more general point deserves mention. Many scientific theories don't involve terms like "__is bigger than__" in any significant way. The hope that we might, at least sometimes, succeed in implicitly defining terms such that they uniquely refer by using an accepted vocabulary would seem most justified in cases where there was a high ratio of accepted terms to unaccepted terms. So we might say, for example, "A *leopard* is a creature that steals goats in the dead of night", thereby implicitly defining "leopard" to some degree, possibly even uniquely if there don't happen to be any other predators in the area. This definition is implicit because I haven't said what a leopard is; I haven't given necessary and sufficient conditions for leopardhood. If the definition results in unique reference, that's a lucky result partly contingent upon environmental factors, and also – plausibly – partly due to the extreme empirical familiarity of all the other terms involved ("night", "goats", etc). But it is open to question whether the implicit definitions that scientific theories are thought to constitute are similarly loaded with familiar terms. Some of our best ones, at the point

when they are introduced, apparently redefine many of the terms they ostensibly inherit (if any), as well as defining new terms (if any). Lewis's claim that mixed *O*-terms (applying to both observational and theoretical entities) are omnipresent is perhaps an exaggeration.

I can hardly pretend to have settled these matters here. My objective has simply been to point out that Lewis's confident characterisation of the problematic assumption in *Putnam's Paradox* is open to question. However his attack on Putnam doesn't rely on this characterisation – a further reason to be suspicious of the assertion that the underlying assumption has been captured, since an attack might be expected to challenge the underlying assumption. To this attack I now turn.

3. Clarifying the Externalist Challenge

Before I try to meet the externalist challenge, I will characterise as best I can the view that I am attacking. It will then become clear what form my attack shall take.

(A) The reference relation must meet requirements imposed by us and our experience.[21]

(B) The reference relation must meet requirements imposed by the world itself, independent of our experience.

(C) There will be at least one relation meeting (A) and (B).

(D) There will be at most one relation meeting (A) and (B).

I will refer to (A)-(D) as Typical Externalism. Let me justify this title by briefly expanding upon (A)-(D).

With (A), I am driving at the sort of constraints that Putnam considers: truth-conditions of whole sentences as determined by operational and theoretical constraints, for example. It's hard to be precise, but the distinguishing feature of this sort of constraint is that it is imposed on the reference relation by the use we make of our language. It won't do to say that these sorts of constraints make essential reference to us; theoretical constraints do not. *Use* comes in, although I suspect it isn't the whole story. I gave my best characterisation of this sort of constraint at the end of the last chapter. Lewis would substitute, for (A), a clause to the effect that reference must make the term-introducing theory come true. I have already discussed the problems with his characterisation. None of this is critical for present purposes, though. The Typical Externalist will accept that some constraint of this sort needs to (or will be) met by a reference relation. Putnam shows how easy it is to meet such constraints, and the Typical Externalist admits the validity of his arguments (see Heller and Lewis) or of some patched version.

[21] Putnam and the realist will take this claim differently. For the realist it would be a substantive claim, for on her view the true theory need not be epistemically ideal and the reference relation between its terms and the world need not meet our empirical constraints. However, I am working on the basis that scepticism is alien to the realist temper. For this

According to the Typical Externalist, it is denying clause (B) that makes Putnam's valid argument unsound. (B) says that there are certain other characteristics that the reference relation must possess – other constraints it obeys. These constraints are not imposed by us, or by our experience, or by our use of language or the contexts in which it is used. This is *contra* Quine's comment, already quoted in the last chapter, that "in point of *meaning* [...] a word may be said to be determined to whatever extent the truth or falsehood of its contexts is determined" (Quine, 1966: 82). Familiar considerations will be cited – for example, that the reference of "electron" was determined even on the lips of Bohr, partially independent of the subsequent tightening of determinacy over the truth and falsehood of its contexts. What determines the reference of "electron", now as back then, is not just constraints of the sort mentioned in (A) – although these play a part. It is also constraints imposed by the world, independently of us. In order to be reference, a relation must meet certain criteria, and these criteria are independent of us and are not imposed by us or by our experience.

Another way of characterising these constraints is to say that they are "direct": they constrain what can count as the reference relation *directly*, rather than indirectly. This is an equivalent characterisation: for direct constraint turns out to mean constraint unmediated by anything of a sort that involves us, or constraint on those properties of the relation that do not involve its role in our lives (preserving truth conditions etc).

Claims (C) and (D) jointly embody the externalist belief that (A) and (B) solve the Quine-Putnam riddle. They do not follow directly from (A) and (B). So are they supported by an argument?

The argument supplied by Lewis is an argument from intuition:

> Putnam's thesis is incredible. We are in the presence of a paradox [...]. It is out of the question to follow the argument where it leads.
>
> (Lewis, 1984: 221)

---

reason I suggest that most realists would also accept the weak epistemological element that this claim embodies for them.

According to Lewis, (C) and (D) are indubitably true, and the job of the conscientious philosopher is to show how Putnam's paradox can be circumvented. This gives Lewis's argument the form of an inference to the best explanation, where the best explanation of the joint truth of (A), (C) and (D) is (B) – the external constraints clause. This attitude is implicit in much of what is written by other writers; and perhaps rightly so. To deny the determinacy of reference outright is paradoxical. (For what, if anything, is being denied?) That does not mean, however, that *asserting* the determinacy of reference, in the way that Lewis does, is the only alternative to denying it. There may be another way. And we might be motivated to find some other way if we found the externalist "best explanation" wanting – in other words, if we find (B) wanting.

