ONE WORLD AND THE MANY SCIENCES:

A DEFENCE OF PHYSICALISM VOL 2.

(Submitted for the degree of D.Phil. in Philosophy, Trinity 1990.)

Andrew Melnyk

Corpus Christi College, Oxford



Part 2

BODLEIAN LIBRARY OXFORD
DEPOSITED
26 IV 91

CHAPTER FIVE

The Explanatory Test Defended: I

The key premiss, (P1), of the positive argument for eliminative type physicalism outlined in the previous chapter states that we are justified in taking a property to exist if, but only if, it is mentioned in the best explanation of the course of our perceptual experience. It expresses, as I noted, one possible version of the explanatory test for reality; and I shall in this chapter and the next be defending it. My strategy, viewed at a high level of abstraction, is to connect causal explanation with justification, rather than with existence. This will have the consequence that, although causal explanation and existence have, logically, nothing to do with one another, so that it is possible for a property to exist, even though it does no necessary causal explanatory work, it is not possible that one be justified in taking such a property to exist.

Since (P1) is biconditional in form, it is logically equivalent to the conjunction of two conditionals, and I shall defend each component conditional in turn. In this chapter,

then, I shall be arguing for the claim that we are justified in taking a property to exist if it is mentioned in the best explanation of the course of our perceptual experience. My defence of this claim will consist in an argument to the effect that it is sufficient to meet the challenge laid down by the sceptic about the external world.

1. Scepticism About The External World

In speaking of the challenge laid down by the sceptic about the external world, I speak advisedly. For it has been traditional to regard the sceptic as offering an argument whose conclusion states, shockingly, that we are not justified in believing in any external objects at all. I shall, however, be breaking with this tradition (following Pollock 1987, pp.5-7). If a reconstructed sceptical argument is valid, then the proper response is to take it as a reductio against one of the premisses. This response is the proper one, because we are more sure of the falsity of the supposed conclusion -- that we are not justified in believing in external objects -- than we are of the truth of the premisses. In the light of this, I think it is best to view scepticism about the external world, not as a serious attempt to persuade us that our belief in an external world is unjustified, but as a challenge of a certain kind. It arises on the assumption, which I share, that if we

are justified in believing in an external world, than it is on the basis, somehow, of our perceptual experience. The challenge is to explain exactly how this can be so, given the traditional sceptical considerations.

These traditional considerations turn upon the observation that, as is evident a priori and as our experience of dreaming and hallucination confirms, it is logically possible for one to enjoy a perceptual experience, when merely dreaming or hallucinating, of a sort introspectively indistinguishable from the perceptual experience enjoyed when one is genuinely perceiving. So, for example, the kind of experience I enjoy when I hallucinate a white rabbit may well be introspectively indistinguishable from the kind of experience I enjoy when I really do see a white rabbit. And it certainly seems to follow from this that it is at least logically possible that all my perceptual experience be introspectively indistinguishable from the way it is now, but that none, or almost none, of the external objects I ordinarily suppose myself to be perceiving exist.

Suppose, then, that it seems to me visually as if there were a white rabbit in front of me. How is this perceptual experience to licence my belief that there really is a white rabbit in front of me? Not by deduction from the occurrence of

a perceptual experience of this type; for, as the possibility of hallucination shows, it is possible for an experience of this type to occur in the absence of any appropriate white rabbit. Nor by <u>induction</u> either; for if the idea is that one concludes on the basis of an observed correlation between rabbit-experiences on the one hand and real rabbits on the other, then the question has simply been begged, since it has been assumed that you already have knowledge of external white rabbits; and if the idea is that veridical perceptual experience somehow wears its character on its sleeve, a small red triangle in one corner of the visual field perhaps, then the correct reply is that this is simply not true -- veridical and hallucinatory perceptual experience can be introspectively indistinguishable. But if not by either deduction or induction, then how? The sceptic's challenge is to answer this question.

Before I set out what I take to be the proper way to meet this challenge, I want briefly to consider a couple of unsuccessful responses to it. Both of them allege that the challenge can be met by rejecting some part of the sceptic's views about the nature of perception. The first such response charges the sceptic with assuming the truth of an indirect realist, or representative, theory of perception according to

which our perceptual contact with the world is mediated by a veil of perception. This theory, it is alleged, is what gives rise to the sceptical challenge, and so should be replaced with some direct realist account according to which the content of a (veridical) perceptual experience is somehow literally constituted by the objects perceived, and on which the sceptical challenge does not arise.

But this response is doubly objectionable. First, indirect realism is to be preferred to the direct realism which the response recommends (see Robinson 1985, pp.170-177). Secondly, even if this were not the case, the adoption of direct realism would not prevent the emergence of the sceptic's challenge. The reason is as follows. The direct realist's account of the content of perceptual experience is obviously going to work only in the case of veridical experience; evidently the content of a hallucinatory experience cannot be constituted by the object perceived, since in such a case there is none. So even if in the case of veridical perceptual experience the objects apparently perceived cannot fail to exist, so that scepticism is impossible, still the introspective indistinguishability of veridical and hallucinatory perceptual experience entails that we cannot tell, on any occasion of our enjoying a perceptual experience, whether it is a genuine instance of perception, in

which case we can lay aside our sceptical doubts, or a hallucination, in which case we cannot. Perception may be as directly realist as you like; still we cannot tell whether what we have is a case of perception. (See Jackson 1977, pp.147-151.)

The second unsuccessful response to the challenge of scepticism is also a protest to the effect that a false view about perception is assumed. This time the claim is that the sceptic assumes that perceptual experience involves something, perhaps its subjective character, over and above a set of beliefs or dispositions to believe (where the beliefs do not somehow involve the subjective character as part of their content); if we abandon this assumption, then we will think of a (say) visual experience as a set of beliefs or dispositions to believe acquired through the eyes.

To this response there are again two replies. First, if the sceptic did make this assumption, then he would be right to do so: reductionist accounts of perceptual experience which omit its subjective character are in my view false (see Goldman 1988, especially pp.88-113 and Jackson 1977, pp.37-48). But, secondly, the sceptic need not make this assumption: he may adopt a reductionist account, and the sceptical challenge can still be made. For the source of this challenge is the

introspective indistinguishability of veridical and hallucinatory perceptual experience, and these two types of perceptual experience remain introspectively indistinguishable even if, as the reductionist in effect proposes, you identify perceptual experience with sets of beliefs or dispositions to believe.

So much, then, for these unsuccessful attempts to meet the sceptical challenge of saying how it is that perceptual experience licences belief in an external world. Let me now introduce my own favoured response. (Defences of it can be found in Maxwell 1968, Ayer 1973, Mackie 1976, pp.62-67 and, at great length, Goldman 1988, especially pp.203-215.)

We need to begin by noting that the sceptic's enumeration of possible ways of moving from perceptual experience to justified belief in external objects is not exhaustive. Clearly he is right to rule out deduction; and if we think of ourselves as enjoying some particular perceptual experience on some particular occasion, then what he says about induction is surely correct also: there is no distinctive feature of a veridical perceptual experience which we can detect and then cite as evidence that the experience is veridical. But suppose that, rather than concentrating upon cases where we enjoy some particular perceptual experience on some particular occasion,

we concentrate instead upon the succession of many perceptual experiences over time, i.e. upon the course of our perceptual experience. If we do so, then the possibility is opened up of our discerning something in that succession of many experiences which is present in no single experience, but which would license some kind of non-deductive inference to the existence of an external world. My favoured response to the sceptical challenge exploits precisely this possibility.

Consider, now , the following minimal realist hypothesis:

> (MRH) There exist some properties which (a) are neither identical with, nor constituted by, one or more mental properties, (b) go on existing even when no one perceives them, and (c) are causally efficacious.

This expresses a recognisably realist hypothesis (see Devitt 1984). But its content is, in an obvious way, indefinite; for the properties which it asserts to exist might be the familiar properties of everyday talk of material objects, or the unfamiliar properties of particle physics, or other kinds of properties. It is clear, however, that we can produce any number of realist hypotheses richer in content than (MRH) simply by specifying in greater detail the characteristics, and in particular the causal powers, of the properties said to exist. We may call these richer hypotheses specific realist

hypotheses. Evidently, different specific realist hypotheses will postulate different sets of properties.

I mentioned a moment ago that the sceptic had overlooked the possibility that there might be some fact about the succession of many perceptual experiences which would licence a non-deductive inference to the existence of an external world. Is there such a fact? There is, and it is the fact that our perceptual experiences are conspicuously ordered and manifest notable regularities. This is made clear by a moment's reflection on how, again and again, some one kind of perceptual experience is associated, either synchronically or diachronically, with some other kind of perceptual experience. These associations are rarely exceptionless, but they still constitute very striking regularities. Since I take this fact to be obvious, I shall not try to argue further for it, though I shall in Section Three examine some alleged difficulties with it. Now, this order and regularity cries out for explanation. And explanation can be provided. For it is plausible to suppose that various specific realist hypotheses constitute candidate explanations of the order and regularity of the course of our perceptual experience; that is, if one of these hypotheses were true, then the properties alleged by it to exist, given the causal powers ascribed to them, would account

for the course of our perceptual experience. These various specific realist hypotheses, however, will not all constitute equally good candidate explanations, and we may reasonably hope that we should be able to pick out one of them as the best such candidate explanation. But then, if inference to the best (candidate) explanation is a legitimate form of non-deductive inference, the fact that some specific realist hypothesis constitutes the best candidate explanation of the course of our perceptual experience will give us reason to believe that hypothesis to be true. Since this hypothesis will state precisely that properties of a certain kind exist, we will have reason to believe in those properties. Of course, any specific realist hypothesis will entail (MRH), which can naturally be viewed as an expression of the claim that there is an external world. So we can see how an application of inference to the best explanation makes it rational to believe in an external world.

But which properties exactly are exemplified in the external world? The answer is implicit in the preceding discussion. We are justified in believing in an external world because this belief -- (MRH) -- is trivially entailed by whatever specific realist hypothesis we accept on the grounds that it constitutes the best explanation of the course of our perceptual

experience. So we are justified in taking to be exemplified in the external world those properties mentioned in that specific realist hypothesis. In other words, we are justified in taking to exist those properties mentioned in the best explanation of the course of our perceptual experience. And this claim is, of course, the relevant component conditional of the key premiss, (P1).

So belief in an external world is justified by means of an inference to the best explanation from the data constituted by the course of our perceptual experience. Both elements of this response to the challenge of scepticism require elaboration and defence. In Section Three I shall consider possible difficulties with the data for the proposed inference. In Section Two, however, which follows, I shall discuss the legitimacy of the pattern of inference involved, that of inference to the best explanation. The objections I shall discuss do not exhaust all those that I can think of; but they are in my view the most important. For responses to other objections I can only refer to the literature already cited.

2. Inference To The Best Explanation

The response to the challenge of scepticism which I have been defending relies upon the legitimacy of that pattern of inference often called inference to the best explanation, but also sometimes abductive inference. It is a familiar enough pattern of inference. Suppose that I hear scratching in the wall and the patter of little feet at midnight, and that my cheese disappears. I will doubtless take these facts as rendering it probably true that there is a mouse in the wainscoting. And this seems to be because, if there had been such a mouse, then, given certain properties of mice, its presence would have accounted for the observed phenomena; and also because, although other explanations are conceivable, involving, for instance, the telekinetic powers of aliens, the explanation involving a mouse is the best. (I take the example from van Fraassen 1980, pp.19-20.)

The general pattern exemplified by this kind of example, and by the many other examples which can easily be devised, seems to be something like this. We begin with a set of phenomena, which are the data for the inference. We then consider all the available candidate explanations for the phenomena -- all the possible circumstances such that, if they did obtain, they would explain the phenomena. Next, we apply certain criteria

for the goodness of candidate explanations to the candidates. Finally, having selected one candidate as best satisfying those criteria, i.e. as the best candidate explanation, we infer to its probable truth, and hence to the ontology which it presupposes. The main unclarity here concerns the criteria by which it is assumed that we can make comparative judgements about the relative goodness of rival candidate explanations: we want to know what they are. Unfortunately, although it is easy to get wide agreement on what to say in such cases as the mouse example, it is notoriously difficult to specify in any sort of detail what the relevant criteria are. But they are likely to include the following: capacity to explain in a unifying way apparently unrelated phenomena, parsimony or simplicity, conservatism, and testability. (See further Quine and Ullian 1970, Chs.5 and 7, Thagard 1978, Glymour 1984, and Lycan 1988, pp.129-130.)

Is inference to the best explanation a legitimate form of non-deductive inference? This is the question which really concerns me, since if it is not, then the response to the sceptic's challenge that I have been defending must be inadequate. In what follows, however, I shall not attempt any sort of fundamental defence of the legitimacy of inference to the best explanation. Such fundamental defence could only be

accomplished after extended discussion of the methodology of such a defence, and for more than one reason I cannot undertake that discussion here (though see, perhaps, Boyd 1985 and Goldman 1988, Ch.13). What I shall do, however, is try to undermine two kinds of scepticism about inference to the best explanation, a global kind and a local kind.

The global sceptic holds that application of inference to the best explanation is <u>never</u> legitimate: the fact that certain possible circumstances constitute the best candidate explanation for a range of phenomena never justifies us in taking those circumstances to be actual. The first thing to be said about this global scepticism is that there is a high price to be paid for its acceptance. For inference to the best explanation is very widely used, both in everyday life (for instance by detectives) and in science (for instance in astronomy). Consequently, it is unclear how much of what we ordinarily think of as knowledge can survive global scepticism about it (cp. Mackie 1982, pp.4-5). This should prompt us to scrutinise the possible reasons for adopting such scepticism. Why believe that inference to the best explanation is never legitimate?

In order to understand such a global doubt, we may begin by considering this question: what on earth has explanation to do

with truth? To this query it is tempting to reply: "Everything!". For it is obvious that, as Nancy Cartwright has emphasised, "causal explanations have truth built into them" (see Cartwright 1983, p.91). The point here, of course, is that a possible set of circumstances cannot be the causal explanation of some phenomenon unless it is true that the circumstances are actual. If there is no mouse in the wainscoting, then the presence of a mouse in the wainscoting cannot be the explanation for the nocturnal noises and missing cheese. Explanations have to be true (even though candidate explanations, of course, do not).

But this reply, although it is correct, does not adequately answer to the global sceptic's worry about the connection, or lack of it, between explanation and truth. For the global sceptic can happily allow the constitutive point that an explanation must be true in order to be one. What he will want to insist upon, however, is the epistemic point that whether a candidate explanation is true is something to which its being a candidate explanation is entirely irrelevant. His question is really this: what on earth does explanatory power have to do with truth, where explanatory power is something which a false hypothesis may have? The force of this question is best brought out by attending to the problem of the

underdetermination of theory by data, perhaps better described as the problem of empirically equivalent theories.

Let two theories be empirically equivalent just in case they entail exactly the same observational consequences. (The distinction between what is observational and what is theoretical is to be made relative to the case in question; intuitively, the observational pertains to the data for some possible inference, the theoretical to its conclusion.) Then, for any theory, there will always be an empirically equivalent rival. This is so because there is a simple method for constructing empirically equivalent theories. To construct a theory, T2, which is empirically equivalent to a theory, T1, all you have to do is to determine the observational consequences of T1 and then list them; the list just is T2. T2 trivially entails exactly the same observational consequences as T1, and hence is, by the above definition, empirically equivalent to it. And since T1 and T2 are empirically equivalent, no possible observation can favour one theory over the other. The significance of this is as follows. If you believe that all that is relevant to the assessment of a theory's truth is its empirical adequacy, i.e. its capacity to entail true observational consequences, then, since no possible observation could confirm T1 without also confirming T2, there

can be no reason to prefer either of T1 or T2 to the other from the point of view of truth. This last qualification is crucial, however, since there may, of course, be pragmatic reasons for preferring T1 to T2; for instance, T1 may be computationally simpler, and hence easier to use in generating useful predictions. But these pragmatic reasons are not reasons for supposing it to be more likely that the theory is true. So although, as seems intuitively obvious, T1 has greater explanatory power than T2 -- T2, after all, simply reiterates the data -- this cannot be permitted to count in favour of T1's truth.

It is now possible to state what the global sceptic about inference to the best explanation is getting at in demanding to know what explanation, or rather explanatory power, has to do with truth (see van Fraassen 1980). His claim is that all that is relevant to assessing the truth of a theory is its empirical adequacy, and hence that any explanatory power the theory may possess, even its constituting the best candidate explanation for some phenomenon, can at best be viewed as a pragmatic, and thus truth-irrelevant, feature of the theory. This, in its turn, has a further consequence. For if it is true, as it seems to be, that for every theory we can construct an empirically equivalent theory, and a theory, moreover, which

only mentions observational consequences, then it follows that we can never be justified in holding true any theory which postulates phenomena which cannot be said to be observational. In short, we cannot get beyond the data. In the case of my favoured response to scepticism about the external world, we will never be justified in accepting as true any hypothesis, such as a realist one, which talks about more than the course of our perceptual experience, even if the hypothesis has enormous explanatory power.

So the challenge laid down by the global sceptic about inference to the best explanation is to show that, and why, a theory which has explanatory power may be thought of as more likely to be true than an empirically equivalent theory which does not have explanatory power, or which has less of it. Regrettably, I am unable to show <u>why</u> this is so; to accomplish that would be precisely to undertake the fundamental justification of inference to the best explanation in which I earlier declined to engage. But I can show <u>that</u> it is so -- or at any rate that we ordinarily take it to be so. My strategy is to consider an example where it seems obvious that explanatory power counts in favour of truth.

My example concerns ancient history -- the attempt to discover what happened in the distant past. We will typically

be faced, as ancient historians, with a varied array of evidence. Some of it will be textual, which is to say that we have before us now (if we are very lucky) manuscripts which claim that such-and-such events took place; and some of it will be archeological -- inscribed blocks of stone, ruined buildings and so on. Now let us imagine that on the basis of an array of evidence of this sort we conclude that there existed a man called Pericles who did certain things. What is to stop us constructing a theory, empirically equivalent to the Pericles theory, simply by listing reports of our current observations of the array of evidence? This new theory will not, of course, be empirically equivalent in the sense that it entails <u>exactly</u> the same observational consequences as does the Pericles theory; but we can easily ensure that that it entails, trivially, every observational consequence which the Pericles theory entails <u>and</u> which we are now in a position to confirm. So it will not matter that the Pericles theory entails observational consequences which only a contemporary of his could have confirmed, and our two theories can certainly be such that they both entail exactly the same observational consequences confirmed or confirmable by us.

Now, if it is correct that all that matters for the assessment of the truth of a theory is its empirical adequacy,

then the ancient historian will have to admit that there is no reason to believe the Pericles theory rather than the new theory (which we can imagine to contain an explicit denial of Pericles' existence). And yet this is absurd, since the historian will surely be justified in believing the Pericles theory to be true. But why will he be thus justified? Not on empirical grounds, to be sure; nor on grounds of simplicity, since the Pericles theory is on any reckoning much less simple than its rival. The obvious suggestion is that his belief is justified by an appeal to the explanatory power of the Pericles theory. The new, rival theory, since it simply reiterates the evidence, appears to have no explanatory power whatever; whereas the Pericles theory, by supposing that there existed a person whose actions had distant effects which we can still observe today, constitutes a candidate explanation for the evidence. So this example seems to be a clear case where a hypothesis is believed to be literally true -- talk of the instrumental usefulness of such a theory as the Pericles theory is ridiculous -- but where the ground for this belief is, because it can only be, the explanatory power of the hypothesis. If it is such a case, then the global sceptic about inference to the best explanation who denies that explanatory power ever counts in favour of truth cannot be right. (For other difficulties in the position of the global

sceptic, see Boyd 1984 and Churchland 1985b, pp.37-40.)