Therefore my attack will be an attack on the explanatory virtues of (B). This is the work of the next section. But my discussion of a possible alternative view of these matters waits for the next chapter.

4. Meeting the Externalist Challenge

I don't think that the externalist challenge is an inconsistent position. I do not have an argument that reduces it to contradiction. What I do think I can show is that externalism is inconsistent with a broader outlook – naturalism. I don't want to get bogged down trying to draw the boundary lines of naturalism precisely (I'm not sure it can be done anyway). Instead I will aim to bring out the aspects of the externalist challenge which clearly cross those vague boundary lines.[22]

A rough characterisation of naturalism is still useful. The idea I am working with is the one explicitly presented by Quine in the opening pages of *Ontological Relativity*, where he aligns himself with John Dewey. Its components include a rejection of the notion that we have anything to learn from a prior metaphysics, and a corresponding approval of the methods of science. It often includes some sort of physicalism or monism – a belief that there is only one sort of thing, and that that sort of thing is the sort of thing that science tells us about – physical stuff, roughly. This bundle is the prevailing starting point for philosophical discussions in our time and place; many modern philosophical problems derive in part from trying to come up with explanations for stubborn phenomena that don't easily or obviously fit the naturalistic framework. Assessing this framework itself is an unmanageably broad task for this work. Nevertheless, a return to the days of systematic *a priori* metaphysics looks unlikely. To this extent at least, we can call naturalism the predominant standpoint – or hue, or tone – of current analytic philosophy, regardless of its various possible difficulties as a fully-fledged position.

At any rate, it is to the broadly naturalistic philosopher that I address my argument. It is difficult to argue with someone who is metaphysically creative, and that is certainly a skill Lewis is willing to exercise. What I want to suggest is that the position Lewis reaches on reference, while consistent, is not plausible. And plausibility is, by and large, relative to the

[22] My argument takes some inspiration from Andersen's work (Andersen, 1993). His way of setting up the problem I find somewhat misleading, in ways that it would be too digressive to discuss here; but there is something interesting in his distinction between *metaphysical* and *empirical* notions of causality, that is close to my argument.

predominant view. In adopting this tactic I am fighting fire with fire, for Lewis's argument is itself an argument from plausibility.

My argument, in a nutshell, is this: The claim embodied by (B) is a claim of transcendental metaphysics, not a claim about the empirical world. It is not an open question whether these external constraints on reference are to be found out by empirical science (*contra* Heller (1988: 125)): it is a closed question. The kind of constraints needed to do the job are empirically inaccessible. The claim that there *are* such constraints is therefore a claim transcending the limits of what our senses can tell us or could tell us in any possible world. Lewis's argument is neither empirically based nor credible as an *a priori* argument: it belongs to the sort of transcendental metaphysics which a naturalistic thinker will find unpalatable.

To see this, first consider that empirical constraints are of the sort mentioned by (A) above. In the last chapter I briefly discussed the difficulty of characterising empirical (operational) constraints. We pre-philosophically feel that experience does offer some direct access to the world, but this access is hard to characterise, without either opening the door to scepticism or introducing an air of magic. Not to say it can't be done: but I don't know how to do it. Luckily, however, I don't need to. For the externalist typically accepts that empirical constraints on reference belong to category (A) above – that is, she typically accepts that the reference relation is underdetermined by empirical constraints. Her proposal is that there is another kind of constraint, *apart* from those imposed by our senses and our grasp of formal theoretical properties (and anything else we can directly grasp). These other constraints I call type (B) constraints.

It does not follow directly, however, from the claim that constraints of type (B) are not *empirical* that they are not empirically *accessible*. It might be the case that we can obtain empirical access to the constraints on reference, but that these constraints do not act via our senses. This is the background view behind theories put forward by the likes of Fodor and Dretske. Dretske thinks that information (including that carried by the reference relation) consists in some sort of *law-like correlation* between states of the organism and states of the world (e.g. see Dretske, 1980).

Plausibly, such a link often acts via the senses – an organism obtains information in this way. But the link is nevertheless to be thought of as *external*, in our present sense. For the primary constraint on reference is that it should be a relation conforming to some particular type, which Dretske then tries to characterise. Whether or not experience is involved is beside the point. And plausibly we might obtain empirical *access* to the mechanisms in various organisms, including our own species, that enforce this constraint. Similarly in the case of Fodor, who believes that when the list of fundamental properties is finally drawn up in a completed science, intentionality won't be on it (Fodor, 1987). Empirical investigation might reveal the way in which a special *asymmetric nomic dependence* is enforced by biological mechanisms. On views like these, there is plausibly room for a non-empirical reference relation, whose instances are or may be empirically accessible. This argument is somewhat analogous to the more familiar claim that even though my visual cortex doesn't constrain my vision by visually appearing, it nevertheless does constrain my vision, and the way that it does is empirically – even visually – accessible.

I want to argue against any such view, so it would be unconvincing for me to ignore these possibilities and move straight from the uncontroversial claim that empirical constraints are of type (A) to the claim that the reference relation is empirically inaccessible.

So the next stage in my argument is to claim that (C) and (D) are also unfounded. That they are not supported by empirical evidence follows from the admission that empirical constraints are included in (A); that it follows is shown by the Quine-Putnam arguments. First take (D), the uniqueness clause. Putnam's permutation argument shows that all empirical evidence[23] is compatible with the reference relation being non-unique. That is, all empirical evidence is compatible with (D) being false. (Do not be tempted to assert that the simplest, most elegant, etc. interpretation of the evidence will be unique: Putnam's inclusion of theoretical constraints in his

[23] Along with some other things, in Putnam's arguments – namely, formal constraints on theory. Also, nothing relevant hangs on whether "all the evidence" means "all actual evidence" or "all possible evidence": the indeterminacy result can be obtained in either case, as discussed in the previous chapter.

argument rules this out.) Therefore any assertion of uniqueness must be entirely uninformed by experience, since the no empirical evidence has any bearing on the question of whether reference is unique – being equally compatible with its non-uniqueness. Yet we have hardly been offered an *a priori* proof, along the lines of mathematics or logic. Now take (C), the existence claim. The claim that there will be at least one reference relation meeting constraints of type (A) is nearly trivial; this is shown by Putnam's model-existence argument. The claim that there is at least one reference relation satisfying (B) is by no means trivial. It cannot be supported by empirical evidence because any empirical evidence will count as a type (A) constraint. Empirical evidence can only ever show that there is (or is not) a reference relation satisfying constraints of type (A). Hence the claim that there is a reference relation satisfying constraints of type (B) must be entirely without empirical basis. Nor has it been supported by any *a priori* argument. This seems a reasonable basis for a charge that the claim is unwarranted.

Putnam also assumes that reference is determinate. However my discussion affects him less, because the things to which he think we refer are, on his view, somehow constituted by or to do with us. In his hands, "determinate reference" means something which we *can* empirically access: it means precisely that our words refer to independent objects up to the extent determined by our experience – which is not very far. He denies that there is any sense in talking about reference in that way; instead, he regards it simply as the relation between words and objects in our experience, subject only to the constraints of experience. His claim becomes an empirical one, for the existence of such a relation is open to empirical confirmation or disconfirmation.

Note that here, there is no parallel move available to the one we ascribed to Dretske and Fodor a few paragraphs back. There, I said that the mere fact that constraints of type (B) are not empirical in their action does not imply that we can't access them empirically. Why can't the externalist also assert that even though the non-existence or non-uniqueness of reference are compatible with empirical constraints, we can have empirical

evidence for them? The externalist can't hold this because it is not a consistent position. If constraints of type (A) are to include all empirical constraints and all theoretical constraints on reference, as Putnam has it, then – according to Putnam's argument – there is no possible way in which we might find or construe any empirical evidence to confirm or disconfirm (C) or (D). Theories like those of Dretske and Fodor are perfectly legitimate on a certain level – on the empirical level. They answer empirical questions about reference. But they simply fail to address the sort of issue we are dealing with at present; for empirical theories can't be extended to metaphysics. If, on the other hand, their theories are to be taken as encompassing a metaphysical element of the sort we have seen Lewis proposing, then they lose their claim to being naturalistic, common-sense, or down-to-earth – and their relation to empirical evidence.

I am arguing that the two sorts of claim are distinct. Either you can have a theory of reference that tells you, along the lines of a natural science, what natural properties and relations are involved in reference. But any such theory is only as good as the rest of scientific theory when it comes to dealing with the Quine-Putnam challenge – it's more grist for the mill. On the other hand, you can propose a theory that seeks to avoid the mill by specifying that the constraints *themselves*, and not our theories of them, do the constraining. But then the claim that is doing work in your theory is the claim expressed most clearly by Lewis. It is a metaphysical claim that, I have argued, is independent of any empirical evidence you may be able to cite in support of your favourite law-like correlation, asymmetric nomic dependence, or whatever. And thus it should raise the hackles of those theorists whose intentions, as in the cases of Dretske and Fodor, are avowedly anti-metaphysical.

What I hope to have shown, therefore, is that of claims (A)-(D), only (A) is capable of any kind of empirical assessment. Nor are (B)-(D) supported *a priori*. Thus they are unwarranted.

Perhaps it will be granted that empirical evidence is beside the point, but argued that Lewis's argument is, in fact, *a priori*. After all, I have admitted that an argument is present somewhere – an inference to the best

explanation. *"A priori"* is not an insult. But notice two things. First, if his argument is *a priori* Lewis's argument from plausibility is an argument from *a priori* plausibility. It is *not* an argument on a par with, for example, Putnam's famous miracle argument for scientific realism. The latter argument suggests that our empirical evidence would be suboptimally explained by anything except scientific realism. Lewis's argument can't be that the empirical evidence concerning reference is likewise suboptimally explained by anything except a single unique reference relation. It can't be that because Putnam shows that the best explanation *can* accommodate non-unique reference, and I have argued that non-existence of a reference relation complying with constraints of both types (A) and (B) can also be accommodated. So Lewis's argument must be that the existence and uniqueness of a reference relation complying with (A) and (B) is plausible *a priori*. This claim is extremely strong.