But perhaps the global sceptic is not finished yet; perhaps the problem of the underdetermination of theory by data still has trouble in store for inference to the best explanation. Following William Newton-Smith, let us call two theories evidentially equivalent iff they are empirically equivalent, in the sense that they entail exactly the same observational consequences, and they enjoy the supraempirical virtues of explanatory power and so on to an exactly equal extent (see Newton-Smith 1983, p.471). The supposition that there might actually be evidentially equivalent theories, which were not mere notational variants, certainly seems to be possible, at least in the sense that the supposition involves no obvious inconsistency (see Newton-Smith 1978 and Wright 1985 for suggested examples). Does this possibility threaten inference to the best explanation? Let us suppose that it is actual: we have two theories, T1 and T2, which are evidentially equivalent. Is there a problem for inference to the best explanation here? As things stand, there is not; for there may be a third theory, T3, which is evidentially superior to both T1 and T2, in which case we should infer to the truth of T3. But let us imagine that there is no such theory as T3, or that for every theory there is an evidentially equivalent rival, so

that the same question would arise for T3 and its evidential equivalent, T4. Is there a problem? Not as far as I can see. For nothing in my exposition of inference to the best explanation ruled out the possibility of their being more than one best explanation for some set of data. Consequently, in the face of the evidentially equivalent theories, T1 and T2, it is open to us to infer to a <u>disjunctive</u> conclusion, [T1 V T2]. This is in effect what Newton-Smith calls the Ignorance Move, and it may be slightly distressing to notice that, as he points out, the adoption of this move involves acknowledging the existence of a fact, namely whether T1 or T2 is true, which we will never be able to establish, since, by hypothesis, neither empirical nor supraempirical considerations favour one of T1 or T2 over the other (see Newton-Smith 1978, p.88, and 1983, p.473). But to anyone without verificationist sympathies this distress will be small. The risk that every realist runs -- the risk that makes it fun to be a realist -- is that the mind-independent world will outrun, either entirely or in part, our capacity to learn about it.

It might be objected that inference to a disjunctive conclusion is all very well where we have only two evidentially equivalent theories, but progressively less plausible as the number of such theories rises until, when the number reaches

infinity, inference to a disjunctive conclusion is not plausible at all. Doubtless this objection is correct, but it forces us to consider the question how widespread evidential equivalence, among theories which are not mere notational variants of one another, is. It is not obvious, in fact, that there are even any pairs of evidentially equivalent theories, still less any infinite sets of them. If there are any mere pairs -- and the possible examples alluded to above are not uncontroversially convincing -- then they are rare. The situation would be different if we had some kind of algorithm for generating evidentially equivalent theories, of the sort I suggested above for generating merely empirically equivalent theories, but we do not. It is worth emphasising also just how difficult the construction of theories is: we should not expect to be able to generate evidentially equivalent rivals in a spare afternoon! So although there is no reason to rule out a priori the possibility of evidentially equivalent rivals, it is very far from clear that the possibility is ever actual, and it seems really most unlikely that there is a case where we have a damagingly large number of such rivals.

So much, then, for global scepticism about inference to the best explanation. Let me turn now to a discussion of a local version of such scepticism. Whereas the global sceptic held

that inference to the best explanation is never legitimate, his local cousin maintains that sometimes it is, though sometimes it is not. In principle, then, there could be any number of local scepticisms, varying in accordance with the precise manner in which they distinguished between the legitimate and the illegitimate applications of the inference. But as it happens I know of only one defender of local scepticism, and it is upon his defence that I shall concentrate.

The local sceptic is Bas van Fraassen, and his claim is that inference to the best explanation is legitimate if, but only if, the conclusion mentions observables, such as tables and elephants; so if it mentions unobservables, such as neutrinos and bosons, the inference is illegitimate (see van Fraassen 1980, pp.19-23). He makes this claim in the course of his critical scrutiny of arguments for scientific realism, and so it might be thought that it was irrelevant to my present concern of defending a certain response to the challenge of external world scepticism. But this is incorrect. The development of my positive argument for eliminative type physicalism is going to require a quite unrestricted use of inference to the best explanation. So van Fraassen's selective scepticism towards it must be overcome.

Van Fraassen's argumentative strategy is to present an

account of the basic form of inference to the best explanation alternative to the one which, following orthodoxy, I presented earlier. This alternative account entails that a proposed inference to the best explanation is only legitimate if its conclusion mentions observables. The account may be illustrated by reference to the mouse example. According to it, we do not infer to the conclusion that there is a mouse in the wainscoting directly. Rather, we infer in the first instance to the weaker claim that all the observable phenomena are as if there is a mouse in the wainscoting. Next, we add a crucial extra premiss stating that a mouse is an observable phenonmenon. And finally, we are able to deduce the conclusion that there is a mouse in the wainscoting. But of course we only reach this conclusion by asserting the extra premiss to the effect that a mouse is an observable phenomenon. And, generalising, the conclusions which we can reach by inference to the best explanation can only commit us to the existence of observables. We may conclude that things are as if unobservables are thus and so, but not that unobservables are thus and so.

To all of this I have two objections, of which the first, and less important, is as follows. There seems to be no positive reason for accepting van Fraassen's reconstruction of what is

going on in everyday cases of inference to the best explanation such as the mouse case. There might have been, for example, an effort to demonstrate its superior ability to handle puzzle cases; but there is not. There are, on the other hand, two reasons for rejecting it. The first is that it is more complicated than the orthodox account, and that counts against it. The second is that van Fraassen's added complication -- the requirement of an extra premiss about the observability of entities mentioned in the conclusion -- is, so far as I can see, quite alien even to intelligent users of inference to the best explanation. It will not do to reply to this point that people often reason in accordance with rules of inference which they do not know how to formulate explicitly. For usually, when presented with the full and explicit form of a pattern of reasoning, people will acknowledge that, if they had been arguing with maximal care and strictness, the full and explicit form is what they ought to have offered. But van Fraassen's expanded reconstruction of what is involved in the ordinary cases strikes no such responsive chord.

Here is the second, and crucial, objection. Van Fraassen's account of the restricted legitimacy of inference to the best explanation depends entirely upon the distinction between observables and unobservables, but this distinction seems to be

epistemically arbitrary. This objection is easily misunderstood, and its force unappreciated. The point is not that there is no distinction between observables and unobservable, since there is, and we can indicate its extension by citing examples, as I did when introducing van Fraassen's view. Nor is the point that the distinction is fuzzy, or that the difference between observables and unobservables is a matter of degree; for fuzzy distinctions, or differences of degree, can, like the difference between night and day, be of practical importance. The point, rather, is that it is entirely obscure what is supposed to be the principled epistemic distinction between observables and unobservables. There is a distinction between what is edible and what is inedible, about as clear as that between what is observable and what is unobservable, but only a lunatic would suggest that there was some special difficulty in getting to know about inedibles. How is the distinction between observables and unobservables better off, from this epistemic point of view?

Observability, it is important to notice, is not the same as actual observation, whose epistemic significance is manifest, since it is evidently one way -- there may be others -- of finding out about things. But here the mouse example may mislead. One obvious respect in which the inference to the

existence of a mouse differs from, for instance, an inference to the existence of a neutrino is that we have actually observed mice but never neutrinos. Thus there is a source of epistemic support for the claim that there is a mouse -- actual observation -- which has no analogue in the case of the neutrino. But this difference is not the same as van Fraassen's distinction between the observable and the unobservable. For precisely this same difference would exist between the mouse example and an inference to the existence of an entity observable, but never actually observed. (Imagine we never actually saw the mouse.) But an inference of this latter sort would, on van Fraassen's view, be perfectly acceptable.

The only epistemic advantage which observability, as opposed to actual observation, might possibly have would be this: in the case of an inference to the existence of an observable, there is a kind of epistemic support -- actual observation -- which the conclusion could enjoy, but which the corresponding conclusion affirming the existence of an unobservable could not. But this suggestion is hopeless. For, first, it is entirely mysterious why merely potential evidence for a conclusion should enable an inference not actually using that evidence to succeed, if otherwise it would not have done. And, secondly, the suggestion leaves open the possibility that there

is a type of epistemic support uniquely available to conclusions affirming the existence of unobservables, a type equal in strength to that deriving, in the case of observables, from observation, and thus capable of negating the disadvantage under which it was suggested that claims about unobservables labour. I am thinking, here, of what practising physicists, in their pretheoretical way, are inclined to distinguish from mere postulation and call <u>experimental detection</u>.

This last remark prompts a final point. "Observable" means "able to be observed". So a philosopher who, as van Fraassen does, lays great weight upon the distinction between the observable and the unobservable owes us an account of what is involved in observing something. Moreover, this account must have the upshot that there is something of great epistemic significance about the relation we stand in when we observe an observable, such as an elephant, which cannot be replicated in any relation in which we may stand towards an unobservable, such as an electron. What this account could be, I have no idea. For it certainly appears that observation of elephants enables us to learn about them because, in observing them, we make ourselves sensitive in a systematic way to their causal effects; in which case, it is hard to see why we cannot place ourselves in such a position that we can be sensitive to the

causal effects of electrons. The intervening causal chain will be longer, and we may not want to call it observation, but these truths are surely irrelevant (cf., perhaps, Goldman 1988, pp.230-232). I conclude, then, that van Fraassen's local and selective scepticism about inference to the best explanation, being based upon an epistemically arbitrary distinction between observables and unobservables, is itself arbitrary and hence unwarranted. (For further criticism, see Churchland 1985b.)

3. The Data For The Inference

It might be thought that the response to the sceptical challenge which I favour cannot work because of some problem with the alleged data for the inference, the data constituted by the order and regularity exhibited by the course of our perceptual experience. One such problem, or class of problems, would be held to arise from the fact that, if this order and regularity is to be described, then the only way in which this could be done is by the use of expressions which are in some sense parasitical upon the language of commonsense realism. For example, I might describe my enjoying of some perceptual experience by saying something like, "It seemed to me visually as if a zebra stood before me". But the term, "zebra", is itself already part of material object talk of the sort which it is the purpose of the proposed use of inference to the best

explanation to legitimate. And this might seem to suggest that the whole enterprise is afflicted by a fatal circularity.

To make this line of objection more specific, let us consider the following quotation from Strawson:

> In order for some beliefs or set of beliefs to be correctly described as a theory in respect of certain data, it must be possible to describe the data on the basis of which the theory is held in terms which do not presuppose the acceptance of the theory on the part of those for whom the data <u>are</u> data. (Strawson 1979, p.45)

In this passage Strawson states an eminently reasonable requirement on the description of data with respect to some proposed theory: the description of the data must not presuppose the <u>truth</u> of the theory. But this requirement is one which the inference to an external world from the data constituted by the course of our experience fully satisfies. The only reason I can imagine for denying that this is so is the observation that the descriptions of the data will involve the material object terms which figure in the commonsense realist talk of everyday life. But this reason is, of course, entirely inadequate. For these material object terms only occur within the scope of the "as if" operator, and it does not follow from the fact that it seems to me visually as if a zebra stood before me that a zebra did indeed stand before me. The

general point here is the simple one that it does not follow from the fact that one uses, or is competent to use, a given concept, either that that concept is instantiated, or that one has to believe that it is. If this were not so, then it would be impossible to speak of unicorns, phlogiston, witches or Santa Claus, perhaps in the course of denying their existence, without its being the case either that these things existed or that you believed that they did. This is why it is perfectly possible to describe the course of our perceptual experience by using the language of commonsense realism without thereby being committed to the truth of commonsense realism (see further Goldman 1988, pp.189-203). And it does not seem to make any difference to this why we must, if indeed we must, describe the course of our perceptual experience using this language. For instance, it does not make any difference if the reason is that the concepts of perceptual appearances can only be had by one who has the concepts of material reality; for one can be competent with the latter concepts without believing that they are instantiated.

Is there another way of understanding the original charge against the theoretical inference in question, that the descriptions of the relevant data are parasitical upon the theory to which we infer? For if the charge is that those

descriptions are parasitical in the sense that they presuppose the truth of the theory, then, as we have just seen, the charge is groundless. Perhaps, however, the thought is simply that the data are theory-laden, and, in particular, theory-laden with a realist theory, namely that of commonsense realism. But this suggestion leads to no sound objection. For the mere fact that data are theory-laden is, as I have already argued, unproblematic (see Chapter Three, Section Four). It may be thought that if the data are laden with the theory of commonsense realism, then that fact prejudices the outcome, so that a commonsense realist conclusion is unfairly favoured. But this strange thought is one that I can safely ignore. For the theory to which I shall eventually, in Chapter Seven, propose that we infer is not that of commonsense realism at all; so no prejudice can be at work here. So if the charge against the proposed theoretical inference is that the descriptions of the relevant data are parasitical upon theory merely in the sense that they are theory-laden, then the charge is harmless. The inference in question has no need of pure, uninterpreted, theory-free, sensory experiences as data.

Let me turn now to a different type of possible objection to the proposed inference to an external reality from the data constituted by the course of our perceptual experience. The

worry here is that a description of the course of our perceptual experience, in talking about sensations, would have to be in a private language, and hence liable to Wittgenstein's, or at any rate a Wittgensteinian, argument against the possibility of such a language.

But I do not believe that any traditional version of the private language argument is going to rule out the possibility of a language fit to describe the data which must be assumed to be describable for the purposes of the response to the sceptical challenge which I am defending (for an exposition of a traditional version, see perhaps McGinn 1984, p.48). For although, if they are sound, the traditional versions of the argument certainly demonstrate that there cannot be a private language in the sense of a language capable of describing sensations, these versions trade upon a certain assumption about the character of sensations, and neither the sceptic nor his opponent needs to accept this assumption. The assumption is that sensations are to be thought of as epistemically accessible only to their owners; the owner of a sensation enjoys this epistemic privilege in respect of his sensation. That such an assumption is made is clear from the following passage from Wittgenstein himself:

> But could we also imagine a language in which a person could write down or give vocal expression to his inner experiences -- his feelings, moods and the rest -- for his private use? -Well, can't we do so in our ordinary language? -But that is not what I mean. The individual words of this language are to refer to <u>what can only be known to the person speaking</u>; to his immediate private sensations. (Wittgenstein 1958, para.243; my emphasis)

And the assumption also appears in the recent development of the argument due to Crispin Wright (see Wright 1986, pp.209-210). Without the assumption, the argument will not go through, since the possibility is thereby opened up of their being an external check upon the consistency or otherwise of the application of the language's terms.

But neither the sceptic nor his opponent has to suppose that the sensations in question -- perceptual experiences -- <u>are</u> such that only their owners can know about them. The sceptical challenge can arise anyway. If this claim is not obvious, then consider the following possibility devised by John Mackie (albeit for a different purpose; see Mackie 1976, p.44). Imagine that humans are routinely born with a box fitted around their eyes in which a television screen is placed. All that they can see, in the ordinary sense, is the interior of the box and the surface of the screen, on which there are television images deriving from a camera mounted on the front of the box. Suppose, too, that no human has ever removed their box. Such

humans could no doubt exploit the information supplied to them by the television screens so as to construct views as to the nature of the world. In fact, I see no point in denying that they can <u>see</u> the world external to their boxes. But they might also entertain sceptical doubts. They might realise that it is possible to be supplied with non-veridical television pictures, perhaps because someone has disconnected your video camera and replaced it with a video recorder playing cartoons. They might then wonder how they could possibly tell that their usual television images were not just as they took cartoons to be -- representative of a fantasy world. So the humans with boxes over their eyes could devise the sceptical challenge. But televisions and television pictures are not such that only their viewers can know about them. So the sceptical challenge can arise without assuming that the immediate objects of perception are epistemically accessible only to their subjects, and the traditional versions of the private language argument cannot undermine it.

It remains possible, however, that the response to the sceptic's challenge that I favour presupposes the possibility of a private language in a different sense, and one against which there may be strong objections. Let us say that a language can be private in the sense that it is what we may

call a <u>born-Crusoe</u> language. A born-Crusoe language is a language devised and then used by a lone person in complete isolation. Its terms refer, let us say, to fully public objects, and there is no reason why, once devised, it may not be taught to and understood by other people. To claim that a born-Crusoe language is possible is to claim that if Robinson Crusoe had been born on his desert island and raised by animals, instead of being shipwrecked as an adult, he might have been able to devise for himself and then use a language with which to describe such things as the palm trees and sand around him.

It is at least plausible to think that the sceptical challenge, and the response to it that I favour, assume the possibility of a born-Crusoe language. For they seem to assume that it is possible to describe one's perceptual experiences without presupposing the existence of any particular external objects such as other humans. But is the assumption of the possibility of a born-Crusoe language correct? Some philosophers, such as Colin McGinn, believe that, intuitively, it is, and I agree (see McGinn 1984, pp.196-197). But I would withdraw my intuitive judgment in the face of a powerful argument to the contrary. So the question is whether there is such an argument.

The traditional versions of the private language argument are of no relevance here, since the objects referred to by a born-Crusoe language can be as public as you like and, as I have already noted, the traditional versions depend upon the assumption that the objects referred to by the terms in the putative language are such that, necessarily, only one person can know about them, and hence private. If what is necessary for there to be a language is genuine rule-following, and hence room for a distinction between really following a rule and merely seeming to do so, then born-Crusoe satisfies this necessary condition. For both his putative linguistic uses and the objects putatively denoted by those uses are in principle liable to independent checking, either by himself or by hypothetical others (see Davies 1988, p.55).

A more plausible candidate for an argument against the possibility of a born-Crusoe language is the argument which Saul Kripke, without endorsement, claims to find in Wittgenstein's writings (see Kripke 1982). Let me sketch Kripke's well-known argument and show how, as I believe, it entails the impossibility of a born-Crusoe language. (Kripke does not consider the case of born-Crusoe, though he has a brief discussion of conventional Crusoe; see Kripke 1982, p.110.)

The argument begins with this question: in virtue of what does a speaker mean one thing rather than another by some utterance? At first glance, there might seem to be several promising answers to this query. But, as Kripke's sceptic about meaning argues, each of these answers is unsatisfactory. Thus the sceptical conclusion is reached, though of course only provisionally, that meaning is impossible. However, all the suggested answers to the initial question propose certain facts only about the speaker as those facts in virtue of which a person means one thing rather than another by some utterance. This opens up the possibility, of course, that further facts, not merely about the speaker, can be conscripted to constitute meaning.

It is in exploiting this possibility that Kripke introduces the community. A speaker's utterance means one thing rather than another in virtue of certain facts, not merely about him, but also about the community. This account, we should note, is said by Kripke not to be one of the truth conditions of ascriptions of meaning to a speaker, but of their assertability conditions. Nevertheless, the account entails that it is only assertable that a speaker means one thing rather than another if certain facts about the community hold. And evidently those facts can hold only if the community exists.

So the community can be brought in, according to Kripke, to save meaning, and therefore language, from annihilation at the hands of the sceptic. But it cannot always be brought in. In particular, it cannot be brought in if no relevant community exists. And it is for this reason that born-Crusoe is in trouble, because by hypothesis he belongs to no linguistic community: there are no others using the same symbols he uses, with whom, for instance, his uses of those symbols might agree. So Kripke's Wittgenstein has to deny the possibility of a born-Crusoe language. (Conventional Crusoe, of course, is safe, because, although he is not in contact with its members, there does in fact exist a linguistic community using the same symbols that he uses.)

Is Kripke's argument a sufficiently powerful argument to override the intuitive judgement that a born-Crusoe language is possible? I propose to argue now that it cannot be, since it entails something unbelievable. Though this will not show where the argument errs, it will show that it errs.

Suppose that there is a nuclear holocaust and that one person alone survives. Such a person, if able to speak and write English before the holocaust, would surely still be able to do so afterwards. But if it is a necessary condition of a person's genuinely operating with a language that there exist a

community of other people using the same symbols, then the holocaust survivor, lacking a community, would not be able to do so. Kripke's argument seems to lead to this conclusion, which I think is unbelievable. And it appears that the possibility of a born-Crusoe language stands or falls with that of the holocaust survivor, the same difficulty afflicting both. Since that of the survivor stands, so too does that of born-Crusoe. It is no use replying that the possible existence of a community suffices to guarantee the survivor's language, since we can easily provide a merely possible community for born-Crusoe, and hence grant him a genuine language.