The second point to note is that we have no argument for accepting this extremely strong claim. Empirical evidence we have already ruled out. If (C) and (D) are meant to be *a priori* plausible, then we might expect some indication of the *a priori* source of this plausibility. The most sincere introspection, in my case, does not reveal it. (Indeed, the plausibility of the claim that there exists a unique reference relation seems to be derived completely from experience: and experience can never support a claim about constraints of type (B).) Perhaps an argument would help: there is nothing wrong with *a priori* arguments as such. But no argument is on the table. Lewis does not argue, as he needs to, that (B), (C) or (D) is on a par with other notions supported by *a priori* plausibility – the law of identity, for instance.

What I am suggesting, therefore, is that the externalist challenge is a logically consistent but extremely strong substantive claim about the world, whose basis is a plausibility claim. I do not accept that the plausibility claim is either empirically or *a priori* supported. The only plausible claim in the vicinity is empirically based, and that claim cannot support the externalist's claims (which transcend empirical bounds). That is the more modest claim that reference is determinate to the extent determined by our experience.

I now turn to a discussion of this more modest claim, in the hope that a more plausible alternative to the externalist picture is available.

Chapter 4

Another Way

I begin this final chapter by summarising the debate up to this point. After Section 1, I assume that I am right about the outcome of that debate; the rest of the chapter is devoted to seeing what consequences my being right would have. The second section explores two of the less attractive ways to understand those consequences. One way is to see them as paradoxical, leading to a contradiction, or at least to a view that is self-defeating. Another is to see them as motivating the abandoning of metaphysical realism in favour of some other metaphysical view. Section 3, the last, presents my own view of the situation. There I try to distinguish between two ways in which determinacy of reference may be asserted; in one way it is trivially true, and in the other way it is false, as shown by the Quine-Putnam arguments.

1. Closing the Dialectic

It may be helpful to summarise the main thread of the dialectic between the Quine-Putnam arguments and the externalist challenge. We will then be in a better position to consider what the implications are if I am right about the outcome of that dialectic.

Quine and Putnam both present arguments sharing certain important features. Their arguments both involve the idea that the assignment of objects to terms, which constitutes reference, is constrained by the way in which we use our language. They then rely on the further claim that nothing else is available to constrain this assignment, to construct arguments to the effect that the available constraints underdetermine reference.

Since they argue that reference is not determined by any of the candidates available, you might expect Quine and Putnam to conclude that reference is in actual fact indeterminate. Yet this isn't quite what happens. Quine draws a conclusion that is something along these lines, although a full discussion must wait until the third section of this chapter. But Putnam chooses to alter his metaphysics instead, in ways which make the world somehow dependent on us. This is meant to thwart the permutation-type arguments because our experience, which is supposed to play the key role in determining reference, also plays a key – even constitutive – role in determining the nature of the world.

But before we continue, externalists, such as Lewis, object. Lewis draws attention to the further claim, relied on by Quine and Putnam, that there is no other sort of constraint available to fix reference apart from the constraints that are imposed by fixing contexts of use. Lewis points out, correctly, that unless supported on independent grounds, this reliance is a weakness in the Quine-Putnam argument.

Quine does not explicitly address this question, but Putnam has a response. He says that the theory describing any constraint proposed to fix reference – regardless of whether that constraint is external – will be liable to similar treatment as any other theory. The Quine-Putnam arguments will still apply to our theory of the reference-saving constraint, just as they apply

to all our other theory. And so proposing an "external" constraint won't help. Proposing anything won't help: whatever is proposed will be subject to the same permutation considerations; and just adding more theory won't block them.

Lewis replies, correctly in the eyes of most, that this is a misunderstanding. The force of the externalist challenge is to suggest that it is the constraint *itself* that fixes reference, not the theory of that constraint. A relation exists between the world and our words, independent of the use we make of those words, which allows those words to refer in the determinate way that we customarily take them to.

In Chapter 3, I constructed an argument that casts some doubt on the plausibility of Lewis's suggestion. The form of Lewis's argument is an inference to the best explanation. For reasons that will be discussed in the next section, Lewis thinks it is absurd to conclude that reference is indeterminate. He is also unwilling to countenance metaphysical adjustments of the sort Putnam proposes (and, in places, sceptical that these really answer Putnam's own arguments). Putnam assumes metaphysical realism, then tries to show it has an absurd consequence – that reference is indeterminate; Lewis also assumes metaphysical realism, and tries to show how indeterminacy of reference is not a consequence. So what he seeks to explain is determinacy of reference, given metaphysical realism and the validity of arguments to the effect that the conjunction of these two is false. He does so by postulating an external reference relation, whose existence would make the valid permutation arguments unsound.

Due to the structure of inference to the best explanation, the degree of confirmation that is conferred on this conclusion depends not only on the merits of the explanation but also upon the plausibility of that which is being explained. In its common scientific and everyday employment, inference to the best explanation is usually inference *from* an obvious and incontrovertible empirical fact – such as the presence of rainbows on particular occasions when the sun shines in the rain. Inference to the best explanation, involving the interaction of sunbeams and raindrops, is licensed in this case by the independent certainty of what is being explained. Without

this, inference to the best explanation does not make much sense. For there is little confirmation to be gained for Theory 1 by showing how Theory 1 explains Data 2, if we are not already sure of the accuracy of Data 2. For instance, the inference from the existence of ghosts to the claim that our souls persist after our bodies perish is unconvincing. It remains unconvincing even if we assume, for the sake of argument, that the claim about souls is the best explanation of the existence of ghosts. For the existence of ghosts is itself doubtful – in contrast to the appearance of rainbows in certain conditions.

Perhaps I should briefly defend my interpretation of Lewis as proposing an inference to the best explanation against a possible alternative. This alternative lacks support in the text but may be thought to avert my argument against Lewis. There is a common sort of reasoning where two theories are postulated together, and their mutual support is considered evidence for their truth (provided they adequately explain whatever low-level data need explaining). For example, the theory that the sea level is rising and the theory that greenhouse gases are warming the planet and melting the ice caps might each gain something by being postulated together. (This is not a historical example.) With no theory about global warming, the data might leave us unable to decide between the theory that the sea level was rising and the theory that the land was subsiding. And similarly, if we didn't think the sea level might be rising, we would have less reason to suppose that the climate was warming up[24] – and thus less reason to believe the theory of global warming. Clearly this is a version of inference to the best explanation: for the theories do ultimately have to explain some data. It's just that the theories also gain a substantial degree of support by explaining each other. And this sort of mutual support doesn't require initial certainty of either theory, at least not in the obvious way that inference to the best explanation require certainty of that (be it theory or data) which is explained.

[24] Of course our thermometer readings would not be in question: rather we would be wondering whether the various readings we had taken in recent years indicated a blip, or a general and global trend.

Perhaps it will be suggested that Lewis's argument runs more along these lines. If so, then Lewis would have to be seen as simultaneously postulating both determinate reference and an externalist reference relation, and relying on their mutual support as evidence for their joint truth. However this is obviously weak. For they would still, together, need to be taken as the best explanation for the data provided by experience. Ultimately this form of reasoning, and hence this interpretation of Lewis, is still be an inference to the best explanation. The difference is that this time two theories are jointly providing explanation for the data. However in the present case the two theories in question clearly do not, even jointly, provide the best explanation for the relevant data. Considering *only* that data – that is, experience – a much simpler explanation is available – namely that experience is the only constraint on reference.[25] It has already been argued, and is surely hard to contest, that experience on its own cannot provide a reason to think reference is determinate beyond the extent to which it is determined in experience. Moreover this "joint postulate" interpretation is clearly deviant from Lewis's text. So I stick by my interpretation of Lewis as proposing an inference to the best explanation of determinate reference. This interpretation attributes to Lewis a far stronger argument.

In Chapter 3, I claim that Lewis's starting point is not independently plausible in the way that it needs to be for inference to the best explanation to work. In particular, I argue that Lewis has provided no independent, non-circular grounds for asserting the determinacy of reference he seeks to explain. I say that the only evidence we might have that reference is determinate, and hence the only reason we can have for asserting that it is, comes from experience, whose role in fixing reference is (the externalist challenger agrees) limited to fixing the contexts of use of the terms in question. Nor do we have *a priori* arguments for the independent plausibility of referential determinacy. Thus I charge Lewis's reasoning with exemplifying speculative transcendental metaphysics, out of keeping with the naturalistic tradition that predominates in our philosophical culture.

[25] This much the Quine-Putnam arguments show, and are admitted to show even by externalists. Otherwise, why postulate a further constraint on reference?

So the externalist position is a consistent one; but its supporting arguments are weak. It remains open that a more plausible resolution of these matters may be found. I believe that the permutation arguments are most clearly understood as I presented them in Chapter 2 – as assuming metaphysical realism and then showing that, on that assumption, reference is not determinate. In the next section we will discuss this conclusion, and find it unattractive. Another way to interpret the situation is to see the arguments as showing that metaphysical realism and determinate reference are not compatible, and then to reject metaphysical realism in order to preserve determinate reference. This is certainly the way Putnam sees the situation, and in the next section we will also cursorily consider his metaphysical alternative to realism. However this will also be found unattractive.

The permutation arguments are taken to be valid on all sides of this debate; and in Chapter 3 I presented my argument that they are also sound. In particular I have defended the most commonly challenged premise, that nothing beyond experience is available to fix reference. I have defended it by arguing that to postulate anything further would be a piece of highly creative metaphysics, out of keeping with naturalistic philosophy. Thus I have argued that, if you are a naturalistic philosopher, you should agree with Quine and Putnam that nothing is available to fix the reference of a word after you have fixed the truth-conditions of its contexts of use. On the basis of Chapter 3 and the present section of Chapter 4, I will take myself to have established this much. The remaining two sections will therefore assume the soundness of the permutation arguments and be devoted to seeing where this leaves us, and to trying to state what I believe to be the most plausible view of our actual situation regarding determinacy reference. The statement of my own positive view will come in Section 3.

2. *Unattractive Alternatives*

If we accept that the Quine-Putnam arguments are sound, then the initially obvious way to see this result is as casting doubt on the determinacy of reference. After all, the arguments focus on showing how reference is underdetermined by the available constraints. However this option has not been explicitly taken by any of the major players; Quine comes closest, but his attitude is oblique – by which I mean that, if you asked him to put his name to the statement, "I believe that reference is indeterminate", he would probably refuse. I will discuss Quine's attitude in the next section, where I will interpret and develop it. In the present section I want to explain why the *prima facie* obvious conclusion of the Quine-Putnam arguments, that reference is not determinate, is so widely considered to be unsatisfactory. Then I will briefly consider the other major alternative that has been proposed – Putnam's metaphysical adjustments. These too I will find unsatisfactory.