One might think to resist this argument by claiming that the case of born-Crusoe and that of the survivor so differ that whereas the place of the community in the assertability conditions of ascriptions of meaning destroys the possibility of a born-Crusoe language, it does not destroy that of the survivor's language. But what could that difference be? A tempting first thought is to say that whereas there was a community in the case of the survivor, there never was in the case of born-Crusoe. But surely that fact would make a difference only if the survivor could be said to be going on in the same way as the defunct community. But according to Kripke's sceptical paradox, in response to which the community

was introduced in the first place, it makes no sense to speak of a single individual going on in the same way; you need a community for that. So the survivor would need another community in virtue of which he could be said to be going on in the same way as the defunct community. What else might the difference be between the two cases? All I can suggest is that in the born-Crusoe case there is meant to be the invention of a language, whereas the survivor does not have to invent a language, but merely use one. But it is hard to see how this is relevant at all. For the argument of Kripke's sceptic was not concerned with some problem about the invention of a language, but with the conditions constitutive of the continuing existence of one.

I conclude, then, that we may safely ignore the upshot of Kripke's private language argument that a born-Crusoe language is impossible. For, by entailing the impossibility of the survivor's language, it has fallen foul of an essential constraint on private language arguments, namely that they do not rule out perfectly respectable instances of (public) language. There must be something wrong with it. And so the data for the proposed use of inference to the best explanation to justify belief in an external world involve no improper dealings with a private language.

CHAPTER SIX

The Explanatory Test Defended: II

The version of the explanatory test for reality expressed by the key premiss (P1) in my positive argument for eliminative type physicalism is biconditional in form; and in the preceding chapter I defended one of the conditionals from which it is composed, the one affirming that it is a sufficient condition of justified belief in a property that it be mentioned in the best explanation of the course of our perceptual experience. In this chapter, I shall defend the other conditional, which is to say that I shall defend the claim that it is a necessary condition of justified belief in some property that the property be mentioned in the best explanation of the course of our perceptual experience.

I take this claim at least to involve the claim that no response distinct from that based on inference to the best explanation is adequate to meet the challenge of scepticism about the external world. Accordingly, my ideal strategy would involve the systematic refutation of all responses to the sceptic about the external world other than the one I favour,

and this would most naturally be thought to require a discussion of so-called transcendental arguments. But my actual strategy will fall far short of this ideal. For example, I shall not discuss any of the traditional, Kant-inspired, transcendental arguments at all, since I endorse the replies to them that have already appeared in the literature (see, for instance, Stroud 1968, Brueckner 1983 and 1984a, Cassam 1987, Goldman 1988, pp.189-203). Instead, I shall do two things: first, I shall discuss a couple of arguments, due to Donald Davidson and Hilary Putnam, which are, or at least can readily be seen as, a priori responses to scepticism about the external world and which, if successful, might warrant our acceptance of a more extensive ontology than is warranted by my preferred response to scepticism; and then I shall criticise a completely different idea to the effect that we should believe, not only in those properties mentioned in the best explanation of the course of our perceptual experience, but also in whichever properties can be constructed, according to certain principles, out of those original properties.

My reason for singling out Davidson and Putnam for special treatment in this way is two-fold. First, their respective arguments are interesting, recent attempts to show that,

contrary to the sceptic's claim, we can rule out the sceptical possibility that we are, for example, brains in vats stimulated to believe what we do believe, even though our beliefs are nearly all false, and that we can rule out sceptical possibilities of this sort simply by reflection on the nature of language and hence without resort to anything like inference to the best explanation. We can know a priori, they claim, that we are not massively deceived in the way the sceptic supposes to be possible, so that, according to them, most of our ordinary beliefs, including our beliefs about what exists, are true. Secondly, however, and consequent upon this last claim, if Davidson and Putnam are right to hold that we know that we are not in massive error, then eliminative type physicalism cannot be correct, since it entails that most of the beliefs of most people, being about non-existent things, are false, so that massive error is not only possible, but actual. Thus I have this second philosophical motivation for wanting to see off the Davidson and Putnam arguments.

1. Davidson

Donald Davidson (1984, Essay 14 and 1986) holds that, if his account of meaning, belief and interpretation is correct, most of our beliefs must be true:

> Once we agree to the general method of interpretation I have sketched, it becomes impossible correctly to hold that anyone could be mostly wrong about how things are. (Davidson 1986, p.317)

And Davidson infers from this the conclusion, astonishing on its face, that most of our beliefs must be true.

Let me begin with my own reconstruction of the Davidsonian argument. His conclusion, of course, is that most of our beliefs must be true. But how is this to be given more rigorous expression? A natural first thought is that, in the useful terminology of possible worlds, he is claiming this:

> (1) In all possible worlds, what we believe in those worlds is mostly true.

But I hesitate to ascribe this very strong claim to Davidson, which is stronger, indeed, than he needs. All we must suppose is that Davidson maintains the weaker claim, entailed by (1), that

> (2) In all those possible worlds in which we believe what we believe in the actual world, what we believe is mostly true.

Whereas (1) expresses the Davidsonian conclusion understood as involving necessity <u>de dicto</u>, (2) takes it to be a claim of necessity <u>de re</u>: our actual beliefs have the property of being

essentially true.

(2) is quite sufficient for Davidson's purposes. For it contradicts the sceptical contention, dramatised in stories about evil demons or brains in vats, that

> (3) There is a possible world in which we believe what we believe in the actual world, but in which what we believe is mostly not true.

Furthermore, from (2), together with the triviality that

> (4) The actual world is a possible world in which we believe what we believe in the actual world,

we can infer the comforting conclusion that

> (5) What we believe is mostly true.

But (2) is not self-evidently correct, and we need to enquire what is supposed to be the argument which supports it. Its first premiss -- to continue with my reconstruction -- is the following Principle of Charity in interpretation:

> (6) Interpretation is possible only if the interpreter assumes the beliefs of the subject(s) to be mostly true.

Since the interpreter's own beliefs represent his best guess as

to what is true, we can say that

> (7) Interpretation is possible only if the interpreter ascribes to the subject(s) beliefs mostly in agreement with his own.

The second premiss appeals to the notion of an <u>omniscient interpreter</u>, an interpreter who believes all the true beliefs there are:

> (8) In all those possible worlds in which we believe what we believe in the actual world, we are interpretable by an omniscient interpreter.

From (8) and (7) it follows that

> (9) In all those possible worlds in which we believe what we believe in the actual world, our beliefs are mostly in agreement with those of an omniscient interpreter.

But since by definition the beliefs of an omniscient interpreter are all true, we can deduce the required conclusion that

> (2) In all those possible worlds in which we believe what we believe in the actual world, what we believe is mostly true.

Davidson's argument, which I hope to have reconstructed

fairly and accurately, is, I think, formally valid. Nevertheless, the argument is unpersuasive, because Davidson offers no satisfactory reasons for accepting its two main premisses, (6) and (8).

I begin with (8), which claims that in all those possible worlds in which we believe what we believe in the actual world, we are interpretable by an omniscient interpreter. Why should we accept this? The claim of possibility that it makes is as weak as it can be if it is to sustain the argument. But it is still, I suspect, rather stronger than Davidson realises, and, crucially, it is not supported by the existence of various other, but less exotic, possibilities. Let us consider some of these milder possibilities.

Of course there could be an omniscient interpreter, if this means simply that there could be an omniscient person capable of <u>doing some interpreting</u> (see Davidson 1984, p.201 and 1986, p.317). But this possibility is of no interest in this connection, because this person might be able to interpret only very few people, perhaps only those who were also omniscient, and this possibility evidently does not ensure that he could interpret us. Perhaps this failure does not matter, because it is obvious in any case that there could be an omniscient person capable of interpreting <u>us</u>. But even this further possibility

is insufficiently strong, because the possible worlds in which omniscient persons are able to interpret us might all be worlds in which we happen to believe only truths. The worlds that are relevant for Davidson's argument are those in which we believe, whether truly or falsely, what we believe in the actual world. And Davidson needs it to be the case that in all <u>those</u> worlds an omniscient interpreter is capable of interpreting us. But this is very far from being obviously true, especially when we appreciate that this alleged possibility is distinct from those that just now I conceded to be genuine. Moreover, it is precisely in regard to the relevant worlds that our hesitation is greatest. For if we suppose, as does the sceptic, that there are worlds in which we believe what we believe in the actual world, but where what we believe is radically false, then we might well think that it is in just those worlds that an omniscient interpreter is <u>not</u> capable of interpreting us (cp. Rasmussen 1987). And nothing said by Davidson either shows or even suggests that this would be a silly thing to think. So the sceptic need feel no duty to accept (8).

A defender of Davidson might at this point suggest that, on the contrary, this <u>is</u> a ridiculous thought, on the anti-realist grounds that to have beliefs <u>just</u> <u>is</u> to be <u>interpretable</u> <u>as</u> having beliefs; so that if, in the worlds in question, we were

not interpretable by the omniscient interpreter, we would *eo ipso* have no beliefs. But this response is inadequate, even though, as it seems to me, it is very much in the spirit of Davidson's philosophy of mind. For although it is true that, in the relevant worlds, we are not interpretable by the omniscient interpreter, there is no reason why we should not be interpretable by persons who are less than omniscient and especially, of course, by those who are as deluded as we are; in which case, even conceding this form of anti-realism about the mental, we would still have to be judged as having beliefs. Perhaps the response could be saved by strengthening the anti-realism so that it claimed that to have beliefs just is to be interpretable by *all* believers as having beliefs; but although this modification would certainly support the conclusion that, in the relevant worlds, we had no beliefs, the strengthened anti-realism required by it seems absurd.

It should now be clear why Davidson is in error when he writes:

> ...it is plain why massive error about the world is simply unintelligible, for to suppose it intelligible is to suppose there could be an interpreter (the omniscient one) who correctly interpreted someone else as being massively mistaken, and this we have shown to be impossible.(Davidson 1984, p.201)

This is incorrect. The believer in the intelligibility of massive error need not suppose that the omniscient interpreter could interpret the persons said to be in massive error at all; hence he need not suppose that the omniscient interpreter would do something -- namely, interpret the persons as holding massively erroneous beliefs -- allegedly demonstrated by Davidson to be impossible.

The second premiss of Davidson's argument, (8), is therefore a highly controversial claim, unsupported by anything Davidson says, and unlikely to be given away by an opponent. It does not seem to me to be a sufficiently robust foundation on which to set the astonishing conclusion that most of our beliefs must be true. (In case you think that Davidson can mount an anti-sceptical argument without this premiss, that is, relying only upon his Principle of Charity, (6), see Brueckner 1986a.)

I turn now to Davidson's first premiss, (6), which expresses his Principle of Charity in interpretation, to see whether it fares any better. It states, of course, that

> (6) Interpretation is possible only if the interpreter assumes the beliefs of the subject(s) to be mostly true.

It cannot be said that Davidson's case for (6), despite its

frequent appearance in his writings, is presented with enormous clarity or precision; and so I lack complete confidence in the success of my attempt to undermine it. Nevertheless, I shall set out and then criticise two lines of argument for it which Davidson appears to endorse. (A different -- evolutionary -- argument for it is well criticised in Stich 1985, pp.256-260.) Of these the first represents (6) as following very naturally from a plausible account of radical interpretation, while the second involves considerations from the theory of reference. I start with the first.

As is well known, Davidson holds that a theory of truth for a language, in the style of Tarski, can serve as a theory of meaning for the language. To make this possible, he introduces a modification to Tarski's Convention T:

> ...an acceptable theory of truth must entail, for every sentence s of the object language, a sentence of the form: s is true if and only if p, where "p" is replaced by any sentence that is true if and only if s is. Given this formulation, the theory is tested by evidence that T-sentences are simply true...(Davidson 1984, p.134)

Davidson's problem, then, is to show how the empirical testing mentioned in this passage is possible in the light of the interdependence of belief, on the one hand, and meaning, on the other. And he thinks that he can do this by building on

the alleged insight that it is possible for a radical interpreter to tell when a subject holds a sentence true, even when the meaning of the sentence is unknown. The empirical confirmation which can then occur is illustrated by this Davidsonian example.

Suppose our theory generates this T-sentence:

> (T) "Es regnet" is true-in-German when spoken by x at time t if and only if it is raining near x at t.

Then we might well have this item of evidence:

> (E) Kurt belongs to the German speech community and Kurt holds true "Es regnet" on Saturday at noon and it is raining near Kurt on Saturday at noon.

And it would be proper, as Davidson notes, to "consider (E) as evidence that (T) is true" (p.135), even though it does not of course constitute "conclusive evidence" (p.136).

If one were to read only this far in Davidson's account of the testing of theories of meaning, then it would appear to be a perfectly standard account of theory testing: theories of meaning are empirical theories which yield testable consequences. But Davidson himself evidently does not see it this way. For he writes:

> We want a theory that...maximises agreement, in the sense of making Kurt [and others] right, as far as we can tell, as often as possible (Davidson, 1984, p.136; my emphasis).

What is extraordinary here is that nothing in the need to confirm T-sentences empirically, or in the preceding account of how that can be done, seems to provide the slightest reason for accepting this new constraint on theories of meaning. Notice, too, that the constraint is not meant as a defeasible one -- part of a strategy successful for the most part, but capable in principle of being overridden -- or else it could not yield a claim as strong as (6). And it needs emphasising that such a constraint would ordinarily be regarded as an outrage if it or anything like it were proposed in the context of the testing of any other scientific theory. For the proposed constraint constitutes a substantive restriction, to be imposed a priori, on the content of any acceptable theory of meaning; whereas if we were devising theories in any other scientific domain no specific theory could be ruled out in this aprioristic fashion, in advance of empirical investigation (cp. Devitt and Sterelny 1987, p.247). We need to be told exactly why it is that theories of meaning do not aim, as other scientific theories do, at successful prediction and explanation, so that they should be tested empirically in the same -- wholly a posteriori

-- way as other scientific theories. (Precisely this view of interpretation is suggested in Grandy 1973, p.442 and in McGinn 1977, p.534.)

On its face, it looks as if (E) is evidence for (T) because the fact reported by (E) would be well explained if (T), together with a number of other plausible assumptions, were true. And it also looks as if among those plausible assumptions is the claim that Kurt's mind and perceptual apparatus are in reliable working order. That is to say, it is partly because we assume Kurt to be a generally reliable detector of the presence of rain that we take the sentence which he is holding true to mean that it is raining. But we do not have to assume this, and it is hard to see why the relaxation of this assumption should disable us from constructing a theory of meaning which, in conjunction with a suitable ascription of propositional attitudes, served to explain the subject's behaviour. In other words, the Principle of Charity might be a defeasible methodological rule of great utility, but it is obscure how it could be anything more than that.

Davidson, however, is adamant that the Principle of Charity is not merely a defeasible methodological assumption. He gives this reason for his stand:

> If we cannot find a way to interpret the utterances and other behaviour of a creature as revealing a set of beliefs largely consistent and true by our own standards, we have no reason to count that creature as rational, as having beliefs, or as saying anything. (Davidson 1984, p.137)

In assessing this passage, we should ignore the word "consistent", since we are not here concerned with any alleged necessity to ascribe rationality or consistency to subjects of radical interpretation. (Such alleged requirements are criticised in Stich 1985 and Levin 1988.) With this excision, are Davidson's claims correct? Let us examine them on the plausible assumption that the essence of having beliefs is that one enjoy states which represent the world, and that one's enjoyment of these states play some role in the causal explanation of one's behaviour (see Fodor 1987, pp.10-16).

Could one have mostly false beliefs and still be rational? Evidently one can be rational and have <u>some</u> false beliefs; and it is hard to see why one would suddenly cease to be rational if <u>most</u> of one's beliefs were false -- one's environment might just be exceptionally deceptive or one's perceptual equipment exceptionally unreliable. Surely we would have reason to count a creature with preponderantly false beliefs as rational just in case, given our beliefs about its environment and perceptual

apparatus, it seemed to have formed its beliefs rationally in the light of the evidence available to it. Could we ever be right to ascribe mostly false beliefs to a creature? Certainly; we would have reason to ascribe mostly false beliefs to a creature if such an ascription was part of the best explanation of the creature's behaviour, which in imaginable circumstances it might be. Finally, if "saying anything" is a matter of communicating, or attempting to communicate, some of the contents of one's inner representational states, then clearly such attempts could be made even by a creature most of whose beliefs were false; and we could surely have evidence that this was happening, especially if we could observe the social interactions of the creature.

So much, then, for Davidson's apparent attempt to show that the Principle of Charity falls naturally out of his account of radical interpretation. It does not. On the contrary, it sits uncomfortably with it.

Let me turn, then, to Davidson's second argument for his Principle of Charity. He often makes claims such as this one:

> ...disagreement and agreement alike are intelligible only against a background of massive agreement. (Davidson 1984, p.137)

Now this evidently needs some filling out, if it is even to begin to seem persuasive, and I think that such expansion is provided by the passage from which the following quotation is taken:

> Before some object in, or aspect of, the world can become part of the subject matter of a belief (true or false) there must be endless true beliefs about the subject matter.....To take an example, how clear are we that the ancients -- some ancients -- believed that the earth was flat? <u>This</u> earth? Well, this earth of ours is part of the solar system, a system partly identified by the fact that it is a gaggle of large, cool, solid bodies circling around a very large, hot star. If someone believes <u>none</u> of this about the earth, is it certain that it is the earth that he is thinking about? (Davidson 1984, p.168)

Perhaps Davidson has in mind the following argument. Suppose it is the case that

> (10) P and Q disagree only if there exists some object to which they can both refer, and to which P and Q respectively ascribe incompatible properties.

Presumably, then, it is also the case that

> (11) P can <u>know</u> that P and Q disagree only if P can know that there exists some object to which P and Q can both refer, and to which P and Q respectively ascribe incompatible properties.

Then if it is also the case that

(12) P can know that there exists some object to which P and Q can both refer only if P and Q share some beliefs about the object,

it follows that

(13) P can know that P and Q disagree only if P and Q share some beliefs about some object.

This argument, it should be noticed, leads to an epistemic conclusion, and not to a constitutive one. It purports to reveal a necessary condition of our knowing that we disagree with another, and not a necessary condition of our simply disagreeing. And that is as well, since it is evidently possible for you and I to disagree about some one object even though we share no beliefs about it at all. For suppose that there is some object which is F and G; then if I think of it as 'the F', and you think of it as 'the G', we can both be referring to it; but we do not then have to share any beliefs about it at all, and I may deny that it is G, and you may deny that it is F.

This conclusion, however, is not, of course, (6), which requires interpreters to find their subjects with mostly true beliefs. But it is close to (7), which was in the original argument inferred from (6), and which is sufficient for

Davidson's purposes.

A first worry about the argument just presented concerns whether its conclusion, (13), is close enough to (7). For according to (7), of course, interpretation is possible only if the interpreter ascribes to the subject beliefs mostly in agreement with his own; whereas (13) merely implies the weaker claim that the interpreter must share some beliefs with the subject found to be in error. How this gap is to be bridged, I do not know. Nor is it insignificant. For the defender of the possibility of brains in vats, stimulated to believe falsely what we believe, can argue that the omniscient interpreter need not fall foul of this weaker constraint in his effort to interpret the brains. For he can ascribe to the brains mostly, or even wholly, true beliefs about the character of their perceptual experiences, but mostly false beliefs about the external world (compare, perhaps, McGinn's proposed method of interpretation in McGinn 1986). It is unclear why this base of true beliefs about the character of perceptual experience would not suffice to make disagreement between the brains and the omniscient interpreter intelligible, even given the truth of (13).

But the argument for (13) is in any case unsatisfactory, since two of its premisses, (10) and (12), are false. (10) is

false because it is not a necessary condition of two persons' disagreeing that they have incompatible beliefs about some one object. There is another, and perhaps more important, type of disagreement, namely ontological disagreement, whereby two persons disagree as to what exists. And ontological disagreement does not require a single object about which those disagreeing hold incompatible beliefs.