Lewis and Putnam assume that reference must be determinate. What they both seem to take to justify determinacy of reference is the apparent paradox of its denial. Such considerations seem to inform Putnam's view of his arguments about indeterminacy of reference as having the form of a *reductio ad absurdum* for the overarching position within which they operate: for indeterminacy of reference is taken to be an absurd result. I also take it that Lewis's title *Putnam's Paradox* is similarly motivated; hence also his claim that reading Putnam is like meeting a man who offers us an argument that no people exist, including that man himself.[26]

Here is the problem they seem to be reacting to. If I claim that my words do not determinately refer, then it appears that in order for me to be making the claim I take myself to be making, I have to assume that at least some of my words determinately refer. This is contradictory.[27]

---

[26] I do not mean to claim that Lewis and Putnam are explicitly concerned with the problem of someone who asserts that reference is not determinate. I merely suggest that they think there is a problem there, and that it is at least one of the reasons that they think rejecting determinacy of reference is not a viable option. They definitely regard the denial of determinacy as having absurd consequences; and they allude to this one.

[27] Perhaps "self-defeating" would be a more accurate term than "contradictory". For the view itself contains no contradiction: that only arises when the view is asserted. But I will

For this reason, taking the Quine-Putnam arguments to show that reference is indeterminate would seem to require a very good lawyer. Yet despite its prevalence, as an argument for determinacy of reference, this reasoning arouses my suspicions. I believe that the contradictoriness of denying determinacy of reference does not show that the permutation results are unsound. Nor does admitting the truth of these results leave us with a hopeless paradox. I will elaborate and argue for my view in the next section.

I have already mentioned that Putnam's own solution is to alter his metaphysical position, by rejecting metaphysical realism. It is worth briefly elaborating how such a move is supposed to work.

What licenses the various technical and not-so-technical devices employed in Putnam's arguments? As I laid the arguments out in Chapter 2, one part of the licence is the assumption that objects are, in some important way, independent of the terms that refer to them. One common intuitive motivation for such an independence is the fact that what word we choose to represent a given thing seems, to a large extent, to be logically arbitrary. Clearly certain factors narrow the choice down – the shape of our vocal chords, the speed at which we can process auditory information, and so on. But within these boundaries, there is considerable room for manoeuvre, as the plurality of natural languages and dialects shows. Another possible motivation for this independence-assumption is the naïve realist view that the world is out there, waiting to be represented correctly or otherwise by what we say about it. The idea that what we say might be right or wrong about the world presupposes, *prima facie* at any rate, a kind of independence between what is said and the way things are.

Independence of this sort is necessary for Putnam's permutation arguments to work because these arguments involve jiggling the assignment of objects to terms in ways that are intuitively surprising. Such jiggling might be blocked if that independence failed to hold. The externalist proposes that, to bridge the troublesome gap between our terms and independent, external objects, there might be an independent, external

---

sometimes loosely refer to the view as contradictory on the basis that it is contradictory whenever it is asserted or held. I exhibit this contradiction in the next section.

relation. Instead, Putnam's own solution is to close the gap by challenging the status of those objects as external.

Putnam's idea is that if objects themselves are somehow constitutively dependent on us and our experience of them, then the fact that reference is only determined up to the extent that it is determined by our experience does not matter. For the objects of reference – things – are similarly only determined up to the extent that they are determined in our experience. The meaning of words is determined by experience; if the world we sought to represent with those words were something above and beyond the world we experience, then our attempts at representation would fail (due to permutation considerations); but if instead the world itself is only determinate to the extent that it is determinate in our experience, then it is determinate to exactly the same extent as our words our. In this way, the gap is closed.

I will not endorse this line. First, I think it is very hard to give sense to the notion of the world being "internal", as may have been guessed by my efforts in the previous paragraph. The internal/external distinction is not an easy one to draw. In particular, drawing the internal/external distinction will be hard if you deny, as Putnam does, that it is intelligible to talk about the external. For then what are we to say the distinction is between? Yet Putnam does need that distinction, at least on the face of it; for otherwise it is not clear how to understand his contrast between his own internal realism and "externalism" – which, in these circumstances, just means metaphysical realism (slightly confusingly).

Moreover I think that the denial of realism is a very delicate matter. I am inclined to think that the way Putnam does it is self-defeating. The problem is with the status of Putnam's own claims. Putnam is on fairly safe ground so long as he argues that metaphysical realism contains a contradiction; for he can exhibit contradiction internally, without "stepping outside" the theoretical framework of metaphysical realism. But as soon as he starts making positive metaphysical proposals, he risks being hoisted with his own petard. He claims that it is impossible to step outside of theory altogether, and speak of things as they really are; yet it is hard to see how he

can avoid doing that himself, if he is to propose an alternative to metaphysical realism. Moreover the notion of stepping outside all theory – however that is to be understood – is one that he regards as close to the heart of metaphysical realism. It appears that he needs to assume the view he is denying in order to assert the truth of his alternative.

Problems like this face many forms of anti-realism. There are moves that can be made in defence, but this is not the place to investigate those moves. I have simply tried to do enough to indicate why Putnam's solution to his problem might itself be thought problematic. That, in turn, I hope is enough to motivate searching for another way of resolving the problem. In the final section, I will propose what I think is the correct solution.