(12) is also false. Let us return to the case where some object is both F and G, and where I think of it as "the F" and you think of it as "the G", and where I deny that it is G and you deny that it is F. In this case, we share no beliefs about the object, but we nevertheless disagree. Why could I not come to know that we disagreed in this way? Suppose that I come to grasp the sense of your referring expression, "the G". Then I could realise that you use the expression to refer to what I refer to as "the F" so long as I was able to conclude that you believed, falsely in my view, that the object I refer to as "the F" was G. (I would not need to accept this belief.) And I could come to the relevant conclusion on the basis of observation of which object tends to provoke your talk of "the G"; I might notice that it was the very object I think of as the F. I could then judge that you were referring to that very object, not in the sense that it satisfied a description you

had, but nevertheless in an intelligible sense. (If you doubt that this is an intelligible sense, or if you suppose that my appeal to it commits me to some causal theory of reference, then I can redescribe the disagreement as an ontological one, whereby you believe that something is G whereas I believe that nothing, or at least nothing relevant, is G.)

I conclude, then, that Davidson supplies no convincing reason for accepting the crucial premiss (6), the Principle of Charity, in his argument for the astounding conclusion that most of our beliefs must be true; and that this fact constitutes another excuse for declining to endorse that conclusion.

2. Putnam

Hilary Putnam (1981, Ch.1) has argued that, despite first appearances and a long sceptical tradition, it is not really possible that we might be brains in vats, stimulated by some super-scientist to believe that we were not; so that the kind of radical error supposedly illustrated by examples of this sort is not really possible either. How does he argue for this fascinating conclusion? And does he do so soundly?

In fact, Putnam argues for two distinct conclusions, both of which seem to hold some terror for the sceptic. I shall deal

with them in turn, displaying Putnam's arguments for them and then indicating why they do not succeed. (For other critiques along similar lines, see, for example, McIntyre 1984 and especially Brueckner 1986b.)

The first conclusion is that

> (14) If we are brains in vats, then we cannot think that we are brains in vats.

And this must cause the sceptic some alarm, since in conjunction with the claim that

> (15) We can think that we are brains in vats,

it entails that

> (16) We are not brains in vats.

And if we can correctly be said to know (14) and (15), then it would be hard to deny that we know (16) also.

Why believe (14)? Because it follows trivially from the claim that

> (17) If we are brains in vats, then we cannot think of anything at all.

Why, then, believe (17)? Because it is entailed by the claim that

> (18) If you can think of a thing, then you must be causally connected to it in the right sort of way,

together with the further claim that

> (19) If we are brains in vats, then we are not causally connected in the right sort of way to any of the relevant things.

Now it is evident that in this Putnamian argument for (14) the contestable premisses are (18) and (19); and I shall in fact contest them both. Their falsity suffices to defuse the argument.

(18), of course, expresses a part of a causal theory of reference, and it is clearly the corner-stone of the whole argument. But it is evidently false. It cannot be a requirement on our being able to think of a thing that we be causally connected to the thing, since we can think of things and stuffs, such as witches, phlogiston, Santa Claus and the aether, to which we cannot possibly be causally connected in any way at all, since they do not exist. Moreover, even if (18) is correct, but only in respect of what might be called successful reference -- reference to existents --, we still

need an account of how we can speak meaningfully of non-existents, that is, how empty terms can nonetheless play a part in meaningful utterances. And such an account manifestly cannot include a requirement of the sort expressed by (18). Moreover, I see no reason to expect that it will have the upshot that, if we are brains in vats, then we cannot think that we are.

(19) expresses the claim that if we are brains in a vat, then we are not causally connected in the right sort of way to any of the relevant things, the relevant things being, of course, those things about which we are supposed to have thoughts. The reason for asserting it would presumably be the correct observation that, according to the sceptical hypothesis, we are brains in vats stimulated to have many beliefs about things which do not so much as exist at the time we think about them -- that is why we can be said to be in massive error --, and if these things do not exist, then we evidently cannot be causally connected to them. But this observation is insufficient to establish (19). For the sceptic can elaborate his hypothesis in such a way that we do enjoy the requisite causal connections, but without denying the observation.

What the sceptic must suppose is this. In the past, perhaps recently, the external world did exist, and we were situated in

it, causally connected in the right sort of way to all the familiar objects of common sense. Consequently, we were able to think and talk about those objects. But at some time since last week, perhaps while we were asleep, the superscientists moved in and did two things. First, they extracted our brains from our skulls and set them up in vats; secondly, they destroyed our familiar external world. In doing these things, the superscientists created a situation in which we have many beliefs about familiar external objects, but in which those objects do not exist. However, and this is the crux, we are perfectly able to have thoughts about the familiar objects because we are causally connected in the right sort of way to objects of the right type, i.e. the objects destroyed by the superscientists. Such historical connections are usually allowed by causal theorists to suffice for reference, which is as well, because if they did not, we would be unable to think and speak today of dodos, which are extinct. The sceptic, then, need not accept (19).

So the argument for (14) fails. But even if (14) were defensible, which I have argued that it is not, it would be at best a paradox of moderate ingenuity, unless it could be combined with (15) to produce an argument for the truth of (16), which states the crucial conclusion that we are not

brains in vats. But could this be done? In particular, is the intermediate premiss, (15), true?

(15) states that we can think that we are brains in vats, and is most simply defended by inferring it from the apparently obvious claim that sometimes we actually do think that we are brains in vats. But although the inference involved in this simple argument is unimpeachable, Putnam's strongly externalist view of meaning and thought -- according to which we may be unable to think that we are brains in vats, even by thinking "we are brains in vats" (see Putnam 1981, p.8) -- renders the truth of the premiss far from obvious. For Putnam's view entails that, if we are to know that we are <u>really</u> thinking that we are brains in vats, and not simply seeming to do so, then we must know that we are causally connected to brains and vats, and hence that such external objects exist. But the assumption that we do have such knowledge can hardly figure as a premiss in an argument against a sceptic about the external world, who will be apt to protest, and with justice, that the question has been crudely begged against him. And, of course, if, in order to preserve the apparently obvious truth of the claim that we sometimes do think that we are brains in vats, we abandon the Putnamian externalism about thought and reference, then we lose all title to the truth of the other premiss, (14).

(15), therefore, may well be true; and surely it is. But it can play no part in a Putnamian argument against scepticism.

Let me turn now to Putnam's second anti-sceptical conclusion, which is that the supposition that we are brains in vats cannot possibly be true, because it is self-refuting (see Putnam 1981, p.7). His argument for it can be given the form of a dilemma, and runs as follows. (Other reconstructions are presented, and then refuted, in Brueckner 1986b.)

Either we are brains in vats, or we are not. If we are not, then, trivially, we are not. But if on the other hand we are brains in vats, then "We are brains in vats" is false, and so we are not brains in vats. If we are not, then we are not; but if we are, then we are not. So we are not.

The controversial part of this argument is, of course, the claim that

> (20) If we are brains in vats, then "We are brains in vats" is false,

and the inference from it of the claim that

> (21) If we are brains in vats, then we are not brains in vats.

I begin with (20). It is represented by Putnam as an implication of the causal theory of reference. His claim is that although, if we were brains in vats, we would not be causally connected to the objects of our everyday ontology, we would still be causally connected to something, perhaps electronic impulses of certain kinds. Therefore, on the assumption that we refer to whatever we are causally connected to in some appropriate way, we would be thinking something, such as "We are electronic impulses". But this thought would, of course, be false, since we would be brains in vats and not electronic impulses.

This argument, however, is fantastic. It turns on the idea that we refer to whatever we are causally connected to in the right sort of way. Notice, first of all, that (18), which can perhaps be plausibly argued for, has no tendency to support this idea, and thence (20), since it states only an alleged <u>necessary</u> condition of reference, whereas what is evidently needed here is a causal account of the <u>sufficient</u> conditions of reference. But no such account of the sufficient conditions of reference is, to my knowledge, in existence. For although, if a causal theory of reference is correct, there will be, no doubt, <u>some</u> specifiable set of causal conditions sufficient for reference, we have at present no idea what it is; and, until

such time as we do, it is the sheerest speculation on Putnam's part to suppose that he knows what envatted brains would be thinking by thinking "We are brains in vats", and even more fantastic for him to claim to know that what they would be thinking would be false. The truth of (20) is something upon which the causal theorist simply has no right at present to pronounce; and so the acceptance of it is about as far from being compulsory as can be. If we are brains in vats, then "We are brains in vats" may be false; but there again, for all that Putnam tells us, it may be quite meaningless or even true.

So much, then, for (20). If it could be established, does it entail (21)? Not in any way, I shall now argue, that assists Putnam's argument.

We need to ask this question: what semantics -- vat or ordinary -- are we to apply to the antecedent of (20) and the whole of (21)? Suppose, first, that we apply ordinary semantics, on the assumption that we are not brains in vats; then the falsity of the vat sentence "We are brains in vats" does not permit the inference from (20) to (21), since the consequent of (21) is an ordinary sentence stating that we are not brains in vats. If, on the other hand, we apply vat semantics, on the assumption that we are brains in vats, then (20) certainly entails (21). But the conclusion ultimately

established can then only be that we are not electronic impulses, since the same semantics must be applied consistently, and this is not what we wanted to establish. So even if (20) were plausible, Putnam's attempt to demonstrate the self-refuting character of the supposition that we may be brains in vats does not succeed.

3. Constructionism

In the previous two sections, I have been defending my claim that it is a necessary condition of justified belief in a property that the property be mentioned in the best explanation of the course of our perceptual experience against arguments intended to show *a priori* that most of our beliefs, and hence presumably most of our ontological beliefs, must be true. In this section I will defend my claim against a quite different line of criticism. According to it, although we are certainly justified in believing in those properties mentioned in the best explanation of the course of our perceptual experience, we are also justified in believing in whatever properties can be *constructed out of* the explanatorily indispensable properties in accordance with certain principles. We may call this line of criticism *constructionism*, and I shall distinguish between two versions of it, conjunctive and disjunctive. The aim of both versions is to show that even if we start with the minimal

physicalist ontology which I favour, we cannot stop there, since we must also admit the constructed ontology.

Constructionism, especially in its conjunctive form, must not be confused with two other, and quite distinct, ideas, which I shall now briefly describe. The first idea is this. Some writers, reflecting on the issue of whether commonsense ontology is compatible with that of microphysics, have thought to resolve any alleged incompatibility by suggesting that, if we take microphysical particles to be <u>parts</u> of <u>wholes</u> which are macro-objects, then it will be fully consistent to believe in both macro-objects and microparticles, just as the belief that dry stone walls exist is in no way compromised by the further belief that, in some acceptable sense, there is no more to a dry stone wall than the dry stones from which it is composed (see Ayer 1973, p.110).

But this suggestion, whatever its merits, is not a constructionist one, since it does not even purport to show that, if one accepts the microphysical ontology, one must accept that of macro-objects as well. It ought, instead, to be addressed to someone who, unlike me, <u>already</u> finds himself inclined to believe in <u>both</u> macro-objects <u>and</u> microparticles, but who is wondering whether these two ontological beliefs are consistent with one another. And of course it is no threat to

my eliminative type physicalist position to learn that they are. For I have not claimed that belief in macro-objects is straightforwardly inconsistent with belief in microparticles; my argument, of course, is that they are rivals at a deeper, epistemic level.

The second idea with which constructionism must not be confused is best introduced by recalling a point made by Quine in response to an argument, very much in the spirit of eliminative type physicalism, once advanced by Grover Maxwell (see Maxwell 1968):

> I disagree, too, when he denies bodies their color because they are collections of submicroscopic particles. Water remains water gallon by gallon, I say, even though its submicroscopic bits are rather oxygen and hydrogen; there is no paradox in this...The qualities of being aqueous and of being smooth and brown are like swarming or waging war: they are traits only of a congeries. This does not make them unreal or subjective. (Quine 1981, pp.176-178)

What Quine says here is perfectly correct, but it is an ignoratio elenchi against Maxwell's views, and would be equally so against mine. For neither Maxwell nor I argue that bodies lack colour or tablehood because they are collections of microparticles. Our claim, rather, is that belief in colour and tablehood is unnecessary to explain why things look

coloured or tabloid, and is therefore epistemically gratuitous. Quine, in effect, falsely accuses Maxwell of trying to argue, from the premiss that, in some acceptable sense, there is no more to a dry stone wall than its dry stones, to the conclusion that dry stone walls do not exist. And, as Quine notes, this inference is certainly fallacious, since it unjustifiably assumes the impossibility of emergent properties, that is, properties had by sets or collections of microparticles. But Maxwell's argument does not depend upon this fallacious inference, and neither does mine.

With these two red herrings cleared from my path, let me turn now to a discussion and criticism of the conjunctive form of constructionism. The intuitive idea behind conjunctive constructionism is this: surely the eliminative type physicalist, since he believes in the existence of microparticles, is committed to believing also in the existence of such things as tables and chairs and elephants and kidneys, things which are obviously not themselves microparticles but which we are inclined to say are made up of microparticles -- for how can one believe in the existence of parts without also believing in the wholes which they make up?

This intuitive idea can be made precise in more than one way. The simplest form of conjunctive constructionism is as

follows. (Later I discuss a more sophisticated form.) We are always justified in believing in sets, or mereological sums, of things we are already justified in believing in, so that, since, as the eliminative type physicalist allows, we are justified in believing in microphysical particles, we are therefore also justified in believing in sets, or mereological sums, of microphysical particles. Since ordinary objects -- tables, chairs, elephant, kidneys -- just are sets, or mereological sums, of microparticles, the eliminative type physicalist must also believe in tables, chairs, elephants and kidneys. And, of course, if all this is correct, then it cannot be a necessary condition of justified belief in a property that it be mentioned in the best explanation of the course of our experience. For we may also be justified in believing in those properties which are the sets, or mereological sums, of the explanatorily indispensable properties.

Against this simplest form of conjunctive constructionism I have two objections. The first is that there seems to be no sense in which it is correct to say that ordinary objects, such as tables and kidneys, <u>just are</u> sets, or mereological sums, of microparticles. The phrase 'just are' denotes numerical identity. But the <u>type</u>, table, cannot be identified with any

set, or mereological sum, of microparticles of specified types, for the simple reason that all sorts of physico-chemically very different things count as tables. Nor can the type, kidney, be identified with any set, or mereological sum, of microparticles of specified types, since physically different organs belonging to different animal species all count as kidneys (not to mention possible artificial kidneys).

The prospects for identifying ordinary objects with sets, or mereological sums, of microparticles do not improve when we turn from types to tokens. For no particular table, nor any particular kidney, can be identified with a set, or mereological sum, of microparticles. This is because the standard identity conditions for sets, or mereological sums, entail that a set, or mereological sum, of microparticles could not lose or gain so much as a single microparticle without thereby ceasing to exist, whereas a particular ordinary object, such as a table or a kidney, can certainly lose or gain an atom or two without thereby ceasing to exist -- indeed, it happens all the time. So the simplest form of conjunctive constructionism has no title to its crucial premiss that ordinary objects just are sets, or mereological sums, of microparticles.

My second, and main, objection, however, to the simplest form

of conjunctive constructionism, is this: the principle of construction at its heart is false. For it does not seem to me in general to follow, from the fact that things of a certain sort exist, that sets, or mereological sums, of those things exist also (see Hornsby 1985). Suppose we take the set, or mereological sum, consisting of my fountain-pen, Christ Church Cathedral and your copy of Word and Object, and let us call this set, or mereological sum, Fred. Does Fred exist, over and above its components? According to my intuition, Fred does not exist. Again, suppose that to be a glurb is to be any fountain-pen, any cathedral and any book. Are there such things as glurbs? I think not. Finally, for fans of events, suppose that we take the set, or mereological sum, consisting of the death of Julius Caesar and the Battle of Hastings and call it the Battle of Julius Caesar. Is there an event to which we can refer using the phrase, "the Battle of Julius Caesar"? Again, I think not. And if you deny the existence of Fred, glurbs and the Battle of Julius Caesar, then you must deny the principle of construction on which the simplest form of conjunctive constructionism depends, since if that principle were correct, your denial would be erroneous.

Perhaps, however, your intuitions about these cases differ from mine. Nevertheless, the simplest form of conjunctive

constructionism is implausible, because it faces two difficulties. First, it seems to licence an apparently unlimited expansion of your ontological commitments, since any number of cases on the pattern of my earlier examples can be constructed; and this is an unattractive prospect. But secondly, an advocate of the simplest form of conjunctive constructionism wants to use it to vindicate some of the special scientific ontology, especially ordinary objects, which eliminative type physicalism rejects. But this vindication can only be of an unsatisfactorily feeble kind. For by the use of the relevant principle of construction one could only hope to show that some particular kidney (say) was as real as Fred, kidneys as real as glurbs, and any kidney-event as real as the Battle of Julius Caesar. And I doubt that this kind of reality is real enough. It seems to do insufficient justice to the intuitions of typical believers in ordinary objects.

So much, then, for conjunctive constructionism in its simplest form, and my objections to it. Conjunctive constructionism can also be defended in a sophisticated form, and it is to consideration of this that I now turn. Sophisticated conjunctive constructionism affirms that to be an ordinary object just is to be any collection of microparticles of any type, so long as the collection constitutes an

individual belonging to a natural kind (or, in the case of artefacts such as tables and chairs, an artefactual kind). The sophisticated conjunctive constructionist can then claim that the principle of construction on which the simpler form of conjunctive constructionism relies -- the principle criticised in my second objection -- fails because it is prone to counterexamples of sets, or mereological sums, of microparticles which do not constitute individuals belonging to either natural or artefactual kinds. So if the principle of construction is refined so that it applies only to collections of microparticles which form individuals belonging to either natural or artefactual kinds, it will be immune to the sort of counterexample which discredited the simpler principle relied upon by the simplest form of conjunctive constructionism.

The sophisticated conjunctive constructionist can also claim to be immune to the first objection which I brought against the simplest form of conjunctive constructionism. That objection was that the needed identity claims -- to the effect that, say, kidneys just are sets, or mereological sums, of microparticles of specified types -- are false. But the sophisticated constructionist need fear nothing from this objection, since he is apparently not committed to the objectionable identity claims. The premiss in his argument which corresponds to the

objectionable identity claims states that to be an ordinary object of a certain sort is to be <u>any</u> collection of microparticles of <u>any</u> type, so long as the collection constitutes an individual belonging to a natural or artefactual kind.

So sophisticated conjunctive constructionism is evidently a powerful form of constructionism. But attractive though it is, I think it is a flawed position. It is flawed because it faces an objection, which I shall soon present. But before I do, let us recall the overall goal of the constructionist project. That goal is to show that the eliminative type physicalist is <u>inconsistent</u> to believe in microparticles but not believe in ordinary objects; even given his initial belief only in microparticles, he ought nevertheless to go on to believe also in ordinary objects. Since this is the overall goal of the constructionist project, it is evidently a constraint on any specific constructionist proposal that it not simply <u>assume</u> the existence of ordinary objects in its argument against the eliminative type physicalist. My objection to sophisticated conjunctive constructionism will argue that it falls foul of precisely this constraint.

Here is the first objection. According to the sophisticated conjunctive constructionist, to be (say) a kidney just is to be

any collection of microparticles of any type, just so long as the collection forms an individual belonging to the natural kind, kidney. Now, it is obvious that there is needed some specification of this natural kind, kidney. But how is this specification to be accomplished? One might note, for instance, that a kidney is a stable collection of microparticles, perhaps one forming a certain shape. But such a specification seems far too broad-grained -- we want the specification to distinguish kidneys from, for instance, kidney-shaped lumps of putty. So I think that the sophisticated conjunctive constructionist will want to specify the natural kind, kidney, by reference to a kidney's relations to members of other natural kinds, and in particular to its causal relations to members of other natural kinds. (The specification will therefore be, in one sense of the word, a functionalist one.) We can expect, I think, that the specification, especially where it mentions causal relations, will be intended to exhibit members of the kind in question as behaving in some lawlike way.

At this stage exactly two options confront the sophisticated conjunctive constructionist. For the 'other natural kinds' just mentioned will be either microparticular kinds (i.e. kinds whose existence is acknowledged by the eliminative type

physicalist) or ordinary kinds (i.e. kinds whose existence is denied by the eliminative type physicalist). So if the kind, kidney, is to be specified by its relations, especially causal ones, to members of other natural kinds, then the specification will either mention only microparticular kinds, or else it will mention ordinary kinds as well. (Into this second category falls any specification mentioning only ordinary kinds.) I shall now argue that neither sort of specification may be permitted to the sophisticated conjunctive constructionist.

Consider, first, an attempted specification of the kind, kidney, in which the only kinds mentioned are microparticular kinds. One difficulty here, of course, is that no one is currently in a position to produce such a specification: it is not, after all, the obvious way to go about it. This evidently does not prove that such a specification is impossible. But it does give some, defeasible, reason to doubt its possibility, perhaps even enough to make it rational to suppose that such a specification is more likely to be impossible than possible. And since it is the constructionist who is picking the fight, the burden of proof lies with him to show that such specification is possible.

But considerations of this sort are inconclusive, and I have two other objections to the notion that the sophisticated

conjunctive constructionist's specification of the kind, kidney, can be accomplished in exclusively microparticular terms. The first objection is this. The sophisticated conjunctive constructionist needs to specify the kind, kidney, in order to meet the first objection I raised against the simplest constructionist, the objection that it does not in general follow from the fact that certain things exist that sets, or mereological sums, of those things exist too. The sophisticated constructionist's reply to this is that it does follow -- so long as the set, or mereological sum, of things constitutes a member of a natural kind. Let us now return to the proposed specification of the kind, kidney, in purely microparticular terms. When one attempts, in imagination, to construct such a specification, one is struck by the diversity in all sorts of ways of the physiology of creatures with kidneys. This strongly suggests that any specification in purely microparticular terms of the kind, kidney, would involve huge numbers of disjunctions, each disjunction describing some microparticular state. The more disjunctive the character of such a specification, however, the less plausible it would be to claim that we had successfully specified a natural kind, in an intuitively strong enough sense of 'natural'. (My suspicion is that what is fishy about the disjunctive specification is that it ascribes no autonomous causal efficacy to the kind in

question: all the causal difference it makes is inherited, as it were, from the microphysical states described in the disjuncts.) This matters, because if we have not specified a sufficiently natural natural kind, then the sophisticated conjunctive constructionist will have failed to meet the first objection made to the simplest conjunctive constructionist, the objection to the required principle of construction.

Now for the second objection to the sophisticated conjunctive constructionist's possible proposal to specify the kind, kidney, by mentioning exclusively microparticular kinds. This objection starts with the assumption defended in the last paragraph that such a specification would be bound to be hugely disjunctive. It then continues by asking how we could get to know such a disjunctive specification, and answers by concluding that we could only get to know such a specification in a way which would beg the question against the eliminative type physicalist. The objection is set out in full when, on page 000ff., I present it also against the disjunctive constructionist. I ask the reader kindly to consult that passage.

In the light of these objections to specifying the kind, kidney, by mentioning only microparticular kinds, the sophisticated conjunctive constructionist had better resort to

the other possible sort of specification, that which mentions ordinary kinds (and perhaps ordinary kinds alone). Now the specification of the kind, kidney, by reference to its relations, especially causal relations, to members of ordinary kinds is certainly possible. Here is an example: to be a kidney is to be any collection of microparticles of any type, so long as the collection normally occurs in the upper abdominal cavity of a vertebrate, and separates water and waste products of metabolism from the blood and excretes them as urine through the bladder. But how is the sophisticated conjunctive constructionist to proceed now against the eliminative type physicalist? (Nothing will have been shown to exist merely by the presentation of a definition!) In order to persuade his opponent that kidneys exist, the constructionist must now claim that it is true that some collections of microparticles normally occur in the upper abdominal cavity of a vertebrate, and separate water and waste products etc. But since this claim mentions such things as abdominal cavities, vertebrates and bladders, the truth of the claim must assume the existence of things of these kinds. But things of these kinds are denied to exist by the eliminative type physicalist: not being microparticles, they do not appear in his ontology. So the eliminative type physicalist is well within his rights to deny the claim which the conjunctive constructionist must

assert as true. Qua eliminative type physicalist, he need not accept the claim. Indeed, he can with justice allege that an attempt is being made to beg the question against him.

So much, then, for the second way of specifying the natural kind, kidney. Since there are only two ways, and since, in the light of the constructionist's overall goal, neither way succeeds, we must conclude that the sophisticated conjunctive constructionist cannot supply the needed specification of the kind, kidney. Since this conclusion about the kind, kidney, could easily be extended to other kinds, including artefactual kinds, we must conclude that, despite its attractions, sophisticated conjunctive constructionism fails.

Because of this objection, then, I doubt that the conjunctive constructionist's project, even in its sophisticated form, of vindicating special scientific ontology by trying to construct it out of a narrowly physical ontology can succeed. The reason why such a project appears possible is, I think, an intuition to the effect that there is no more to (say) a table than a collection of microparticles. But I do not have to reject this intuition; on the contrary, I share it. But I account for it in a different way. For I believe that, pace the constructionist, the true content of the intuition is that volumes of space which are commonly thought to contain tables

in fact contain nothing but swarms of microparticles clustered together. This is, I believe, the only sense in which it is true to say that ordinary things, such as tables and kidneys, 'are' collections of microparticles. (For the sense of 'are' here, see Rorty 1965.) But there is nothing here to threaten the eliminative type physicalist.

So much, then, for conjunctive constructionism. Let me turn now to its disjunctive cousin. Disjunctive constructionism is a possible view inspired by (though not stated in) some remarks made by John McDowell. In the course of his attempt to defend the reality of commonsense colour against those who object that nothing like commonsense colour features in our best scientific explanations of why things look coloured, he proposes a modification to the explanatory test for reality:

> ...it would be obviously wrong to suppose that someone who gave such an explanation [sc. a scientific explanation in terms of the physical properties of the putatively red object] could in consistency deny that the object was such as to look red. The right explanatory test is not whether something pulls its own weight in the favoured explanation (it may fail to do so without thereby being explained away), but whether the explainer can consistently deny its reality. (McDowell 1985, p.118)

The modified explanatory test for reality most naturally suggested by McDowell's remarks can perhaps be formulated like

this. A property P is real iff either (a) P is invoked in some favoured explanation or (b) the denial, of some thing, that it has P would be inconsistent with the truth of some favoured explanation (given, I take it, some account of what is involved in having P). And this modification can naturally be seen as leading to a constructionist proposal along these lines: we are justified in believing, not only in those properties which are mentioned in the best explanation of the course of our perceptual experience, but also in those properties the denial of whose existence would be inconsistent with acceptance of the explanatorily indispensable properties.

But I have two objections to this new constructionist proposal. The first is that, like the simplest form of conjunctive constructionism, it rests upon an implausible principle of construction. For if the principle were endorsed, then it would licence an almost limitless expansion of our ontological commitments. The principle leads to this consequence because we can invent any number of gerrymandered properties, defining them in terms of disjunctions of explanatorily indispensable, and hence ontologically respectable, properties, and we will then be committed by the principle, implausibly, to believing in the gerrymandered properties. Suppose that "G" names some gerrymandered

property, and that "P1", "P2",..."PN" name explanatorily indispensable properties. Then we will be able to make claims like this one: X has G iff either X has P1, or X has P2, or..., or X has PN. If we then find out that a thing has, say, P2, it will be impossible, given the definition of what it is to have G, in consistency to deny that the thing also has G, and hence that G is a real property. But this result seems absurd: G is not a real property. And so the disjunctive constructionist proposal can be seen to commit us to a counter-intuitive extension of our ontology.

But there is a second objection to the disjunctive constructionist proposal, an objection which only surfaces when we spell out a little more carefully how the proposal is meant to work. Suppose, then, that we let "G" name some fairly ordinary, special scientific property -- that of being a kidney, say -- which we hope to restore to full ontic respectability, and that in accordance with disjunctive constructionism we define it in terms of some disjunction of microphysical properties. We can then ask the following epistemic question about the resulting definitional claim: how are we supposed to get to know it?

Perhaps it will be said that the claim is analytic, and hence knowable a priori. But it is implausible to say that the claim

is analytic, if only for the reason that neither you nor I nor any living physiologist could even make a start on giving the disjunctive definition, in microphysical terms, of what it is to be a kidney, just by calling upon our grasp of the English word, 'kidney'.

Perhaps it will be said that the definitional claim is a stipulative definition. But then the disjunctive constructionist will have succeeded at best in demonstrating the existence of these stipulatively-defined entities, "kidneys", and will have no warrant for thinking that he has done anything to vindicate the existence of kidneys.

So it very much looks as if the disjunctive constructionist's so-called definitional claim must not be construed as a definition at all, but rather as an empirical claim about kidneys, to be confirmed by the systematic examination of kidneys with a view to determining their constituent parts. The disjunctive constructionist's idea will then have to be that empirical investigation of many kidneys has validated his claiming that to be a kidney is to be either a microphysical thing of type A, or a microphysical thing of type B, or a microphysical thing of type C, or... and so on; for no kidney has yet been found which was not a microphysical thing of type A, or type B, or type C, or ... and so on. But if the

disjunctive constructionist's claims about what it is to be a kidney do have this empirical character, then the eliminative type physicalist will once again be well within his rights to disallow such a premiss from appearing in an argument intended to convince him that, given his narrowly physicalist ontology, he is committed to countenancing also some special scientific ontology. For if such an empirical premiss about kidneys can only be known as a result of the empirical investigation of kidneys, then the existence of kidneys has already been assumed. So the premiss cannot be introduced into a constructionist argument against the eliminative type physicalist, whose narrowly physicalist ontology does not include kidneys, without begging the question against him. The disjunctive constructionist proposal is therefore liable to this second objection.

I conclude, then, that constructionism, in whichever form, is a hopeless cause, and poses no threat to the main premiss, (P1), of the positive argument for eliminative type physicalism.

CHAPTER SEVEN

The Best Explanation: PHYSICS

My defence of the claim that we are justified in taking a property to exist if, but only if, it is mentioned in the best explanation of the course of our perceptual experience is now complete. This claim is, of course, the key premiss, (P1), in the positive argument for eliminative type physicalism. It is time now to turn to the second premiss of that argument, (P2), which states that PHYSICS is the best explanation of the course of our perceptual experience.

My aim in this chapter is the comparatively humble one of making this premiss plausible, which is to say more probable than not. As I emphasised in Chapter Four, it states an empirical claim, and there is no doubt that an ideal defence of it would involve the actual production of PHYSICAL explanations for tracts of perceptual experience. But the production of such explanations lies, unfortunately, beyond my scientific competence. Also, it perhaps assumes something like a grand unified theory of the fundamental forces which we do not as yet possess (see Joseph 1980). So what I shall be trying to do is

to make it reasonable to think that <u>there</u> <u>are</u> such explanations, even though I do not give them. A strategy of this sort can work, since it can be reasonable to think that there are explanations of a given sort, even if one is not presented with them. For instance, it is reasonable to suppose that there is a folk psychological explanation of everything you did yesterday afternoon; but we do not need to give such explanations of what you did for that supposition to be reasonable.

What I have to say on behalf of (P2) falls into two parts. First, I argue that it is at least possible that PHYSICS constitutes the best explanation of the course of our perceptual experience -- nothing of an aprioristic nature rules this out. Secondly, I advance considerations intended to make it more probable than not to think, not only that PHYSICS is a candidate explanation of the course of our perceptual experience, but that it is the best such candidate, i.e. that it is, as we usually put it, the best explanation.

As a final preliminary before proceeding I should note that throughout this chapter I continue to rely upon the assumption noted and discussed in Chapter Four (Section Three), the assumption that it is possible to give a realistic interpretation to quantum mechanics.

1. PHYSICS As A Candidate Explanation

An initial source of resistance to the idea that PHYSICS even could be a candidate explanation of the course of our perceptual experience is the observation that, as a matter of obvious historical fact, physics was not developed explicitly in order to explain perceptual experience. On the contrary, it came into being to explain the behaviour of precisely those macroscopic objects, including laboratory instruments, whose existence I dispute. But resistance to (P2), if derived from this source, is easily overcome. For I do not need to deny the historical facts about the history of physics alluded to. My claim is that physics, and particularly microphysics, supplies the materials for the construction of PHYSICS, and that PHYSICS is at least a candidate explanation for the course of our perceptual experience. The actual historical origins of physics, like the flowers that bloom in the spring, have nothing to do with the case.

But how can this construction be accomplished? By the taking, I believe, of two steps. First of all, it is necessary to take all the evidential statements which tend to confirm physics -- all the statements about the behaviour of putative macroscopic objects, including such things as Geiger counters

and bubble chambers -- and rewrite them using either the "as if" locution or the "so-calling" operator (see Chapter Three, Section Four, pp.110). This will give us a mass of new evidential statements, but these statements will now describe, not the behaviour of macroscopic objects, but the course of our perceptual experience. Consequently, PHYSICS will at least stand a chance of constituting an explanation of this experience. Now for the second step. If PHYSICS is to provide such an explanation, then we must be able to devise an appropriately PHYSICAL way of bridging the gap, as it were, between the putative macroscopic objects and our perceptual experiences. We have, by hypothesis, a physical, and hence acceptably PHYSICAL, explanation of the behaviour of the macroscopic objects. But we need to extend this explanation so that it reaches straight to perceptual experience, without proceeding through the macroscopic middle-men. We need, in other words, a PHYSICAL account of perception. The second step, then, is to suppose that there could be such an account.

Let us consider a ludicrously over-simplified example. Suppose that it seems to me visually that I am seeing a trace on a sensitive plate. The conventional way of explaining this would involve two stages: first, there would be the physical explanation in terms of a particle collision (or whatever) of

the state of the plate; then, there would be an explanation of the perceptual experience in terms of the reflectance properties of the plate and light reflecting off the plate and subsequently striking the retina and thereby triggering a string of neural changes culminating in the experience. My proposal is that we could give a PHYSICAL version of all this. The explanation in terms of the particle collision would stay the same, though it would explain the state of the particular swarm of microparticles which the conventional account calls "the plate": no plates would need to be mentioned in the PHYSICAL version. But then, instead of the conventional account, we would go on to explain the perceptual experience in terms of streams of photons bouncing off the relevant swarm of microparticles and then striking the PHYSICAL equivalent of my retina and thus initiating a string of microphysical changes culminating in my visual experience. I find it plausible to conjecture that there could be a PHYSICAL account along these lines, and it would not matter, in the light of my understanding of the PHYSICAL, if it became necessary to postulate a few new properties in order to obtain it. The last stage of this account assumes the existence of physico-phenomenal causal laws. But later on I explain and defend this assumption, and discuss the ontological status of perceptual experiences (see Chapter Eight, especially Section

Four).

I conclude, then, that PHYSICS is at least a candidate explanation for the course of our perceptual experience. The historical origins of physics, and especially microphysics, do not cast doubt upon this claim, since there appears to be no reason why it should not be possible in principle to construct PHYSICAL explanations of perceptual experiences.

2. PHYSICS As The Best Explanation

I must now argue that PHYSICS is, in addition to being a candidate explanation for the course of our perceptual experience, also the best such explanation. To do this, I shall begin by discussing very briefly the criteria by which we may judge one candidate explanation to be superior to another.

As as often been noticed, it is not easy to say anything very precise about the proper grounds for preferring one candidate explanation to another, though there are a few sketches in the literature (see Quine and Ullian 1970, Chs. 5 and 7; Thagard 1978; Glymour 1984; Lycan 1988, pp.129-130). Nevertheless, it is often intuitively very clear, at least in the simple sort of case considered by philosophers, what we should say. An example of such a case is this. As you are reading, you hear some children playing cricket outside your window and then hear

the sound of smashing glass. Your immediate inference is that a stray ball has broken your window, and you do not even consider the hypothesis that it was broken by a telekinetic Martian with a grudge against you.

What we seem to seek in these cases is a combination of two explanatory virtues, which we can call consilience and economy. Consilience is intended as a measure of how much an explanation explains, and it counts heavily in favour of an explanation that it explains, and thus unifies, a wide variety of data which one would not otherwise have taken to be linked. Economy is a matter of a candidate explanation's involving the fewest kinds of entity, the simplest laws and the fewest ad hoc hypotheses. But consilience and economy are apt to conflict, because a richly consilient explanation may buy its consilience at the price of a reduction in economy. We should therefore prefer whichever candidate explanation enjoys the best balance between the virtues of consilience and economy.

What are the other candidate explanations of the course of our perceptual experience which are rivals to PHYSICS, and whose inferiority to PHYSICS I must show? No one will suggest, I trust, that any one of the special sciences should serve as a rival explanation all by itself -- it would simply not explain enough. More plausibly, we might, in the spirit of pluralism,

take all the special sciences together, and let the conjunction stand as PHYSICS' rival. But this could be improved by the addition to it of PHYSICS, which would enable the explanation of more unusual experiences, such as those as of functioning solar-powered calculators, magnetic phenomena and of course experiments in physics laboratories, as well perhaps as deeper explanations of the behaviour of the postulated special scientific ontology. So the candidate explanation most likely to be pitted against PHYSICS is that consisting of the conjunction of all the special sciences together with PHYSICS.

Now for the adjudication between the two rivals. One thing at least is very clear: whatever exactly economy (or simplicity) comes to, PHYSICS is a more economical explanation than the conjunction of PHYSICS with all the special sciences -- at the very least it involves the postulation of so much less ontology. But it follows from this, given our criteria for the assessment of competing candidate explanations, that we should prefer PHYSICS' rival only if that rival achieves greater consilience than does PHYSICS, and indeed sufficiently greater consilience to make up for its weakness from the point of view of economy. In short, we should prefer PHYSICS' rival only if it explains a great deal more than PHYSICS alone. But this necessary condition is surely not satisfied, and so we

should not prefer PHYSICS' rival but instead stay with PHYSICS unadorned.

PHYSICS certainly looks competent to handle all our everyday perceptual experiences. Suppose, for instance, that it seems to me visually as if there were an orange and a tomato before me. There could be a PHYSICAL explanation of this experience, in terms of the reflection of streams of photons against swarms of microparticles and the ensuing collision of those photons with further swarms of microparticles -- those comprising the PHYSICAL equivalent of my retina and brain -- in which PHYSICAL changes arise which culminate in my visual experience. And fundamentally the same type of explanation would account for both the orange and the red experience. Similarly, there could be a satisfyingly unifying PHYSICAL explanation of why on some occasion it seemed to me, both auditorily and tactually, as if I were tapping a drum with my finger.

In the light of this, it is tempting to suggest the conjecture that PHYSICS could in principle supply an explanation for <u>every</u> experiential event, and not merely the everyday ones; and it not at all clear why we should not give in to this temptation. If we do, then it will not be the case that the conjunction of all the special sciences together with PHYSICS explains <u>anything</u> more than does PHYSICS alone, still

less a great deal more. But even if, more modestly, we allow that for certain experiential events there is no PHYSICAL explanation, it does not follow that PHYSICS' rival can explain more than PHYSICS, since it too might be unable to explain the recalcitrant experiences in question. Indeed, what <u>is</u> there in the course of our perceptual experiences that PHYSICS' rival <u>does</u> explain that PHYSICS alone does not? (There may be phenomena currently inexplicable -- but my question seeks phenomena explicable by PHYSICS' rival but not by PHYSICS.) It is unclear, indeed, that there is anything at all.

The customary tack for the defender of the special sciences would be to reply at this point that PHYSICS' rival explains regularities in the behaviour of special scientific entities. But this reply will not do. It will not do because it is illegitimate to introduce the antics of special scientific entities as explananda. It is illegitimate because, as I argued in Chapters Five and Six, we are justified in taking a property to exist if, but <u>only if</u>, it is mentioned in the best explanation of the course of our <u>perceptual experience</u>, and, with the exception of certain psychological entities, it is uncontroversially true that special scientific entities are <u>not</u> a part of the course of our perceptual experience. For such things as tables, kidneys, rates of interest and ridges of high

pressure are not perceptual sensations.