3. Saying What Can't Be Said

The problem with what I want to conclude is that it is very hard to assert without falling into contradiction or triviality. I think that the permutation arguments, whose soundness I accept, show something important about reference. I even want to claim that they show that, in a particular sense, reference is not determinate. But I am also sensitive to the fact that, put thus baldly, I risk contradicting (or at least defeating) myself. For to assert that reference is not determinate is contradictory. My strategy in this concluding section will be to start by bringing out this contradiction. Determinacy of reference is a trivial truth, in that it follows from the denial of its denial. Then I will try to indicate the sense in which I deny determinacy of reference, without contradiction. I will indicate how I think this denial is compatible with the trivial truth of the bald assertion of determinacy of reference; and I will do so by emphasising the triviality of the latter.

It is a common inferential method to prove the truth of some statement "q" by showing how adopting "¬q", plus in most cases some auxiliary premises, results in a contradictory claim such as "(p & ¬p)". How is the *reductio* of the assertion of indeterminacy of reference supposed to work? Let us formulate:

(DET) Reference is determinate.

I have already briefly argued for the claim that asserting indeterminacy of reference leads to contradiction; here I make the argument more explicit. If (DET) is false, then what objects our words refer to is not determinate. Since "reference" is a word, it follows that there is no determinate fact concerning to which object(s), property or relation it refers. The same goes for the words "determinate" and "is". So not-(DET) cannot be regarded as expressing the indeterminacy of reference. For it cannot be regarded as mentioning either determinacy or reference.

Now we insert a premise to the effect that the person denying (DET) takes herself to be referring to reference and to determinacy, and to be denying that reference is determinate. My justification for this premise is simply that if we do not accept the premise – if we believe that she does *not*

take herself to be talking about determinacy and reference – then the entire dialogue breaks down. It is hard to explain why we are so exercised to dispute with her, and so interested in what she says, if we admit that we don't know what she's saying or trying to say.[28]

Now we have a contradiction. For not-(DET) denies that reference is determinate. But in order for not-(DET) to be asserted in a way that merits anyone's attention, it must be assumed that the asserter believes that the words she is currently using to make the denial have determinate reference. In other words, she must take (DET) to be true of here assertion of not-(DET). Yet it is a consequence of not-(DET) that it is not.[29]

So asserting not-(DET) leads to contradiction. Therefore, we should conclude that the negation of not-(DET) is true: it is not the case that it is not the case that reference is determinate. We have shown a claim of the form "¬(¬p)". This, in most contexts, suffices for showing that p. Thus we have shown that (DET) is true: reference is determinate, in at least one case – the case where we deny (DET).

So how can I go on to deny the determinacy of reference? And moreover, why should I do so? The answer to the latter question is that I accept the Quine-Putnam arguments: and I believe that they show that, in some important sense, reference cannot be determinate. I will now try to elaborate that sense, and distinguish it from the sense in which it can be shown, by the foregoing *reductio*, that reference is determinate.

The trick is to use analogy. This is Quine's approach. Quine notes that we can't distinguish different models of a theory from within the vocabulary of that theory. And if our "background theory" – our overarching, semi-explicit theoretical framework – is suitably similar to the various theories we consider when we make such assertions as the foregoing, then we can see how our vocabulary might ultimately be in the same position as the vocabularies of those less inclusive theories. One way

[28] I ignore the possibility that we are talking to her because we take her to be talking about something else that is much more interesting.

[29] In the foregoing argument I suppress the assumption that not-(DET) is equivalent to "Reference is indeterminate." I think it implausible that there is some third possibility,

to see the argument of the present work is as trying to show that there is nothing about our overarching background thesis to set it apart in any relevant respect. I have done this by arguing that experience is the ultimate determinant of reference.[30] Nothing further is available, I have argued, to prevent the kind of reassignments that are commonplace in model-theory.

Quine puts it like this:

> It is thus meaningless within the theory to say which of the possible models of our theory form is our real or intended model. Yet even here we can make sense of there being many models. For we might be able to show that for each of the models, however unspecifiable, there is bound to be another which is a permutation or perhaps a diminution of the first.
>
> (Quine, 1969: 54)

The sense in which reference is not determinate is thus accessible only by analogy. For we have already seen that denying determinacy of reference directly leads to contradiction. Let me now distinguish more clearly between the sense in which reference must be determinate and the sense in which it can't be.

The viewpoint from which reference is not determinate is the realist, God's-eye, external viewpoint. As previously discussed, the permutation considerations assume this viewpoint for their starting point. The contradiction that arose from trying to deny determinacy of reference was due to the fact that we were trying to deny it of our own actual situation. The permutation arguments, on the other hand, are indifferent to situation. They merely show that any attempt to refer, that possesses certain features and meets certain conditions, will fail. If our actual situation shares those features and meets those conditions, the permutation arguments tell us that we also fail to determinately refer. And I have argued that our actual situation does meet those conditions.

---

besides determinacy and indeterminacy of reference. I take it that these alternatives exhaust the logical territory.

[30] For fluency I have slipped to speaking as if experience constrains reference directly. It will be remembered from my previous discussion that I think *use* plays a mediating role. Reference is fixed to the extent that contexts of use are fixed; and contexts of use are fixed in experience.