Why do I make a partial exception of psychology? The answer is that I explicitly affirm the existence of perceptual experiences. These experiences involve what I call phenomenal properties, and it might reasonably be held that these properties are psychological, and hence special scientific, on the grounds that they are properties mentioned by the part of psychology dealing with perception. Nevertheless, these phenomenal properties, unlike other special scientific properties, <u>are</u> part of the course of our perceptual experience: indeed, they constitute it! So they are, on my own showing, perfectly legitimate explananda -- <u>the</u> explananda, in fact. (However, this concession does not entail that I acknowledge any property that is not PHYSICAL. For in Part 3 (Chapter Eight) I argue that phenomenal properties <u>are</u>, i.e. are type-identical with, PHYSICAL properties.)

So the diseconomy of PHYSICS' rival finds no compensation in any greater consilience. The suggestion that it explains regularities in the behavior of special scientific entities is illegitimate and hence fails. We should thus accept PHYSICS alone as constituting the best explanation of the course of our perceptual experience: premiss (P2) of my positive argument for eliminative type physicalism has been defended.

Part 3

CHAPTER EIGHT

Eliminative Type Physicalism and Phenomenal Properties

My task in the three chapters comprising Part 2 was to defend a positive argument for eliminative type physicalism, and that task is now complete. What I want to do in this final chapter is to discuss, and to refute, a traditional objection to physicalism. According to this objection, it is impossible to see how physicalism could accommodate what I shall be calling phenomenal properties. It is these properties that philosophers are refering to when they say that there is something it is like to see a red object or to taste a lemon or to have a headache (see Nagel 1979; Robinson 1982; and Jackson 1982 and 1986), or when they speak of raw feels or qualia or the subjective character of experience. Some of the philosophers who speak in such terms are no doubt in favour of an act-object analysis of mental events of this kind (or these kinds), so that they will be quite happy to talk of sensations, that is, of mental particulars, to which we may stand in some relation of awareness. But they are not obliged to rely on

such an analysis, and the problem they raise, in so far as it is genuine, arises also on an adverbial analysis (see Double 1985).

Eliminative type physicalism states, roughly, that every property is a physical, or rather PHYSICAL, property. But the astute reader will have noticed that the positive argument for eliminative type physicalism outlined in Section Three of Chapter Four does not quite lead to that conclusion. For the positive argument takes its start from the problem of scepticism about the external world, and the key premiss, (P1), is a general principle, not about <u>all</u> properties which we may be justified in believing in, but, more narrowly, about <u>external</u> properties only. So the proper conclusion of that argument, and hence, by my stipulation, the proper statement of eliminative type physicalism, is this: we are justified in taking an <u>external</u> property to exist if, but only if, it is mentioned in the laws of PHYSICS. And it follows from this that I am not in fact committed to holding that internal properties, such as my own phenomenal properties, are PHYSICAL.

This observation might at first sight seem to absolve me from any obligation to discuss the objection from phenomenal properties, on the grounds that I am not committed to affirming the PHYSICAL character of phenomenal properties. In fact, I

suspect that first appearances here may be deceptive. But be that as it may, I do want to extend the thesis of eliminative type physicalism so that it becomes a claim about all properties, and not just the external ones, and this certainly brings me into collision with the traditional objection.

So does the existence of phenomenal properties constitute an insuperable obstacle to the acceptance of the extended thesis of eliminative type physicalism? My eventual answer to this question will be that it does not, since there is no good reason for denying, and one good reason for affirming, that, in a sense to be explained, phenomenal properties <u>are</u> physical, or rather PHYSICAL, properties. (I shall usually speak of physical, rather than PHYSICAL, properties, except where something important hangs on it, in order to reduce distraction.)

1. Solution By Elimination?

The quickest way with the problem of phenomenal properties is simply to deny their existence: if phenomenal properties are not real, then the type physicalist need not undertake the task of showing that they are physical properties. But this way is too quick, and for two reasons I cannot endorse it. First, the denial of the existence of phenomenal properties seems to me to

be absurd. Towards philosophers who dispute their existence I am inclined to adopt the attitude adopted by Aristotle towards certain sceptics: such persons need punishment, not counterargument! In drawing attention to phenomenal properties one is drawing attention to something utterly familiar; so familiar, in fact, that it is perhaps easy to overlook it. (For an argued response to those theories of perception which deny the existence of phenomenal properties, see Jackson 1977 and Goldman 1988, pp.88-113.) Secondly, we need, in considering physicalism of any kind, to keep constantly in mind the question of how we could be justified in believing physicalism to be true. And my view here, of course, is that, as the positive argument defended in Part 2 revealed, the ultimate evidence on the basis of which it is reasonable to accept eliminative type physicalism is in the end constituted by the phenomenal properties exemplified in one's perceptual experience. It follows from this that I am in no position to deny the existence of phenomenal properties, on pain of sawing off the branch on which I am sitting.

It might be said that my way with the denial of the existence of phenomenal properties has itself been too quick. Can anything plausible can be said in favour of it? I know of only two lines of argument here. Neither, as I shall now argue, is

successful.

The first line of argument depends upon an analysis of ordinary language, and seems to underlie many things recently said by Peter Hacker (see Hacker 1987, especially ch.6). The claim about ordinary language is that, if we attend carefully enough to our usage of such words as "looks", "appears" and "seems", we will see that we never use them to refer to anything like phenomenal properties (or individuals). It is then inferred from this claim that there are no phenomenal properties (or individuals). The first point to note is that the linguistic premiss applies only to phenomenal properties of perceptual experience; it could not be used, and doubtless would not be used, in any attempt to deny the reality of a sensation, properly so-called, such as pain. But is the premiss true? I am inclined to doubt it. But however that may be, the most serious problem with the argument is the inference from this premiss. It is evidently unsound, because it is not in general true that properties cannot exist unless and until we have words, in ordinary language, with which to refer to them. Linguistic analysis can doubtless show that some incoherence is involved in the supposition that a certain property exists; for instance the property of being a round, square cupola. And it can also perhaps help to reveal the

ontological commitments of what we ordinarily say; though we may, upon learning of those commitments, decide to revise what we ordinarily say. However, the analysis of language, whether ordinary or extraordinary, is surely incapable by itself of determining which properties exist. It might be said that the aim of the ordinary-language argument is not to prove that phenomenal properties do not exist, but merely to undermine any ordinary-language reason for thinking that they do, by revealing the ontological commitments of ordinary talk to be less extensive than might otherwise appear. But those philosophers who believe in the existence of phenomenal properties typically do not do so, and certainly need not do so, on the basis of an analysis, which might be faulted, of ordinary language. So this first line of argument must be rejected.

The second line of argument, which is perhaps to be found in Churchland (see his 1985a), suggests the possibility that, if we undergo a large enough revision in the concepts in terms of which we understand mental phenomena, and in particular if we come to understand mental phenomena in terms of a mature neuroscience, we will somehow cease to conceive of such phenomena in the ways which tempt us to think of them as exemplifying phenomenal properties; we will instead, perhaps,

think of them spontaneously as brain states of the appropriate kind.

But this possibility, intriguing though it is, does not seem to me to be genuine (cf. Newton 1986). It is certainly possible to learn to call things by new names, so that I could with a little effort learn to complain, upon burning my fingers, that my C-fibres were firing. But if I did learn to do this, it would in no way diminish my inclination to suppose that being in pain was, or at least involved, the exemplification of a phenomenal property. Something else that is possible is that I should learn to discriminate between types of pain between which I could not previously discriminate; and I might learn this with the help of neuroscience. My situation would then closely resemble that of a wine taster (Churchland's example) who has learnt to identify the chemical constituents of a wine by taste alone. But the effect of this training on me would surely be to give me reason to believe in <u>new</u> kinds of phenomenal property which before I had lumped together under the broad heading of pain; and so once again we could continue to believe in phenomenal properties. However, although these two things certainly seem possible, and although the second possibility in particular may illustrate the truth that the very character of the phenomenal properties exemplified in

one's (say) perceptual experience may undergo limited change as a result of training, the acquisition of new information or the adoption of a new theory, still it does not seem possible that any of these factors could bring it about either that one's experiences henceforth ceased to exemplify phenomenal properties at all (for the qualifications here, see Goldman 1982, pp.146-7), or that one would believe that they had never in the past done so. If all this were possible, then one ought to be able to think of models for it, or at least to imagine such models; but none of Churchland's examples seem to be instances of the relevantly strong possibility. This second line of argument, then, seems also to fail.

2. Two Kripkean Arguments

If every property is a physical property, and if phenomenal properties exist, then phenomenal properties must <u>be</u> physical properties: we must be prepared to assert type (or property) identities between the phenomenal and the physical. Is this conclusion shown to be false by any argument reconstructible from Kripke (see Kripke 1980)?

It is possible to reconstruct a number of arguments on the basis of Kripke's text. But I shall consider here only two, which are in my view the most powerful. The first argument

rests upon the doctrine of the necessity of identity; and if successful it refutes any attempted <u>type</u> identity between the phenomenal and the physical. The second argument is an essentialist argument, and is intended to exclude any attempted <u>token</u> identity between the phenomenal and the physical. Since type identities entail token identities, this second argument can be used against type identities too.

Here is the first argument. We need to assume that "pain" and "B" are rigid designators which refer, respectively, to pain and the physical property, presumably some type of brain state, with which it is proposed to identify pain. The first premiss is then this:

> (1) If pain = B, then (a) in every possible world in which pain exists, pain = B and (b) in every possible world in which B exists, pain = B.

The second premiss may be either of the two following propositions:

> (2a) There is a possible world in which pain exists, but in which it's not the case that pain = B.
>
> (2b) There is a possible world in which B exists, but in which it's not the case that pain = B.

And (1), together with either (2a) or (2b), entails the falsity

of the supposition that pain = B.

The argument is evidently valid; but can the premisses be supported? The defence of (1) need not concern us now. Both (2a) and (2b) must presumably be said to follow from the general principle that the conceivability of something entails its logical possibility, and the specific claims that we can conceive of instances of pain where there are no instances of B, and that we can conceive of instances of B where there are no instances of pain; and so they do. Kripke now adds an ingenious twist to the argument. He considers an objector who argues that, if Kripke were right, then eminently respectable scientific property identities, for example the claim that heat = molecular motion, would also be refuted, since we can conceive of instances of heat without molecular motion and vice versa. Kripke replies that there is a crucial disanalogy between the claim that heat = molecular motion and the claim that pain = B. For in the case of heat we cannot, despite appearances, really conceive of instances of heat without molecular motion; what we are really imagining are situations in which something which affects us epistemically just as heat does, but which is not identical with heat, occurs in the absence of molecular motion. In the case of pain, however, an analogous move is not, as they say, available. For any

situation in which something affects us epistemically just as pain does <u>is</u> a situation in which there is pain. Therefore our apparent ability to conceive of instances of pain without instances of B cannot, as it can in the case of heat and molecular motion, be explained away as illusory.

Thus the first Kripkean argument. Let me now explain what is wrong with it. The nub of the matter is whether we really can conceive what Kripke must say that we can conceive if (2a) and (2b) are to be defensible; and the nub of this nub, as it were, is whether it is possible to explain away our apparent conceivings of pain without B, just as, according to Kripke, we can explain away our apparent conceivings of heat without molecular motion. To explain the apparent conceivings away, of course, will be to show that, despite appearances, we are not really conceiving the things in question at all. The resources of Kripke's own theory, I shall now argue, show how we can accomplish precisely this.

Kripke concentrates upon pain, arguing, plausibly, that anything that feels like pain just is pain: this is the source of the crucial disanalogy he alleges between the claim that pain = B and the claim that heat = molecular motion. But if we attend instead to B, the type of brain state with which it is proposed to identify pain, we discover that we can explain away

our apparent conceivings in the case of pain as well as in the case of heat. To see this, let us reflect for a moment upon how we are thinking of B when we engage in the relevant sort of conceiving. We do not, I guess, think of it in any very clear way at all. (Compare the similar protest lodged in Hardin 1987, pp.285-286.) Kripke speaks of the firing of C-fibres, but since most of us have no idea of what C-fibres are like, this leads to no impressive gain in clarity. I imagine, then, that we probably all think of instances of B in some very vague fashion as portions of grey matter assumed to be undergoing certain internal changes. Now for the point. Given the way in which we are bound to think of B and its instances, how are we so sure that we are conceiving of instances of B, when we take ourselves to be conceiving of instances of B without instances of pain, rather than conceiving of instances of some property which is distinct from B, but which affects us epistemically just as B does? And when we appear to be conceiving of instances of pain without instances of B, how are we so sure that those instances are not identical with instances of B, but that in the possible worlds in question B does not affect us epistemically in the way it does in the actual world? (Cf. Maxwell 1978.) The answer to both of my questions is, of course, that we cannot be so sure; just as, according to Kripke, we cannot in the case of heat and molecular motion be

so sure that we are conceiving of heat rather than of something distinct from heat which affects us epistemically just as heat does. By concentrating upon pain, and upon the unusual features both of it and of the term "pain", Kripke has overlooked the fact that the other term of the identity statement, "B", is a rigid designator which, like "heat", has its reference fixed and then passed on by means of definite descriptions which are accidentally, that is, non-essentially, true. It is this which opens up the possibility of explaining away our apparent conceivings of pain without B and B without pain. If B, and "B", were like pain, and "pain", then this Kripkean argument would work; but they are not, so it does not.

I turn now to the second argument reconstructible from Kripke's text. Assume that "M" and "P" are rigid designators referring, respectively, to a token of pain (a pain event) and to a token of the physical type (a physical event) with which it is proposed to identify M. Suppose, now, as the physicalist claims, that

(3) M = P

Since

(4) M is essentially a pain event

whereas

(5) P is not essentially a pain event,

it follows by an application of Leibniz's Law that (3) is false. There is a modal property had by M, namely that of being essentially a pain event, which is not had by P; hence M is not identical with P.

One might entertain a general doubt about this argument on the grounds that events or, if you think this is different, instances of properties simply do not have essential properties (cf. Lycan 1987, pp.16-17). But let us not adjudicate that difficult issue here. The argument is valid; but are (4) and (5) true? The Kripkean appeal must here also be to our apparent conceivings. Let us grant that no particular pain event, say my current headache, could exist without its being a pain event (though it could differ in other ways); this is not a crazy thought. What about (5)? At this point we need to ask a question analogous to that asked in connection with the previous Kripkean argument: how are we thinking of the brain event, P, when we speculate on its essential properties? No doubt the correct response to this query is: in very vague

terms indeed, since few of us know much about the brain. But then our speculation to the effect that P is not essentially a pain event can hardly carry much weight. Indeed, it is on a par with a microphysical layperson's speculation about the essential properties of bosons. In fairness to the token identity theorist, then, we ought to pick upon a really plausible candidate for identification with M, study it seriously, and then consider whether it is essentially a pain event. But then, I suggest, we might well lack the intuition needed to support the relevant premiss analogous to (5), and certainly our current untutored intuitions do not show that we would not. So our current intuitions do not support Kripke's argument. But even if, as I doubt, one retained the relevant intuition having followed the improved procedure that I suggest, this would not really suffice. For if Kripke's argument is to enjoy the requisite generality, if, that is, his argument is to show that M is identical with no physical event of any kind, then the argument must be such that P could designate any physical event at all, and not just your favoured candidate for identification, and that P should still seem to us to lack the property of being essentially a pain event. But how can Kripke guarantee this result? Only, it seems to me, by adopting some general account of what it is to be a physical event, an account which rules it out that a physical event

could be essentially a pain event. But if, as seems overwhelmingly likely, this general account cannot be defended on grounds independent of the current dispute, then the question has simply been begged against the physicalist, and the Kripkean argument loses all its force. Kripke could claim that his intuition was a quite general one to the effect that no physical event of any kind, even of a kind of which he knows nothing, could have the modal property of being essentially a pain event; but my reply to this is to call his bluff, and to deny that he or anyone else really has such an intuition. So Kripke does not give us any reason for thinking that there just could not be a physical event which, as a matter of brute fact about the world, had the property of being essentially a pain event; and if the token identity theorist is right, there is for every pain event a physical event with just this property.

My response to the second Kripkean argument has raised the question of what account we can give of the physical; the question of what is involved in something's being, say, a physical event or a physical property. The question is manifestly of the first importance, both to the defender and to the opponent of the claim that phenomenal properties are physical properties, and I shall presently be recalling my earlier discussion of the issue (see Chapter Four, Section

One). But I want first to consider the fascinating Knowledge Argument against physicalism devised by Frank Jackson (see Jackson 1982 and 1986).

3. The Knowledge Argument

Jackson's argument is advanced in support of the claim that "one can have all the physical information without having all the information there is to have" (Jackson 1982, p.130). His method is to tell us a plausible story from which, he claims, an anti-physicalist moral can be drawn. Here is the story. Mary is a brilliant scientist who learns all the completed physics, chemistry and neurophysiology which there is to know. Perhaps we can add, if it makes sense of such a person, that she is a specialist in the physics etc. of human colour vision. But she has not learnt her science in the usual way. For she has been confined for some reason in an entirely black and white environment, and has learnt only by reading books printed with black and white words and diagrams and by watching black and white television pictures. Then, once upon a time, she is released from her monochrome environment and for the first time in her life sees a ripe tomato. This experience, very naturally, comes as something of a shock to her. It shocks her because she has learnt something new, namely that the visual experiences of the people whom she studied from

inside her black and white world involve something which she did not know they involved, despite her knowledge of all the physical facts about those experiences.

Jackson claims that on the basis of this story we can run the following argument (Jackson 1986, p.293):

> (1) Mary (before her release) knows everything physical there is to know about other people.
>
> (2) Mary (before her release) does not know everything there is to know about other people, since she <u>learns</u> something about them upon her release.
>
> (3) Therefore, there are truths about other people (and herself) which escape the physicalist story.

There are two lines of objection to Jackson's argument which have appeared in the literature, but neither of them is in my view successful. The first line of objection (see, for example, Lewis 1983c, pp.130-2) challenges the truth of (2) by insisting that, upon her release, Mary gains no new knowledge, or at least knowledge-that, at all, but instead gains a set of new <u>abilities</u>; so her knowledge of physics is not thereby shown to have been incomplete. But this carries little conviction. For although no doubt Mary does acquire new abilities upon her release, it does not seem as if this is all she acquires: as the story is told, it seems simply wrong to maintain that she does not in addition acquire new knowledge (see Conee 1985a).

According to the second line of objection, when Mary is released she does not learn of the existence of a new, nonphysical property, but rather is presented in a novel way with, or perhaps gains a new perspective on, a familiar and hence physical property; Jackson's argument therefore requires no supplementation of the physicalist ontology (see Horgan 1984).

Now it is probably true quite generally that one may in some sense have a complete description of a thing and yet fail to recognise it (cf. Warner 1986, p.150). But this cannot serve to account for Mary's surprise upon her release. For she is hypothesised to know, not merely everything about the particular property which (it is suggested) she nonetheless fails to recognise, but everything there is to know about every property there is. But if, as the physicalist claims, she really does have this complete knowledge, then she ought, before her release, to have known all about the many modes of presentation of, and various perspectives on, the property in question. So however we look at it, Mary's knowledge of physics, complete though it is, still seems to have left something out (see Conee 1985a). This line of objection, then, fails.

The objections to Jackson's argument so far considered have been attempts to challenge (2). I propose to challenge (1) (cf. Conee 1985a, p.301): before her release Mary does not know everything physical there is to know about other people. Why does Jackson say that she does? His appeal is plausible enough: "You do not need color television to learn physics or functionalist psychology" (Jackson 1986, p.295). Expressed less tersely, Jackson's defence seems to be that since a tremendous amount of physics can be taught and learnt in black and white, it must all be susceptible of teaching and learning in that way. But we must remember that what we are talking about here is the physics of the brain, and that the brain is on everyone's account one of the most extraordinary objects in the universe. It is therefore rash to assume that what goes for the physics of the extra-cranial world (which is the physics we tend to have in mind) goes also for the physics of the brain.

Jackson needs a better argument than this crude inductive extrapolation. Such an argument might seem to be provided by a consideration of the objectivity of physics. The claim will then be that unless the whole of physics could be learnt from a monochrome television it will not, as it must, be objective in the relevant sense. But can flesh be put onto the bones of

this argument?

Perhaps in this way. It is part of the normative ideal regulating anything we would count as physics that it should be comprehensible in principle to any cognitive agent; but since there could evidently be cognitive agents who lacked colour vision, it follows that there could not be a genuine physics any part of which was fully comprehensible only to cognitive agents who enjoyed colour vision.