But it is important that this failure to refer is understood by analogy only. The permutation results tell us that it is equally impossible to assert as to deny determinacy of reference in the absolute sense. For if it is impossible to distinguish models from within the vocabulary of our theory, then it is equally impossible to assert that there is exactly one intended model as it is to deny that there is exactly one. Both attempts assume something impossible: the ability to distinguish between models from within the vocabulary of the theory in question.

How is this to be reconciled with the truth, recently proved, of the assertion, "Reference is determinate"?

The answer is that in this sense, where no analogy is being employed, reference is indeed determinate. "Rabbit" refers to rabbits, and so on. The sense in which reference is *not* determinate is accessible only by analogy; and then, the previously exhibited *reductio* won't work, because it relies on direct assertion. The point of employing the device of analogy is to avoid direct assertion, thus avoiding the *reductio*.

From the perspective of this analogy, we can see that reference is determined to the extent that it is determined in and by experience. This is another way of putting what I have argued, that experiential constraints are the only constraints on determinacy of reference. When we truly assert (and prove) "Reference is determinate", we are operating within the framework of our experiential constraints. When we employ the analogical device that Quine suggests and I endorse, we are seeking to escape that framework.

So the position I finish with is one that is very difficult to clearly express. The bald denial of determinacy of reference is problematic, indeed false. Reference, the reference with which we are familiar, is determinate. The force of the considerations presented in the course of this thesis is rather to limit the ambition of anyone who asserts this truth. For we can show, by analogy, that were we able to view our actual situation from without, we would see that it is one where objects are assigned to terms with just the determinacy required for our experiential constraints on that assignment to be satisfied: and that is not very far.

I am fully aware that my position as stated falls far short of a thoroughly formulated, fully expressed philosophical view. But I introduced this work by promising not to introduce a new view. To produce one would require another work of similar or greater length. There are two substantial issues within this debate on reference. One is whether the Quine-Putnam arguments are sound. The other is how we are to react to them if they are sound. I regard debate on the second question as futile unless the first is at least properly understood, and preferably settled. And the question of whether the Quine-Putnam arguments are sound is widely regarded to have been settled in what I feel is the wrong way. So I have focussed on these arguments; and I believe that I have defended them.

Bibliography

Andersen, D.L. (1993): *What is the Model-Theoretic Argument?* in *The Journal of Philosophy* 90, 311-322.

Benacerraf, P. (1965): *What Numbers Could Not Be*, in *Philosophical Review* 74: pp. 47-73.

Blackburn, Thomas (1988): *The Elusiveness of Reference* in French, Uehling and Wettstein (eds.), *Midwest Studies in Philosophy Volume XII: Realism and Anti-Realism.*

Boghossian, P. A. (1991): *Naturalizing Content*, in G. Rey and B. Loewer (eds.), *Meaning in Mind: Fodor and his Critics*: pp. 65-86.

Carnap (1956): *Empiricism, Semantics and Ontology*, in *Meaning and Necessity.*

Davidson, Donald (1969): *True to the Facts*, in *The Journal of Philosophy* 66: pp. 748-764.

Devitt (1983): *Realism and the Renegade Putnam*, in *Nous* 17: pp. 291-301.

Dretske, Fred (1980): *The Intentionality of Cognitive States*, in French, Uehling and Wettstein (eds.) *Midwest Studies in Philosophy V*: pp. 281-294.

Field, Hartry (1972): *Tarski's Theory of Truth*, in *The Journal of Philosophy* 69: pp. 347-375.

Fodor, Jerry A. (1987): *Psychosemantics: The Problem of Meaning in the Philosophy of Mind.*

Hale, Bob and Wright, Crispin (1997): *Putnam's Model-Theoretic Argument Against Metaphysical Realism*, in Bob Hale and Crispin Wright (eds.), *A Companion to the Philosophy of Language*: pp. 427-457.

Heller, Mark (1988): *Putnam, Reference and Realism* in French, Uehling and Wettstein (eds.), *Midwest Studies in Philosophy Volume XII: Realism and Anti-Realism*: pp. 113-127.

LePore, E. and Loewer, B. (1988): *A Putnam's Progress*, in French, Uehling and Wettstein (eds.), *Midwest Studies in Philosophy Volume XII: Realism and Anti-Realism*: pp. 459-473.

Lewis, David (1970): *How to Define Theoretical Terms*, in *The Journal of Philosophy* 67: pp. 427-446.

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

Putnam, Hilary (1978): *Meaning and the Moral Sciences*. Includes *Realism and Reason*: pp. 123-140.

Putnam, Hilary (1980): *Models and Reality*, in *Journal of Symbolic Logic* 45: pp. 464-482.

Putnam, Hilary (1981): *Reason, Truth and History*.

Quine, W.V.O. (1960): *Word and Object*.

Quine, W.V.O. (1966): *Truth by Convention*, in *The Ways of Paradox*.

Quine, W.V.O. (1969): *Ontological Relativity*, in *Ontological Relativity and Other Essays*: pp. 26-68.

Winnie, John A. (1967): *The Implicit Definition of Theoretical Terms*, in *The British Journal for the Philosophy of Science*: pp. 223-229.

Zalabardo, José L. (1998): *Putting Reference Beyond Belief*, in *Philosophical Studies* 91: pp. 221-257.

Zalabardo, José L. (2000): *Introduction to the Theory of Logic*.