But the principle underlying this argument is hard to believe. It is said that there are parts of modern physics which are incredibly difficult to understand. Presumably, then, one has to be a person of remarkable intelligence in order to understand them. Are we to accept that the objectivity of modern physics is put in doubt by the existence of people who, though undoubtedly cognitive agents, are simply insufficiently intelligent to grasp it? Surely not, though this is what the principle implies. There seems, indeed, to be no incoherence in the idea of a cognitive agent, capable of knowledge, who is nonetheless unable, and constitutionally so, to understand the complexities of modern physics.

Let us try another way. Richard Warner argues that no <u>particular</u> sensory capacity must be needed "in order to be able

to assess the truth or falsity of claims of physics" (Warner 1986, p.251; but note his disclaimer on p.252). For if this were not so, then physics "would not be subject to 'the test of independent and impartial criteria, recognising no authority of persons in the realm of cognition'" (Warner 1986, p.251, quoting Israel Scheffler). He continues that "having to rely on, e.g., the authority of red-seers would be to recognise the authority of persons in the realm of cognition. This is why no particular capacity for sensation is required to assess the truth or falsity of physics" (Warner 1986, p.252).

My first objection to this argument is that there is no good reason why having to rely on the observational reports of cognitive agents enjoying a particular sensory capacity lacked by us entails that we are then recognising the authority of persons in the realm of cognition. For to recognise the authority of persons in the realm of cognition is presumably to take the sheer say-so of a person or group of persons as by itself constituting good evidence for some claim; and this would certainly be an abandonment of a normative ideal regulating physics. But we do not have to do this if we rely on the observational reports of red-seers. For their observational reports may be just that, genuine observational reports, and hence quite as satisfactory as our own,

monochrome, observational reports. And there may of course be a large number of red-seers, so that they are able to corroborate one another's reports in whatever way exactly it is in which we non-red-seers corroborate our observational reports. So having to rely on the observational reports of those who enjoy a particular sensory capacity lacked by us has not been shown by Warner to fall foul of any genuine constraint on what may count as physics.

My second objection is that Warner cannot be right to conclude that no particular capacity for sensation is required to assess the truth or falsity of statements of physics. For let us consider any experimental observation which is necessary for the assessment of the truth or falsity of some statement of contemporary physics. We can, I think, conceive of a cognitive agent who, as a matter of brute though regrettable fact, would be <u>destroyed</u> by making this experimental observation. Suppose we discovered such an agent; we would then learn that our own sensory capacities were special, and that our physics required for its assessment the enjoyment of particular sensory capacities had by us but not by our alien friend, even though he was a cognitive agent with sensory capacities. But, contrary to Warner's conclusion, we would not dream of inferring from all this that our physics was somehow

insufficiently objective or undeserving of the laudatory appellation "science".

My response to Jackson's argument, then, consists in refusing to accept his first premiss that, before her release, Mary knows everything physical there is to know about other people. Jackson himself provides no good reason for accepting this premiss, and neither, I have argued, does anyone else. But since this is so, I must protest that the question has simply been begged against the physicalist: Mary cannot know everything physical there is to know about other people, because phenomenal properties are physical properties and she does not know about <u>them</u>!

Though this reply is already adequate to defuse Jackson's argument, let me nevertheless elaborate a little on the charge of question-begging. There are certain physical properties, I say, namely the phenomenal properties, to which Mary has never been exposed in her monochrome world; for she has seen no polychromatic objects and has never been stimulated (we must suppose) to enjoy polychromatic visual sensations. It is as if it were claimed that she had learnt all physics even though she had never been exposed to, say, a magnetic field. This same point can be put by saying that <u>there is a certain kind of experimental observation, namely of those properties which are</u>

phenomenal properties, which Mary has not been permitted to make. Now it may well sound odd to describe Mary's deprivation in this way. But it is not really so. Our concept of observation is, understood most broadly, the concept of being placed in such a way that one can systematically gain information about what is being observed through one's being sensitive to some of the thing's causal consequences. (Perhaps we should say "direct causal consequences", though I doubt that there exists a very principled criterion of directness.) If, therefore, phenomenal properties are, as is plausible, physical properties of the brain, then it is altogether unsurprising that the person best placed to observe them should be the person whose brain it is. As things currently stand, I am simply not causally connected in the right way to your brain for me to be able to observe the phenomenal properties instantiated in your brain; whereas I am appropriately connected to the relevant parts of my own brain.

"As things currently stand": does this mean that it is logically possible for me to observe (to have?) your sensations? I say it is. If there existed a sufficiently ingenious observational instrument by which I could be connected to your brain -- and for all I know this could turn out to be physically or technologically impossible -- , then I

could enjoy the very experiences which you enjoy, and perhaps even at the same time as you do. Of course, such an observational instrument does not now exist, and so Jackson is well within his rights not to include Mary's use of one in his story about her. But this does not entail that she cannot observe phenomenal properties. For she can, in a perfectly intelligible sense, observe her own. But Jackson, question-beggingly, does not allow her to do so. (This reply to Jackson, or something very similar to it, would also tell, I think, against McCullogh 1988.)

4. Phenomenal Properties As Physical Properties

So far I have argued only that two attempts to show that phenomenal properties could not possibly be identical with physical properties do not succeed. I want now to explain in more detail the sense in which, I claim, it is very likely that phenomenal properties are physical, or rather PHYSICAL, properties.

My claim is going to exploit the loose link which I built into my account of what it is to be a PHYSICAL property. It will be recalled that, according to that account, a property is PHYSICAL iff (roughly) it is mentioned in the laws of a science most of whose properties are mentioned in current physics (see

Chapter Four, Section One and, for a qualification inessential here, Section Three). The loose link is signalled by the insertion of the word, "most", which opens up the possibility of including as PHYSICAL properties somehow added to the list of those mentioned in the laws of current physics.

So the next matter to address, before the loose link can be exploited, is the question of the conditions under which we would rightfully add a new property, i.e. admit a new property (a property, that is, either newly discovered or maybe just newly considered) as PHYSICAL. The following seems to me to be correct: a new property may be admitted as PHYSICAL if it is lawfully linked -- linked by a causal law -- to standardly physical properties, i.e. those mentioned in the laws of current physics. If we consider why some recently discovered particle is at once allowed to be physical, then I think we will see that it is for just this reason: the property of being the particle is linked by a causal law to other, standardly physical, properties. And, conversely, if we reflect upon why we do not regard the property of, say, being a kidney as a physical property, then we will see that it is because the property of being a kidney is <u>not</u> lawfully linked -- linked by a causal law -- to standardly physical properties: no law of physics, or of any other science, states that, as a matter of

causal necessity, everything which is a kidney has some standardly physical property. So we do not decide whether a new property is physical by somehow determining the essence of the physical and then seeing if the new property exemplifies it: uncontroversially physical properties are too heterogeneous a group for that to be possible.

According to this formulation of a condition sufficient for a new property's being admitted as PHYSICAL, phenomenal properties can be, and very likely are, PHYSICAL properties. For, as Conee notes in this connection (Conee 1985a, p.301), phenomenal properties seem both to be caused by and to cause standardly physical properties of the brain. And it is presumably because they play this causal role in the brain that we are inclined to say that they help to explain the behaviour of the person. Given their apparent causal role, then, and given also a principle to the effect that events related as cause and effect must each fall under some law (though not necessarily the same law), phenomenal properties are very likely to be linked by a causal law to standardly physical properties, and hence, in accordance with the sufficient condition formulated above, to <u>be</u> PHYSICAL properties. This, then, is the sense in which, I claim, phenomenal properties are PHYSICAL properties. (On the significance of the causal

efficacy of qualia, compare, perhaps, Horgan 1987; and on phenomeno-physical lawlike connections, see Clark 1985, especially p.386, where he speaks of "physiophenomenological generalisations".) It might be objected that the causal laws backing the phenomenal/physical causation do not subsume phenomenal events qua phenomenal, which is what my argument requires, but only qua something else, perhaps physical. But this is most implausible. For, as Crane and Mellor point out (Crane and Mellor 1988, pp.13-14), whole industries depend upon our being able reliably to bring about the same type of phenomenal effect by applying the same type of physical cause. This strongly suggests that phenomeno-physical causal laws exist which subsume the phenomenal qua phenomenal: they link phenomenal with physical properties.

Yet some dissatisfaction may still be felt at this conclusion. Even though, as I have argued, phenomenal properties are on my conception of the physical perfectly respectable physical properties, and hence, trivially, identical with physical properties, it might be thought that what I should have done is show that phenomenal properties are identical with, as it were, pre-existing physical properties; that is, that I should have shown that pain is the firing of C-fibres or some such thing. It should be noted, however, that

nothing I have said is actually inconsistent with the possibility that phenomenal properties are identical with what I have referred to, in a calculated solecism, as pre-existing physical properties. But my main point is that such identity is not necessary: if phenomenal properties are physical properties, then they are identical with physical properties, namely themselves, even though they may be sui generis.

This same dissatisfaction can be expressed in a slightly different fashion. Joseph Levine asks: "What is explained by learning that pain is the firing of C-fibers?" (Levine 1983, p.357) He remarks further that "... there seems to be nothing about C-fiber firing which makes it naturally 'fit' the phenomenal properties of pain.... it makes the way pain feels into merely a brute fact" (p.357). What is Levine getting at here?

I think we should ignore his apparent requirement that identity statements be explanatory. For although identity statements may well figure in acts of explanation, as when we explain the state of some water by first noting that it is a collection of H_2O molecules, it is to laws and not identity statements that we look when we seek explanatory insight. What really concerns Levine, however, is, I think, just this: suppose we have a law which connects instances of pain with

instances of some physical property, e.g. $\Box\forall x[Px \rightarrow Mx]$; then it seems inconceivable that we should be able to come up with a fundamental explanation, at a deeper level than either pain or the physical property, of why this lawlike connection holds; it is left as, in Levine's words, a brute fact. Thus he urges that "... it is the non-intelligibility of the connection between the feeling of pain and its physical correlate that underlies the apparent contingency of that connection" (p.359).

But the demand for intelligibility, or for an explanation of the holding of phenomeno-physical lawlike connections at a deeper level, is not always reasonable to make. There must be a point at which we do just have to accept it as a fact, as if you like a brute fact, that physical properties are lawfully linked to one another in certain ways. My suggestion is precisely that phenomenal properties and physical properties are lawfully linked in this brutal fashion. Is there any reason to think that we just could not have an instance of this here? I think not, though Levine mentions a relevant consideration. Since, as he claims, "... the phenomenon of consciousness arises on the macroscopic level" (p.358), we should expect there to be an account of it at a lower level.

Now I agree with Levine entirely that it is incredibly hard to understand how something like pain, in all its phenomenal

aspects, could be given any sort of plausible account in terms of the motions of microphysical particles; though not all physical properties are particles, and we would doubtless do better to think of those fundamental physical properties that are phenomenal properties more on the model of, say, a magnetic or gravitational field (cf. Maxwell 1978 for the suggestion; and see Sellars 1963 for the problem, especially pp.26, 34-6.). But Levine is wrong to expect such a microphysical account of consciousness, even if it is a good principle to expect microphysical explanations of all macroscopic phenomena, because there is no particular reason to claim, as he does, that consciousness arises on the macroscopic level. Of course, the phenomena of consciousness are manifested in the form of their causal consequences at the macroscopic level -- in bodily movements; but then so are microphysical phenomena. (Indeed it is precisely because of this that the macroscopic can be explained in terms of the microphysical.) But in ordinary language, which is all we have to guide us in this matter, phenomenal properties are ascribed to persons, rather than to human bodies; and since it is a controversial thesis that persons just are bodies, it is very far from obvious at what level we should take persons (as opposed to bodies), and hence their phenomenal properties, to be.

Levine's demand that phenomeno-physical lawlike connections should be intelligible is encouraged by certain theoretical identities with which we are all familiar; for example, the claim that water = H_2O, or that heat = molecular motion, or that lightning = electrical discharge. Since these theoretical identities are all microreductions, and involve explaining the moderately-sized in terms of the very small, it is tempting to conclude that all theoretical identities must be like this, including mental/physical identities. But this is not a temptation to which there is, I think, any good reason to succumb: not all theoretical identities have to resemble the familiar examples.

Perhaps I should add that although, as I have been arguing, the admitted non-intelligibility of phenomeno-physical lawlike connections does not make it incorrect to regard phenomenal properties as physical properties, I do not mean to suggest that these connections raise no philosophical problems. On the contrary, they have been taken by some philosophers to constitute the basis of an argument for theism (see Swinburne 1979, ch.9 and Adams 1987, ch.16). Nothing I have said prevents these philosophers from being right.

5. Phenomenal Properties and Multiple Realisability

I have been defending type physicalism in respect of phenomenal properties, and I want to conclude by considering a standard argument against such a view which is often advanced by functionalists. This is the argument from multiple realisability. (Cf. Lycan's "shyness" argument against Sellars in his 1987, pp.103-4.) It is said that if, say, being in pain is identified with some physical property, then we will be debarred from being able to claim that creatures whose physiologies differ significantly from our own can be in pain. Hence type physicalism cannot cope with the fact that a given mental state, say, pain, can be realised in many ways.

Let us allow, as approximately true, that the type physicalist is genuinely committed in the way suggested by this argument. The argument can still be rejected by denying the alleged fact of the multiple realisability of phenomenal states. For what reason do we have for supposing that there do exist pain-feeling creatures with physiologies relevantly and significantly different from our own? It is customary at this point to mention molluscs and Martians. But this should hardly persuade us. It is very far from clear that we want to say that molluscs feel pain (whatever states functionally equivalent to our own they may be in) in the way that we humans do; and of course we have never met any Martians. Perhaps it

will be replied that it is only the logical possibility of such Martians which is being claimed. This may well be so; but then the type physicalist response will be that type physicalism is perfectly compatible with this logical possibility. For phenomeno-physical lawlike connections, in common with all lawlike connections, are alleged to hold only contingently and not as a matter of logical necessity. So the lawlike connections which, as things actually are, hold between pain and some physical property instantiated by human brains might, in some possible world, hold between pain and some other property instantiated only in Martian brains. (And this same contingency, incidentally, also entails that the type physicalism I have been defending is fully consistent with the logical possibility of the inverted spectrum.)

Finally, I cannot forbear from emphasising that, and why, functionalism, which is generally adopted by supporters of the argument from multiple realisability, does not handle phenomenal properties in a satisfactory way. The difficulty for functionalism is not so much the question of whether there can be a non-circular functionalist analysis of phenomenal properties, a question which has been discussed more than enough (see, however, Conee 1985b). The problem, rather, is whether a functionalist can, without resorting to ad hoc type

identity claims, give an explanation of what would otherwise be the very puzzling fact that although, as everyone allows, the realisers of simple functional states, such as Coke machines, do not instantiate phenomenal properties, the realisers of exceedingly complex functional states, such as (it is said) persons, do instantiate such properties. For, like others, I cannot see what this explanation, if it deploys only resources available to the functionalist qua functionalist, could possibly be.

6. Conclusion

My defence, both negative and positive, of the extended thesis of eliminative physicalism -- the thesis that every property, and not just every external one, is a PHYSICAL property -- is now complete. On the negative side, I have been arguing, with specific reference to Kripke and Jackson, that there is no good reason -- no reason founded upon a principled account of what it is to be physical -- to deny that phenomenal properties are physical properties. And then, positively, I argued that an independently plausible account of the physical, that which identifies it with the PHYSICAL, permits, and very probably obliges, the identification of phenomenal with physical, or rather PHYSICAL, properties. Eliminative type physicalism is not refuted by the existence of phenomenal

properties.

CONCLUSION

> ...science is the measure of all things, of what is that it is, and of what is not that it is not. (Sellars 1963, p.173)

Eliminative type physicalism, like other brands of physicalism, is nothing less than a large-scale view as to the nature of reality. And so in aiming in this thesis to defend it, I have set myself an ambitious task. In this Conclusion, then, I intend to take stock -- to summarise what I think has been achieved, and to indicate some of those areas where further work must, or might, be done.

By exhibiting the varieties of physicalism and anti-physicalism, in the Introduction, as responses to the quite general problem of the many sciences, I set up a standard which any view supposed to supplant eliminative type physicalism would have to meet -- that of constituting a response to the problem. And in Part 1 I exploited this standard. My overall strategy was to argue that in rejecting eliminative type physicalism a critic must adopt one of three comparably general rival views, and that each of these rival views was defective. Specifically, I argued in Chapter One against token physicalism, first, that it leads to belief in unappealingly massive overdetermination and, secondly, that

there seems to be no satisfactory account of how we could come to know the claims of token identity affirmed by the token physicalist. In Chapter Two I argued that the key notion in supervenience physicalism, that of supervenience itself, cannot do the work expected of it, since the metaphysical mystery it involves can only be resolved by construing the special scientific properties as unreal projections. Finally, in Chapter Three, I described the pluralism to which one is led by a systematic desire to avoid being a physicalist, and argued that it leads to an indefensible anti-realism and anti-empiricism.

I do not claim to have refuted any of these rival positions -- refutations in philosophy are rare. I do claim to have revealed the costs, perhaps unexpectedly high, of rejecting eliminative type physicalism and hence embracing one of its rivals. If eliminative type physicalism seemed initially to be a non-starter, then I hope that Part 1 established it firmly in the race.

But of course I suspect that eliminative type physicalism deserves to win, and so I devoted Part 2 to an exposition and defence of a positive argument for it. Since I understand the varieties of physicalism and anti-physicalism as large-scale claims concerning what there is, I founded the argument on the

traditional problem of scepticism about the external world. Indeed, I believe that any defensible alternative to my view must mesh with an epistemology adequate to address traditional sceptical concerns. The positive argument comprised two stages. In the first, I gave reasons to adopt a strong version of the explanatory test for reality according to which only explanatorily indispensable properties can justifiably be said to exist. In the second stage, I proposed that this explanatory test, when conjoined with an empirical speculation, supports eliminative type physicalism.

This positive argument was only really sketched, though some parts of it were supplied with tougher argument than others, and I am aware that at many points it is open to objection. But I am also aware that it would have been a waste of time to elaborate it in greater detail than I did if the general idea lying behind it was not made plausible. My aim was to establish the plausibility of the general idea. Moreover, as I emphasised at the time, the positive argument rests squarely, at its second stage, upon empirical considerations which may simply fail to turn out to vindicate my position. Should the evidence go against me, I would have simply to retire gracefully.

In Part 3 I tried to confront directly the widely-supported

objection, to any kind of physicalism, that the world contains such things as qualia, or the subjective character of experience, which are not physical. I argued in response that no critic of physicalism has successfully articulated an interesting understanding of what it is to be physical from which it follows that phenomenal properties (as I called them) could not be, or merely are not, physical. And I offered a reason, resting, as does the whole of Part 2, upon an empirical conjecture, for maintaining that phenomenal properties <u>are</u> physical properties.

There are many topics which a project of this ambitious sort ought to touch upon, but which I have in this thesis omitted to discuss: I mention only four. First, I have taken for granted throughout the thesis some version of the nomological principle of causality. Any thesis must start with some assumptions, and this is one of mine. More specifically, I assume that an adequate account of the nature of causality will imply some suitable version of the nomological principle of causality. So there is further work to be done here.

Secondly, I give no clear indication of what view I take -- or even should take -- of the semantics of, in particular, theoretical terms. In view of the unusually large number of empty terms I take our language to contain, and of several

other distinctive features of my overall position, the development of an appropriate account of the semantics of theoretical terms may not be easy. But I take it that I am committed to the possibility of such an account. For the record, I should indicate my sympathy with the account offered by David Lewis (see his 1983, pp.78-95).

A third topic for further research would be the question of how, if at all, an eliminative type physicalist can accommodate the epistemic normativity on which I lay such heavy emphasis. The issue here, of course, is that of whether epistemology can be, in a suitably generous sense, naturalised. The issue has its own obvious interest, but I believe in addition that its resolution may also make it possible to defend my response to external world scepticism against a certain class of objections.

Finally, I should mention the propositional attitudes. Although I defend a view as to the place of phenomenal properties in an exclusively physical world, I do not, of course, indicate what I propose to do with the propositional attitudes; and their right of abode in the physical realm is not obvious. But the questions raised by the attitudes are profound and difficult -- it is no accident that the problem of physicalism has traditionally been considered to pertain

especially to the philosophy of mind -- and I myself see no immediate prospect of any successful answers to them, still less any answers congruent with the strict and exotic demands of the eliminative type physicalist.

Howard Robinson suggests that materialism is part of a picture of the world that is "hypnotising but terrifying: the world as a machine of which we are all insignificant parts" (1982, p.125). And of course he is quite right. But what if materialism is true? This is the awful possibility which I hope I have shown we must at least consider.

> I always think that when one feels one's been carrying a theory a little too far, then's the time to carry it a little further.
>
> A little further? Good heavens man! Are you growing old? (Max Beerbohm)

Bibliography

Adams, R. M. The Virtue of Faith. Oxford: Clarendon Press, 1987.

Ayer, A. J. The Central Questions of Philosophy. Harmondsworth: Penguin, 1973.

Benacerraf, P. "Mathematical Truth," in P. Benacerraf and H. Putnam (eds.) Philosophy of Mathematics. Cambridge: Cambridge University Press, 1983.

Blackburn, S. Spreading the Word. Oxford: Clarendon Press, 1984.

Bonevac, D. "Supervenience and Ontology." American Philosophical Quarterly 25 (1988): 37-47.

Boyd, R. "The Current Status of Scientific Realism," in J. Leplin (ed.) Scientific Realism. California: University of California Press, 1984, 41-82.

-----. "Observations, Explanatory Power, and Simplicity: Toward a Non-Humean Account," in P. Achinstein and O. Hannaway (eds.) Observation, Experiment, and Hypothesis in Modern Physical Science. Cambridge, Mass.: The MIT Press, 1985, 47-94.

Brand, M. "Particulars, Events and Actions," in M. Brand and D. Walton (eds.) Action Theory. Dordrecht, Holland: D. Reidel Publishing Company, 1976. 133-157.

Brueckner, A. L. "Transcendental Arguments I." Nous 17 (1983): 551-575.

-----. (a) "Transcendental Arguments II." Nous 18 (1984): 197-225.

-----. (b) "Putnam's Model-Theoretic Argument Against Metaphysical Realism." Analysis 44 (1984): 134-140.

-----. (a) "Charity and Scepticism." Pacific Philosophical Quarterly 67 (1986): 264-268.

-----. (b) "Brains In A Vat." Journal of Philosophy 83 (1986): 148-167.

Cartwright, N. How the Laws of Physics Lie. Oxford: Clarendon Press, 1983.

Cassam Q. "Transcendental Arguments, Transcendental Synthesis and Transcendental Idealism." Philosophical Quarterly 37 (1987): 355-378.

Churchland, P. M. "Eliminative Materialism and the Propositional Attitudes." Journal of Philosophy 78 (1981):

67-90.

-----. (a) "Reduction, Qualia and the Direct Introspection of Brain States." Journal of Philosophy 82 (1985): 8-28.

-----. (b) "The Ontological Status of Observables: In Praise of the Supraempirical Virtues," in P. Churchland and C. A. Hooker (eds.) Images of Science. Chicago: University of Chicago Press, 1985, 35-47.

Clark, A. "Qualia and the Psychophysiological Explanation of Color Perception." Synthese 65 (1985): 377-405.

Cohen, G.A. Karl Marx's Theory of History. Oxford: Oxford University Press, 1978.

-----. "Functional explanation, consequence explanation and game theory." Inquiry 25 (1982): 27-56.

Conee, E. (a) "Physicalism and Phenomenal Qualities." Philosophical Quarterly 35 (1985): 296-302.

-----. (b) "The Possibility of Absent Qualia." Philosophical Review 94 (1985): 345-366.

Crane, T. and Mellor, D. H. "There Is No Question Of Physicalism" (1988) in a forthcoming collection edited by R. Warner and S. Wagner.

Davidson, D. Essays on Actions and Events. Oxford: Clarendon Press, 1980.

-----. Inquiries into Truth and Interpretation. Oxford: Clarendon Press, 1984.

-----. "A Coherence Theory of Truth and Knowledge," in E. LePore (ed.) Truth and Interpretation. Oxford: Basil Blackwell, 1986, 307-319.

Davies, S. "Kripke, Crusoe and Wittgenstein." Australasian Journal of Philosophy 66 (1988): 52-66.

Dennett, D. "Three Kinds of Intentional Psychology," in R. Healey (ed.) Reduction, Time and Reality. Cambridge: Cambridge University Press, 1981.

Devitt, M. Realism and Truth. Oxford: Basil Blackwell, 1984.

-----. "Rorty's Mirrorless World." Midwest Studies in Philosophy, Vol. XII: Realism and Antirealism (1988): 157-177.

Devitt M. and Sterelny, K. Language and Reality. Oxford: Basil Blackwell, 1987.

Double, R. "Phenomenal Properties." Philosophy and Phenomenological Research 45 (1985): 383-392.

Dupre, J. "Materialism, Physicalism, and Scientism." Philosophical Topics 16 (1988): 31-56.

Eddington, A. The Nature of the Physical World. Cambridge: Cambridge University Press, 1929.

Elster, J. "Review of Cohen 1978." Political Studies 28 (1980): 121-8.

-----. "Marxism, functionalism and game theory." Theory and Society 11 (1982): 453-82.

-----. "Further thoughts on Marxism, functionalism and game theory." In Roemer, J. (ed.) Analytical Marxism. Cambridge: Cambridge University Press, 1986.

Feyerabend, P. "Explanation, Reduction and Empiricism," in his Realism, Rationalism and Scientific Method. Cambridge: Cambridge University Press, 1981.

Field, H. "Mental Representation." Erkenntnis 13 (1978): 9-61.

-----. Science Without Numbers. Oxford: Basil Blackwell, 1980.

Fodor, J. "Special Sciences," in N. Block (ed.) Readings in Philosophy of Psychology. London: Methuen, 1980. I: 120-133.

-----. "Observation Reconsidered." Philosophy of Science 51 (1984): 23-43.

-----. Psychosemantics. Cambridge, Mass.: The MIT Press, 1987.

-----. "A Reply to Churchland's 'Perceptual Plasticity and Theoretical Neutrality'." Philosophy of Science 55 (1988): 188-198.

Forrest, P. (a) "Supervenience: The Grand-Property Hypothesis." Australasian Journal of Philosophy 66 (1988): 1-12.

-----. (b) Quantum Metaphysics. Oxford: Basil Blackwell, 1988.

Glymour, C. "Explanation and Realism," in J. Leplin (ed.) Scientific Realism. California: University of California Press, 1984, 173-192.

Goldman, A. H. "Epistemic Foundationalism and the Replaceability of Observation Language." Journal of Philosophy 79 (1982): 136-154.

-----. Empirical Knowledge. California: University of California Press, 1988.

Goldman, A. I. A Theory of Human Action. Englewood Cliffs, NJ: Prentice Hall, 1970.

Goodman, N. Problems and Projects. Indianapolis: Bobbs Merrill, 1972.

-----. Ways of Worldmaking. Sussex: Harvester Press, 1978.

-----. "On Starmaking." Synthese 45 (1980): 211-215.

-----. Of Mind and Other Matters. Cambridge, Mass.: Harvard University Press, 1984.

Grandy, R. "Reference, Meaning and Belief." Journal of Philosophy 70 (1973): 439-452.

Grimes, T. "Explanation and the Poverty of Pragmatics." Erkenntnis 27 (1987): 79-82.

-----. "The Myth of Supervenience." Pacific Philosophical Quarterly 69 (1988): 152-160.

Hacker, P. M. S. Appearance and Reality. Oxford: Basil Blackwell, 1987.

Hardin, C. L. "Qualia and Materialism: Closing the Explanatory Gap." Philosophy and Phenomenological Research 48 (1987): 281-298.

Harman, G. The Nature of Morality. New York: Oxford University Press, 1977.

Haugeland, J. "Weak Supervenience." American Philosophical Quarterly 19 (1982): 93-103.

Hill, C. "Why Cartesian Intuitions Are Compatible With The Identity Thesis." Philosophy and Phenomenological Research 42 (1981): 254-265.

-----. "In Defence of Type Materialism." Synthese 59 (1984): 295-320.

Honderich, T. "The Argument for Anomalous Monism." Analysis 42 (1982): 59-64.

Hooker, C.A. "Towards A General Theory of Reduction." Dialogue 20 (1981): 38-59; 201-36; 496-529.

Horgan, T. "Token Physicalism, Supervenience, and the Generality of Physics." Synthese 49 (1981): 395-413.

-----. "Supervenience and Microphysics." Pacific Philosophical Quarterly 63 (1982): 29-43.

-----. "Jackson on Physical Information and Qualia." Philosophical Quarterly 34 (1984): 147-152.

-----. "Supervenient Qualia." Philosophical Review 96 (1987): 491-520.

Hornsby, J. "Which Physical Events Are Mental Events?" Proceedings of the Aristotelian Society LXXXI (1980-1): 73-92.

-----. "Physicalism, Events and Part-Whole Relations," in E. LePore and B. McLaughlin (eds.) Actions and Events. Oxford: Basil Blackwell, 1985, 444-458.

Humphreys, P. The Chances of Explanation. Princeton: Princeton University Press, 1989.

Jackson, F. Perception. Cambridge: Cambridge University Press, 1977.

-----. "Epiphenomenal Qualia." Philosophical Quarterly 32 (1982): 127-136.

-----. "What Mary Didn't Know." Journal of Philosophy 83 (1986): 291-295.

Joseph, G. "The Many Sciences and the One World." Journal of Philosophy 78 (1980): 773-791.

Kim, J. "Events as Property Exemplifications," in M. Brand and D. Walton (eds.) Action Theory. Dordrecht, Holland: D. Reidel Publishing Company, 1976. 159-177.

-----. "Supervenience and Nomological Incommensurables." American Philosophical Quarterly 15 (1978): 149-156.

-----. "'Strong' and 'Global' Supervenience Revisited." Philosophy and Phenomenological Research 48 (1987): 315-326.

-----. (a) "The Myth of Nonreductive Materialism." Proceedings and Addresses of the American Philosophical Association 63 (1989): 31-47.

-----. (b) "Explanatory Realism, Causal Realism, and Explanatory Exclusion." Midwest Studies in Philosophy 12 (1988): 225-240.

Kitcher, P. and Salmon, W. "Van Fraassen on Explanation." Journal of Philosophy 84 (1987): 315-330.

-----. (eds.) Scientific Explanation. Minnesota Studies in the Philosophy of Science (Vol. 13). Minneapolis: University of Minnesota Press, 1989.

Klagge, J. "Supervenience: Ontological and Ascriptive." Australasian Journal of Philosophy 66 (1988): 461-470.

Kordig, C. The Justification of Scientific Change. Dordrecht, Holland: D. Reidel Publishing Company, 1971.

Kripke, S. Naming and Necessity. Oxford: Basil Blackwell, 1980.

-----. Wittgenstein on Rules and Private Language. Oxford: Basil Blackwell, 1982.

Levin, J. "Must Reasons Be Rational?" Philosophy of Science 55 (1988): 199-217.

Levin, M. "Kripke's Argument Against the Identity Thesis." Journal of Philosophy 72 (1975): 149-167.

-----. "The Extensionality of Causation and Causal-Explanatory Contexts." Philosophy of Science 43 (1976): 266-277.

-----. (a) Metaphysics and the Mind-Body Problem. Oxford: Oxford University Press, 1979.

-----. (b) "On Theory-Change and Meaning-Change." Philosophy of Science 46 (1979): 407-424.

Levine, J. "Materialism and Qualia: The Explanatory Gap." Pacific Philosophical Quarterly 64 (1983): 354-361.

Lewis, D. (a) "New Work for a Theory of Universals." Australasian Journal of Philosophy 61 (1983): 343-377.

-----. (b) "An Argument for the Identity Theory," in his Philosophical Papers. Oxford: Oxford University Press, 1983, I.

-----. (c) "Postscript to 'Mad Pain and Martian Pain'," in his Philosophical Papers. Oxford: Oxford University Press, 1983, I.

-----. "Putnam's Paradox." Australasian Journal of Philosophy 62 (1984): 221-236.

-----. (a) "Causal Explanation," in his Philosophical Papers. Oxford: Oxford University Press, 1986, II.

-----. (b) On the Plurality of Worlds. Oxford: Basil Blackwell, 1986.

Lycan, W. G. Consciousness. Cambridge, Mass.: The MIT Press, 1987.

-----. Judgment and Justification. Cambridge: Cambridge University Press, 1988.

Macdonald, G. and Pettit, P. Semantics and Social Science. London: Routledge and Kegan Paul, 1981.

Mackie, J. L. Problems from Locke. Oxford: Clarendon Press, 1976.

-----. Ethics: Inventing Right and Wrong. Harmondsworth: Penguin, 1977.

-----. The Miracle of Theism. Oxford: Clarendon Press, 1982.

Maxwell, G. "Scientific Methodology and the Causal Theory of Perception," in I. Lakatos and A Musgrave (eds.) Problems in the Philosophy of Science. Holland: North Holland Publishing Company, 1968, 148-160.

-----. "Rigid Designators and Mind-Brain Identity." Minnesota Studies in the Philosophy of Science Vol. 9: Perception and Cognition C. Wade Savage (ed.) (1978): 365-403.

McCulloch, G. "Scientism, Mind, and Meaning," in P. Pettit and J. MacDowell (eds.) Subject, Thought and Context. Oxford: Clarendon Press, 1986.

-----. "What It Is Like." Philosophical Quarterly 38 (1988): 1-19.

McDowell, J. "Values and Secondary Qualities," in T. Honderich (ed.) Morality and Objectivity. London: Routledge and Kegan Paul, 1985, 110-129.

McGinn, C. "Charity, Interpretation and Belief." Journal of Philosophy 74 (1977): 521-535.

-----. Wittgenstein on Meaning. Oxford: Basil Blackwell and The Aristotelian Society, 1984.

-----. "Radical Interpretation and Epistemology," in E. LePore (ed.) Truth and Interpretation. Oxford: Basil Blackwell, 1986, 356-368.

McIntyre, J. "Putnam's Brains." Analysis 44 (1984): 59-61.

McMichael, A. "Creative Ontology and Absolute Truth." Midwest Studies in Philosophy Vol. XII: Realism and Antirealism (1988): 51-74.

Mellor, D. H. "Physics and Furniture" in N. Rescher (ed.) Studies in the Philosophy of Science. Oxford: Basil Blackwell and the University of Pittsburgh, 1969, 171-187.

Miller, R. W. Fact and Method. Princeton: Princeton University Press, 1987.

Nagel, T. "What is it Like to be a Bat?" in his Mortal Questions. Cambridge: Cambridge University Press, 1979.

Newton, N. "Churchland on Direct Introspection of Brain States." Analysis 46 (1986): 97-102.

Newton-Smith, W. "The Underdetermination of Theory by Data." Proceedings of the Aristotelian Society: Supplementary Volume

52 (1978): 71-91.

-----. "Trans-theoretical Truth Without Transcendent Truth?" in D. Henrich (ed.) Kant oder Hegel?. Stuttgart: Klett-Cotta, 1983, 466-478.

Nussbaum, Martha C. Aristotle's De Motu Animalium. Princeton: Princeton University Press, 1978.

Owens, D. "Levels of Explanation." Mind 98 (1989): 59-79.

Peacocke, C. Sense and Content. Oxford: Clarendon Press, 1983.

Petrie, B. "Global Supervenience and Reduction." Philosophy and Phenomenological Research 48 (1987): 119-130.

Plantinga, A. "How To Be An Anti-Realist." Proceedings of the American Philosophical Association 56:1 (1982): 47-70.

Pollock, J. Contemporary Theories of Knowledge. London: Hutchinson, 1987.

Primas, H. Chemistry, Quantum Mechanics and Reductionism. New York: Springer-Verlag, 1981.

Putnam, H. "A Philosopher Looks at Quantum Mechanics," in R. Colodny (ed.) Beyond the Edge of Certainty. Englewood Cliffs,

New Jersey: Prentice-Hall, 1965, 75-101.

-----. "Realism and Reason," in Meaning and the Moral Sciences. London: Routledge and Kegan Paul, 1978.

-----. "Philosophy and Our Mental Life," in N. Block (ed.) Readings in Philosophy of Psychology. London: Methuen, 1980. I: 134-143.

-----. Reason, Truth and History. Cambridge: Cambridge University Press, 1981.

-----. Realism and Reason. Cambridge: Cambridge University Press, 1983.

-----. The Many Faces of Realism. Illinois: Open Court, 1987.

Quine, W. V. Theories and Things. Cambridge, Mass.: Harvard University Press, 1981.

Quine, W. V. and Ullian, J. S. The Web of Belief. New York: Random House, 1970.

Rasmussen, S. A. "The Intelligibility of Abortive Omniscience." Philosophical Quarterly 37 (1987): 315-319.

Resnik, M. "You Can't Trust an Ideal Theory to Tell the Truth." Philosophical Studies 52 (1987): 151-160.

Robinson, H. Matter and Sense. Cambridge: Cambridge University Press, 1982.

-----. "The General Form of the Argument for Berkeleian Idealism," in J. Foster and H. Robinson (eds.) Essays on Berkeley. Oxford: Oxford University Press, 1985, 163-186.

Rorty, R. "Mind-Body Identity, Privacy, and Categories." Review of Metaphysics 19 (1965): 24-54.

Ryle, G. Dilemmas. Cambridge: Cambridge University Press, 1954.

Salmon, W. Scientific Explanation and the Causal Structure of the World. Princeton: Princeton University Press, 1984.

Scheffler, I. "The Wonderful Worlds of Goodman." Synthese 45 (1980): 201-209.

Sellars, W. Science, Perception and Reality. London: Routledge and Kegan Paul, 1963.

Stebbing, L. S. Philosophy and the Physicists. Harmondsworth: Penguin, 1944.

Stich, S. "Could Man Be An Irrational Animal?" in H. Kornblith (ed.) Naturalising Epistemology. Cambridge, Mass.: The MIT Press, 1985, 249-267.

Strawson, P. F. "Perception and Its Objects," in G. F. Macdonald (ed.) Perception and Identity. New York: Cornell University Press, 1979, 41-60.

-----. Skepticism and Naturalism: Some Varieties. London: Methuen, 1985.

Stroud, B. "Transcendental Arguments." Journal of Philosophy 65 (1968): 241-256.

-----. "The Physical World." Proceedings of the Aristotelian Society 87 (1986-7): 263-277.

Swinburne, R. The Existence of God. Oxford: Clarendon Press, 1979.

-----. "Are Mental Events Identical With Brain Events?" American Philosophical Quarterly 19 (1982): 173-181.

Thagard, P. "The Best Explanation: Criteria for Theory Choice." Journal of Philosophy 75 (1978): 76-92.

Van Fraassen, B. The Scientific Image. Oxford: Clarendon Press, 1980.

Vision, G. Modern Anti-Realism and Manufactured Truth. London: Routledge, 1988.

Warner, R. "A Challenge to Physicalism." Australasian Journal of Philosophy 64 (1986): 249-265.

Wiggins, D. "Truth, Invention and the Meaning of Life." Proceedings of the British Academy 62 (1976): 331-378.

-----. Sameness and Substance. Oxford: Basil Blackwell, 1980.

Wilkes, K. Physicalism. London: Routledge and Kegan Paul, 1978.

Wittgenstein, L. Philosophical Investigations. Oxford: Basil Blackwell, 1958.

Wright, C. "Does Philosophical Investigations I. 258-60 Suggest A Cogent Argument Against Private Language?" in P. Pettit and J. McDowell (eds) Subject, Thought and Context. Oxford: Clarendon Press, 1986, 209-259.

Wright, J. "Realism, Verificationism, and Underdetermination." Southern Journal of Philosophy 23 (1985): 503-529.